Evo-devo of flowers and flowering genes: *Salvia* as a case study

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN
(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften
des Karlsruher Instituts für Technologie (KIT)
genehmigte

DISSERTATION
von

Sascha Wetters

aus
Freudenstadt

Referent: Prof. Dr. Peter Nick
Korreferent: Prof. Dr. Tilman Lamparter
Mündliche Prüfung: 20. Juli 2020

Die vorliegende Dissertation wurde am Botanischen Institut des Karlsruher Instituts für Technologie (KIT), Lehrstuhl I für molekulare Zellbiologie, im Zeitraum von Juli 2016 bis Juni 2020 angefertigt.

I. Statement

Hiermit erkläre ich, dass ich die vorliegende Dissertation, abgesehen von der Benutzung der angegebenen Hilfsmittel, selbständig verfasst habe.

Alle Stellen, die gemäß Wortlaut oder Inhalt aus anderen Arbeiten entnommen sind, wurden durch Angabe der Quelle als Entlehnungen kenntlich gemacht.

Diese Dissertation liegt in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vor.

Karlsruhe, den 10. Juni 2020

Sascha Wetters

II. Acknowledgement

To quote Bilbo Baggins when he left the Shire: “I’m going on an adventure.”

And how else can one as a young man describe the start of a PhD thesis if not as an adventure. Indeed adventurous; personally, scientifically and socially. Challenging, often laborious and exhausting, but in the end almost four years of continuous inspiring days and incessant possibilities to grow above oneself.

Week after Week, I had the privileged opportunity to see and hear Prof. Dr. Peter Nick’s tireless enthusiasm for plant science, which was a constant and everlasting inspiration from the very first day of this PhD thesis and still is to the present day.
I would like to thank Peter Nick for offering me the possibility to be part of this. Especially for letting me find my own way.

I’d like to thank Prof. Dr. Tilman Lamparter to be the corrector of the thesis, as well as all the other members of the committee.

Wholeheartedly I would like to thank my sincere friend Dr. Vaidurya Sahi, whose passionate mind broadened my knowledge of the world; the world of science and even more the world we live in.

The greatest knowledge of botany I have ever met (and most likely ever will) in one single person, I would like to thank Dr. Max Seyfried, especially for providing plant material of different *Salvia* species and his encouragement for teaching. In the context of teaching encouragement I would also like to thank Dr. Beatrix Zaban.

I got the chance to meet a lot of inspiring colleagues from all parts of our world, above all my dear roommates of office 517, namely Daniela, Karwan, Xin and Pallavi. Thank you.

My special thanks go out to the hard working azubis and the gardeners who took so good care of my plants and experiments, namely Anne Kuppinger, Kerstin Huber, Lisa Weiler, Elena Schneider and Lara Potthoff.

For administrative and technical support, I would like to thank Renate Herberger-Biester, Joachim Krüger, Sabine Purpur, Ernst Heene, and Nadja Wunsch.

I’d like to thank Trent, Maynard and his crew and also Tom.

Furthermore, I’d like to thank Stephen Hillenburg (deceased 2018), who brought an indescribable joy into my life and into the lives of my closest friends. May he rest in peace, and, though in the shadow of his creation, remain unforgotten.

Speaking of friends, yes, every single conversation in pubs and elsewhere has lead me to the very point where I am now. Thank you.
And of course, my family. Words just wouldn’t catch it.

III. Abstract

The Plantae kingdom and its autotrophic representatives shape the surface of planet Earth and especially flowers contribute to a great amount to the astonishing beauty of nature. Naturally, flowers are not primarily intended to please humans, but they represent the sexual reproductive organs of the angiosperms, the flowering plants. The ingenious interaction of flowers with pollinators is an example for co-evolution and results in fine-scaled adaptations and complex floral shapes and structures.
The unique staminal lever mechanism that is limited to the species rich genus *Salvia* from the Lamiaceae family is an extraordinary example for plant-pollinator interaction. The species of genus *Salvia* possess zygomorphic flowers, which have an amazing variety of petals and stamens, resulting in melitto- and ornithophily and worldwide distribution.
In order to conduct a robust investigation on the evo-devo of *Salvia* flowers, the aim of the study was to take as holistic a view as possible on the genus *Salvia* by interweaving morphological, molecular and developmental approaches and put this against the background of speciation.
Flowers of 24 *Salvia* species from different parts of the world were investigated with regard to their floral geometry. Measurements of corolla and reproductive traits were analyzed by principal component analysis, which unveiled the flower size (represented by 79,21% of data in PCA) and bilabiate or tuberous floral shape (represented by 11,34% of data in PCA) as dominating floral features. In this context, the greater floral variety of New World *Salvia* species (concerning the dimensions size and shape) could be demonstrated, a result that is congruent with the elevated species number of genus *Salvia* on these continents. Finally, the statistical evaluation of floral traits appears to be a suitable tool to morphologically classify *Salvia* species.
For characteristic genes from the ABCDE model of flowering orthologues were identified for the genus *Salvia* and used for gene expression experiments that were conducted in two distant (concerning origin, floral habitus or pollinator) related *Salvia* species, *S. pratensis* and *S. elegans*, and here for closed flower buds and fully opened inflorescences, respectively. Key finding in this experiment is a general gene expression pattern that for the most part corresponds with the ABCDE model demonstrating the models conservation for the biggest genus of the Lamiaceae. Additionally, most of the evaluated flowering genes have been shown to be higher expressed in developing tissue, compared to opened flower tissue. B-class genes display huge elevation in expression in *Salvia pratensis* and *Salvia elegans* in corolla and stamen tissue, the organs that are of crucial importance for attracting pollinators and the way of reproductive isolation in *Salvia*.
The two B-class genes GLOBOSA and DEFICIENS have been used for an in-depth Bayesian based phylogenetic inference including more than 30 different species, which is the first examination of a floral trait-related marker for the genus *Salvia* and therefore directly linked with speciation in this genus. Different levels of gene duplication were unraveled, with a duplication event of the GLOBOSA gene that is limited to New World *Salvia* species, whereas the DEFICIENS gene displays a clear duplication that is evident for the whole genus *Salvia*. A model based on the multiplication of transcription factor interactions of flowering genes has been elaborated, hypothesizing that this increase of interactions synchronizes with the species radiation of genus *Salvia* in the New World and thus might one of the driving forces for this radiation. Significant posterior probability ($> 95\%$) values underline the results and also proves the suitability for phylogenetics using these genes in the genus *Salvia*, especially for the Eurasian *Salvia* species.
Plotting the amino acid sequences of B-class genes in the wider context of ABCDE genes including sequences from *Arabidopsis*, lead to a strongly supported ABCDE cluster, that shows a remarkable split into a B-class clade and an ACDE-class clade. This split was interpreted as an early divergence of division of responsibility into plant-based intrinsic factors (ACDE-class), like inward seed growth, and extrinsic factors (B-class), like outward interaction with pollinators, in *Salvia* flowers or maybe for flowers in general.
Taken together, the evo-devo approaches conducted in this thesis and applied to the species rich genus *Salvia* give an insight on the driving forces behind the speciation process.

Contents

"Do you know, I cannot understand how anyone can pass by a green tree, and not feel happy only to look at it!"

- Prince Lev Myshkin in Fyodor Dostoevsky's "The Idiot"

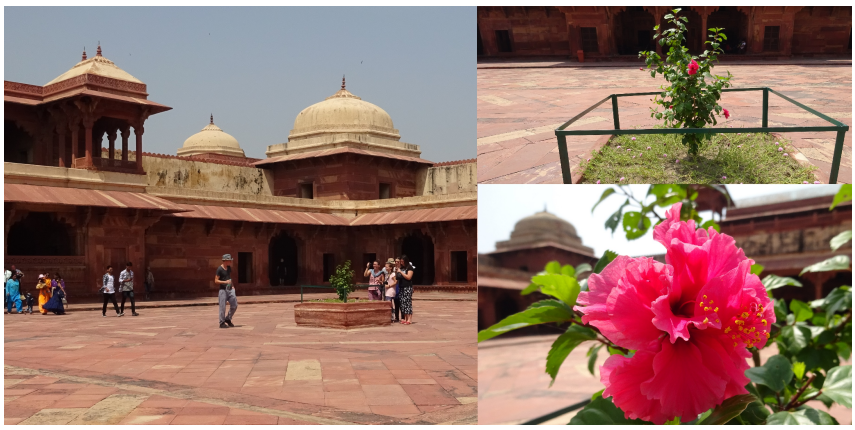
1 Introduction

1.1 Flowers – the aesthetic and functional crowning glory of a successful kingdom

Plants, as the first organisms to conquer solid land on planet Earth [1], had and obviously still have a vital impact on almost all other organisms, including the ones that currently (self-proclaimed) dominate that very same solid land.
The true age of this solid land remained unclear for a long time and was dominated by religious beliefs. The inviolability of the sanctitic creation has been axiomatic until the 18^{th} century, when doubts on the Earth's age slowly started to erode this belief. Those doubts were grounded on geological findings, which pointed to a much older Earth than presumed by the predominant zeitgeist [2].
Today it is assumed that the Earth was formed about 4.5 billion years ago [3]. After one billion years as a seething desert, LUCA, the Last universal common ancestor (a putative thermophilic, anaerobic, CO_2-fixing, H_2-dependent and N_2-fixing archetype) [4], was formed, marking the beginning of life on planet Earth about 3.5 billion years ago [1]. It took another three billion years until the colonization of the solid land in the Ordovician around 450 million years ago began. This colonization marks also the start of the success story of plants, that, with several setbacks though, has its flowering time with the development of flowering plants that originated at the end of the Jurassic or at the beginning of the Cretaceous period (see introduction chapter 1.1.2) [1][5][6] and nowadays are the dominant class of the Plantae kingdom.

1.1.1 Flowers and humans - beauty and science walking hand in hand

Often still recognized as the green somethings in the cracks of the asphalt, the human dependency on plants is undiminished, as they remain primary producers for this planet [1]. And as the species that categorizes nature for its own understanding, we, as a species, have been carried on the shoulders of plants as they initially promoted quantum leaps to the various plateaus of our cultural development (1) [7] and subsequently were used to gather an epistemic understanding of natures mechanisms (2).
(1) Fruits for early survival (nutrition for early Hominoids).
Seeds for germination of culture (the start of the neolithic revolution, when humans started to cultivate land and domesticated plants, the circle closes and humans became (plant-like?) sessile to some extant).
Flowers for perception of astonishing aesthetics of nature.
(2) Furthermore, flowers have been objects of scientific experiments leading to the description of fundamental laws of biology, like genetic inheritance [8], the latency problem [9] and plant sexuality facilitated by insects [10], done by Mendel, Kölreuter and Sprengel, respectively. Thus, flowers have helped to increase the understanding of biology to an overwhelming extant, and have been used for before mentioned experiments long before they themselves became part of the understanding process.



(a) Hibiscus (*Hibiscus spec.*) - as a prominent feature of moghal gardens in an exposed place (Fatepur Sikri, India)

(b) Plant-Pollination interaction depiction on marble (Amber Fort, India)

Figure 1: Example of the worship of flowers by man.

Flowers are short shoots with limited growth of the apical meristem, functioning as reproductive organs in angiosperms (flowering plants, Magnoliophyta) [1][11][12]. Flowers show different levels of symmetry, with radial symmetry (or actinomorph, many sections planes yield mirror-inverted halves, e.g. the Asteraceae family) and dorsoventral symmetry (or zygomorph, only one sectional plane yield mirror-inverted halves, e.g. the Lamiaceae) as the two most dominant patterns. There are asymmetric flowers as well, however, they are rather rare [11].
The floral organs are transformed leaves, which has been understood by Johann Wolfgang von Goethe, who was not only an extraordinarily gifted writer but also a talented natural scientist. In his "Metamorphose der Pflanzen" Goethe described floral organs (calyx, corolla, stamen and carpels) as modifications ("Metamorphosen") of leaves that are arranged in whorls. Hereby the most outer whorl is still most likely to be recognized as leaf. Sepal leaves (from Latin *sepalum*: "covering") can be fused and form the calyx, that often keeps the green color and protects the flower bud before opening. The second whorl is noticeably colored and serves for pollinator attraction. The petal leaves (Latin: *petalum* for "to spread out") taken together are known as the corolla and can be fused as well. The variety of corolla composition is manifold within the angiosperms and is the main reason for the colour variety of nature. The two outer whorls are also referred to as perianth (meaning "around the flower") and represent the sterile part of the flower, but functionally clearly belong to the reproductive parts of the flower. The third whorl represents the male part of the flower and its pollen producing stamen are collectively called androecium (Greek: "the house of the man"). Every single stamen is build up by a filament that carries the anther, which is composed of two theca that are connected by the sterile connective. The theca contain two pollen sacs respectively, which enclose the pollen grains, that are usually released by a longitudinal crack of the theca. The forth whorl is composed of carpels, that are the female part of the flower and as a whole are called gynoecium (Greek: "the house of the woman"). The single carpels can be fused and consist of a stigma, a style and an ovary. The often elongated style connects the ovary and the stigma. The latter has an adhesive surface where dispersed pollen can attach. Ovaries contain the ovules that will be part of the fruit in later stage of development [1][11][12][13].

But how did flowering plants actually originate, and how did they become the most important group of plants and could shape the surface of this planet?

1.1.2 The dawn of angiosperms - evolutionary history of flowering plants

The most recent common ancestor of the angiosperms (flowering plants sensu stricto) existed 140 to 250 million years ago [5][6]. At that time, dinosaurs were still roaming the surface of Pangaea, the paleo-continent that began to fall apart around 175 million years ago [14]. Generally, angiosperms are described as plants that have ovules enclosed within ovaries, with subsequent seed production and fruit development [1][11]. The dominant feature of the angiosperms, the flower (as described above), attracts organisms from completely different realms of the tree of life, including insects, birds or bats. Insects (including bees) existed for hundreds of million years before the first flower arose, however, a rapid co-evolution of flowering plants and insects (with artful plant-pollinator interactions) has most likely contributed to the worldwide dispersal of the angiosperms [15]. But the temporal first appearance and the habitus of the first flowering plants and the actual first flower is part of an ongoing debate that is addressed by different approaches. In 1998, a Jurassic (estimated age of 125 million years – early Cretaceous) angiosperm fossil from China with above mentioned characteristics has been described and titled as *Archaefructus* [16](see figure 2). This name is a reference on *Archaeopteryx* (a fossil which shows a transition form of dinosaurs and birds and ultimately proved Charles Darwins theory of the origin of species right), which is often called the "Mona Lisa of evolutionary science"; as botanist we could refer to the *Archaefructus* fossil as the "sunflower (referring to van Gogh) of the dawning angiosperms".
A pre-Cretaceous origin of the angiosperms has been proposed by Fu *et al.* in 2018 based on a series of plant fossils, in which the existence of structures that resemble to ovaries enclosed ovules beneath a putative perianth has been ascribed. According to this finding the dawn of the angiosperms would be 50 millions earlier then previously presumed [17]. However, the interpretation of Fu *et al.* was criticized by other groups that in summary pointed out that those structures could be interpreted as conifer cones as well [18][19].
Apart from the fossil findings there are computer-based approaches for estimation of age and especially the habitus of the first angiosperm flower based on recent floral structure traits [20](see figure 2), an approach, which has been under critic as well, meaning the physical and temporal appearance of the first angiosperm is still a heated debate.

It is undoubted though, that the flowering plants are a success story of evolution, in view of their worldwide dispersal, which took place in an evolutionary eye blink.

The perhaps most well-known quote concerning this topic from one the greatest biologist, Charles Darwin himself, is not to be missed here:

"The rapid development as far as we can judge of all the higher plants within recent geological times is an abominable mystery."

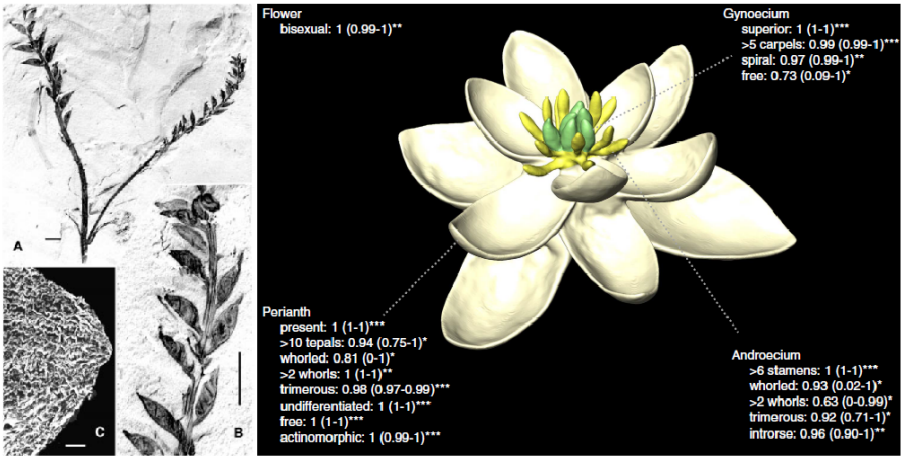


Figure 2: Fossil finding of *Archaefructus* with the carpels clearly visible in picture B (from [16]) and the computer based model of the putative first angiosperm flower (from [20]).

1.2 Lamiaceae - a zygomorphic species rich plant family

The Lamiaceae family belongs to the most diverse taxa in the angiosperms and comprises 236 genera [21][22]. This number is widely accepted, however, the arrangement of these genera into subfamilies is under continuous revision. Recent family-wide chloroplast-based molecular phylogeny of the Lamiaceae revealed twelve strongly supported clades which can be seen as the phylogenetic backbone of the whole family [23]. Ten of these clade were assigned to subfamilies (e.g. Nepetoideae or Lamioideae), the other two clades correspond to the genera *Callicarpa* and *Tectona* [23]. One year later, the revision of the two genera to form own subfamilies was proposed [24].

Although the statement, that the taxonomic confusion of the Lamiaceae is "a disgrace to botany", which is attributed to the British taxonomist David Lindley, was made almost 200 years ago, it appears to be still relevant today notwithstanding the growing knowledge of genes and genomes.

Commonalities of this family are the tetragonal stem and the bilaterally symmetrical, strongly zygomorphic flowers that are build up by five fused sepals and petals that form a characteristic lip shape [11]. The older name Labiatea (which is still used sometimes today) refers to this shape and derives from the Latin term *labia* (= lip). Plants of the Lamiaceae family grow as shrubs or trees and are widely known and cultivated for their aromatic compounds or ornamental features, e.g. essential oils of mint (*Mentha*) and basil (*Ocimum*) [1] or colorful leaves (coleus – *Solenostemon*) or flowers (scarlet sage – *Salvia splendens* Sellow ex Roem. & Schult). Pollination in Lamiaceae is an extraordinary example for plant-pollinator interaction and co-evolution, whereby both, entomophily (here especially melittophily) and ornithophily, is observed [1]. Co-evolution with pollinators on the one hand and the evolutionary history of numerous whole genome duplications [25] might have been the driving forces for the extensive diversification of this family. Within the Lamiaceae, the Nepetoideae is the largest subfamily with around 3500 species, which makes up half of the estimated 7200 species of the whole family [21][22]. The subfamily of Nepetoideae is thought to be originated 52,3 to 57,6 million years ago around the transition from Paleocene to Eocene [26] (based on the 95% confidence intervals of plastidic analysis [26] and fossil records of the Lamiaceae, e.g. fruits of *Melissa spec.*, which were described 1926 [27] and reviewed 2010 [28]).

In the years during this PhD, the availability of whole genome sequence information has grown significantly, which is evident for the Lamiaceae as well. For example, groups encrypted the plastidic genome sequence of *Lavandula angustifolia* Mill. [29] or the whole genome of model plant for flowering *Antirrhinum majus* [30]. In 2018, the Mint Evolutionary Genomics Consortium sequenced and provided genomic and transcriptomic information of 48 Lamiaceae species, including four species of *Salvia* sensu lato [31].

1.3 *Salvia* - an ungraspable genus?

Salvia is a species-rich genus from the Lamiaceae family and many of its members are well-known for several properties like common sage (*Salvia officinalis* L.) usage as tea, Chia-seeds (*Salvia hispanica* L.) as so called Superfood, or other species as ornamental plants (e.g. *Salvia splendens*, *Salvia nemorosa* L. or *Salvia farinacea* Benth.). The amount of cosmopolitan occurring species of this genetically paraphyletic genus has been numbered with around 1000 described species, with the genera *Rosmarinus*, *Perovskia*, *Zhumeria*, *Meriandra* and *Dorystaechas* (in total 15 species) disrupting the monophyly of *Salvia* [32]. The paraphyletic *Salvia* has been considered a monophyletic genus, because two of the originally four stamen that are typical for the tribe Mentheae were functionally lost [33] and the modification of the stamens into lever-like structures can be seen as a "key innovation" for this genus [34].

1.3.1 Dispute on the classification of *Salvia*

However, lately there are opposing aspirations to either "salvinise" the disturbing genera to create a large monophyletic *Salvia* genus [35]; or to split up the former *Salvia* genus into six smaller genera, with the eponymic species *Salvia officinalis* as part of *Salvia* sensu stricto [36] (see figure 3).

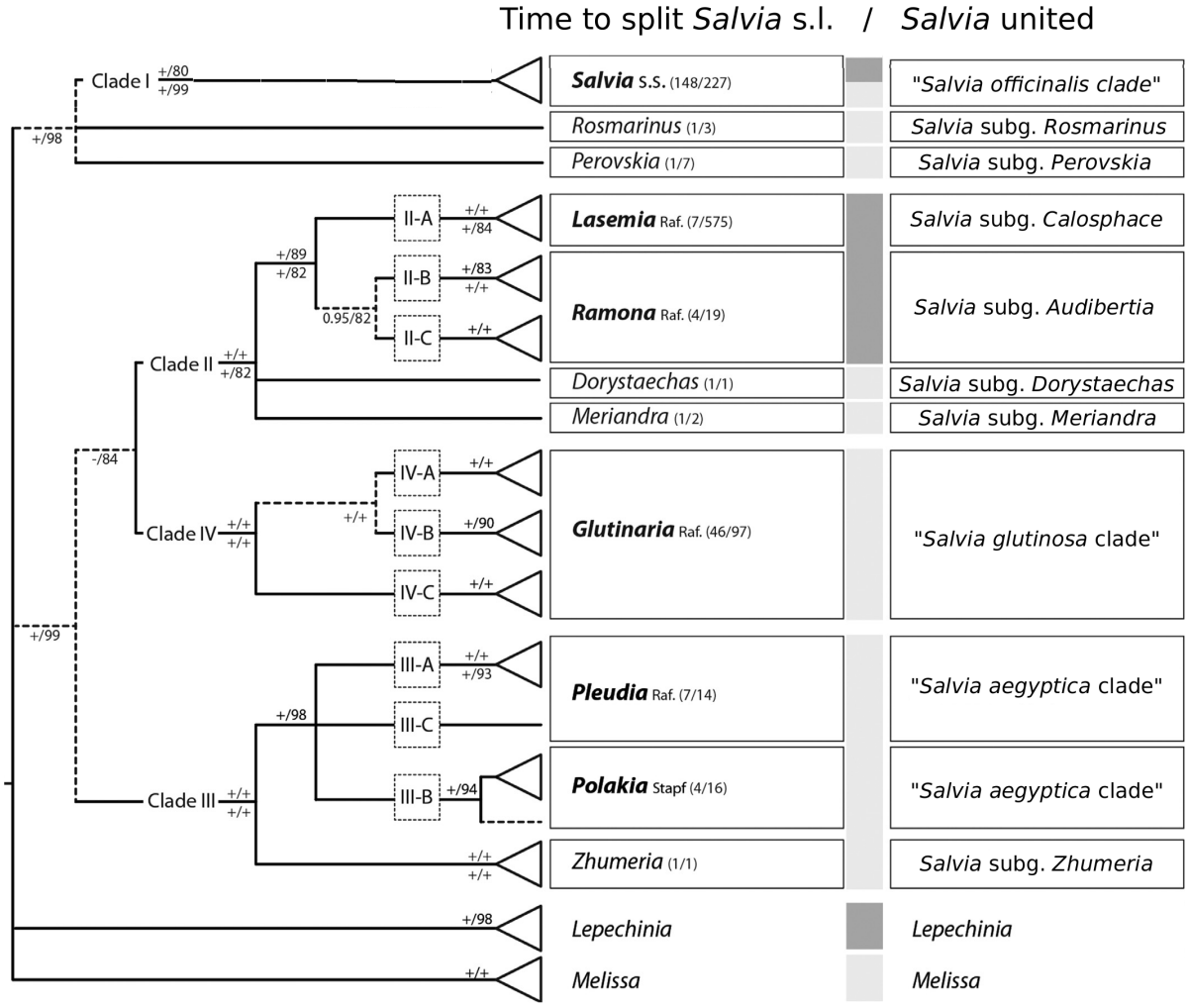


Figure 3: Proposed nomenclatures for the genus *Salvia* sensu lato from Claßen-Bockhoff and Will 2017 [36] and Drew *et al.* 2017 [35]. The phylogenetic tree is modified from Claßen-Bockhoff and Will 2017 [36] and nomenclature from Drew *et al.* 2017 [35] has been added on the right side (labeled with *Salvia* united). Dark grey bars = New World species, light grey bars = Old World species.

Arguments for integrating *Rosmarinus*, *Perovskia*, *Zhumeria*, *Meriandra* and *Dorystaechas* into *Salvia* with a subsequent renaming of the concerning 15 species with the genus name *Salvia* are listed in Drew *et al.* 2017 and include morphological arguments (e.g. *Salvia* as the only group in the Mentheae possessing two anterior fertile stamen), practical arguments (e.g. renaming 15 species is easier compared to renaming of three quarter of *Salvia* and a subsequent relabeling of countless herbariums) and phylogenetic arguments [35].
Opposing to this approach, the splitting of *Salvia* sensu lato into the six genera *Salvia* sensu stricto (including *Salvia officinalis* and around 250 European *Salvia* species), *Lasemia* (including the ornamental blue flowered *Salvia patens* and other New World species), *Ramona* (including *Salvia columbariae*), *Glutinaria* (including *Salvia miltiorrhiza* and other Asian *Salvia* species), *Pleudia* and *Polakia* has been suggested by Will and Claßen-Bockhoff 2017, on the grounds that the current nomenclature can no longer be maintained with respect to the morphological diversities in *Salvia* sensu lato [36]. Up to this point there is no final consensus on this taxonomical dispute and since Drew *et al.* suggested updated circumscriptions for the species of the five genera nested within *Salvia* (e.g. *Rosmarinus officinalis* to *Salvia rosmarinus*) there might be confusion about the nomenclature, since some groups started to use these updated species names. For this thesis the accepted names from the plant list [37] were used, however, in table 1 the possible synonyms that resulted from this dispute are listed.

1.3.2 *Salvia* floral morphology and pollination

The basic modules of the *Salvia* stamen are the filament that is laterally attached to the corolla, the strongly extended connective which separates the theca of each stamen and the theca itself [38]. The length and the angle of the filament varies considerably in different species and also the differences in the elongated anterior upper arm of the connective which ends with the fertile theca, that is build of two pollen sacs, has a major impact on the staminal structure [38]. Overall there is a huge intrageneric variance of the stamen building blocks, however, the general bauplan is restricted to the *Salvia* genus, which means a clear morphological border delineates *Salvia* from other genera, while the delineated bauplan brings enough flexibility to extensively vary within the genus, leading to the success that is projected by huge species number and worldwide distribution.

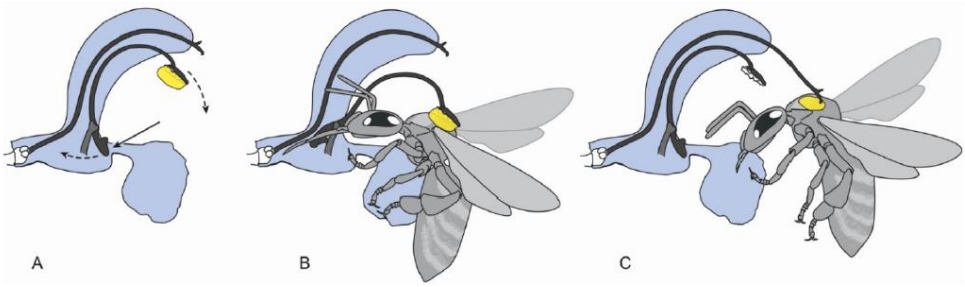
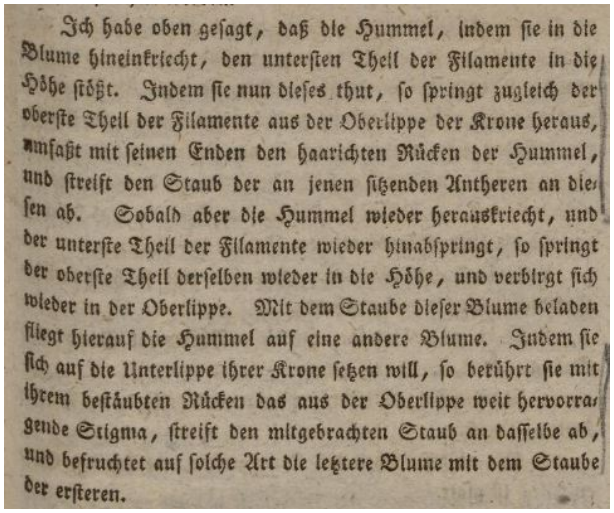


Figure 4: *Salvia pratensis* schematic pollination (from [38])

Figure 5: *Salvia pratensis* pollination of an indigenous individual, displaying step A and B from figure 4.

The stamen structure and the functional staminal lever mechanism with the precise pollen deposition on the pollinators body is most likely the reason for the global dispersal of the genus *Salvia* s.l. [34]. This seemingly simple mechanism (see figures 4 and 5) has been described first by Christian Konrad Sprengel in the 18^{th} century, who also proposed that insects are the assistants of plant sexuality in his groundbreaking (delayed however, during his lifetime rejected) work "Das entdeckte Geheimnis der Natur im Bau und in der Befruchtung der Blumen" [10]. On page 61 of this work Sprengel meticulously describes the process of pollen transfer in *Salvia pratensis*, which is definitely worth quoting as a whole (see figures 4 and 5 for the process in nature and figure 6 for Sprengels description).



Ich habe oben gesagt, daß die Hummel, indem sie in die Blume hineinkriecht, den untersten Theil der Filamente in die Höhe stößt. Indem sie nun dieses thut, so springt zugleich der oberste Theil der Filamente aus der Oberlippe der Krone heraus, umfaßt mit seinen Enden den haarichten Rücken der Hummel, und streift den Staub der an jenen sitzenden Antheren an diesen ab. Sobald aber die Hummel wieder herauskriecht, und der unterste Theil der Filamente wieder hinabspringt, so springt der oberste Theil derselben wieder in die Höhe, und verbirgt sich wieder in der Oberlippe. Mit dem Staube dieser Blume beladen fliegt hierauf die Hummel auf eine andere Blume. Indem sie sich auf die Unterlippe ihrer Krone setzen will, so berührt sie mit ihrem bestäubten Rücken das aus der Oberlippe weit hervorragende Stigma, streift den mitgebrachten Staub an dasselbe ab, und befruchtet auf solche Art die letztere Blume mit dem Staube der ersteren.

Figure 6: Ich habe oben gesagt, daß die Hummel, indem sie in die Blume hineinkriecht, den untersten Theil der Filamente in die Höhe stößt. Indem sie nun dieses thut, so springt zugleich der oberste Theil der Filamente aus der Oberlippe der Krone heraus, umfasst mit seinen Enden den haarichten Rücken der Hummel, und streift den Staub der an jenen sitzenden Antheren an diesen ab. Sobald aber die Hummel wieder herauskriecht, so springt der oberste Theil derselben wieder in die Höhe, und verbirgt sich wieder in der Oberlippe. Mit dem Staube dieser Blume beladen fliegt hierauf die Hummel auf eine andere Blume. Indem sie sich auf die Unterlippe ihrer Krone setzen will, so berührt sie mit ihrem bestäubten Rücken das aus der Oberlippe weit hervorragende Stigma, streift den mitgebrachten Staub an dasselbe ab, und befruchtet auf solche Art die letzter Blume mit dem Staube der ersteren.

-Translation of figure 6.
"I said above that the bumblebee, by crawling into the flower, pushes up the lowest part of the filaments. As it does so, the uppermost part of the filaments leaps out of the upper lip of the crown, embraces the hairy back of the bumblebee with its ends, and wipes off the pollen of the anthers sitting on them. But as soon as the bumblebee crawls out again, the uppermost part jumps up again, and hides itself again in the upper lip. Loaded with the pollen of this flower, the bumblebee flies onto another flower. When it wants to sit on the lower lip of its crown, it touches with its pollinated back the stigma protruding from the upper lip, wipes off the pollen it has brought with it, and in this way fertilizes the last flower with the pollen of the former."

The optimization of plant-pollinator interaction that has (regarding the Lamiaceae family) its peak efficiency in *Salvia* can be ascribed to the staminal lever mechanism [39]. Distinct realms of nature (namely plants and insects) have undergone a long history of co-evolution resulting in an symbiotic-like manner, whereby the short interaction of pollinator and flowers results in a nectar reward for the animal and an almost guaranteed reproduction success for the plant. A reward-for-dispersal system with symbiotic characteristics. In *Salvia* the interaction of pollinators is so precise that some bee-pollinated Old World species can even ignore geographical, phenological and pollinator specific barriers by using different parts of the same pollinators body to build up their reproductive isolation barriers [34].

One quarter of *Salvia* genus is ornithophilous, however, this way of pollination is restricted to the New World, with only three exceptions of bird-pollinated *Salvia* species in South Africa [40].

The morphological consensus (based on floral structure and pollination mechanism) conflicted with first molecular studies [32] that shows the paraphyletic nature of *Salvia*, which, though surprising in the beginning, has been proved several times and is commonly accepted (examples [33][36][35][40]).

It is clear what flowers actually *are*, how they are build up, how they became a model of success of evolution as well as how they functionalize their mode of action in the genus *Salvia*.
But how does a single flower or the inflorescence of a plant actually develop?

1.4 The role of plant hormones in the flowering process

Plant hormones orchestrate all developmental processes throughout the complete lifetime of an individual plant.

Auxins are key regulators in various developmental processes of the plant, such as cell elongation, cell division, apical dominance and many others [1][41][42][43]. The main natural auxin in plants, indoleacetic acid (IAA), is biosynthesized from L-tryptophan and controls major aspects of flower development [1]. During the flowering process, the unfolding of a single flower and thus the stretching of all its flower organs is of crucial importance. This is particularly evident for petals, stamens and style development, where cell elongation is promoted by auxins and gibberellic acid, another plant development hormone. Likewise, the elevated cell division in the floral organs is orchestrated by auxins and cytokinines. After the flower has opened and unveiled its reproductive organs, most likely it does not take a long time for a pollination event to take place. The ripening and the tearing of the anthers by contact to a pollinator to release the pollen grains is mediated by jasmonic acid, which is itself induced by auxin. After successful pollination the floral senescence process sets in (mainly of corolla and stamen), which is under the control of the auxin-induced plant hormone ethylene [44].

1.4.1 Auxin Response Factors in model plants

As summarized in the previous lines, auxin is crucial for flower development, however, the plant hormone is dependent on transcription factors to shift its mode of action to the genetic level. The transcriptional response of auxin is mediated by the Auxin Response Factor (ARF) family [45]. In a low auxin level state ARFs are bind by Aux/IAA proteins that repress the ARF activity and therefore inhibit the expression of auxin-responsive genes [46]. When the level of auxin rises, the Aux/IAA protein becomes ubiquinated and subsequently degraded by the 26S proteasome (see figure 7).

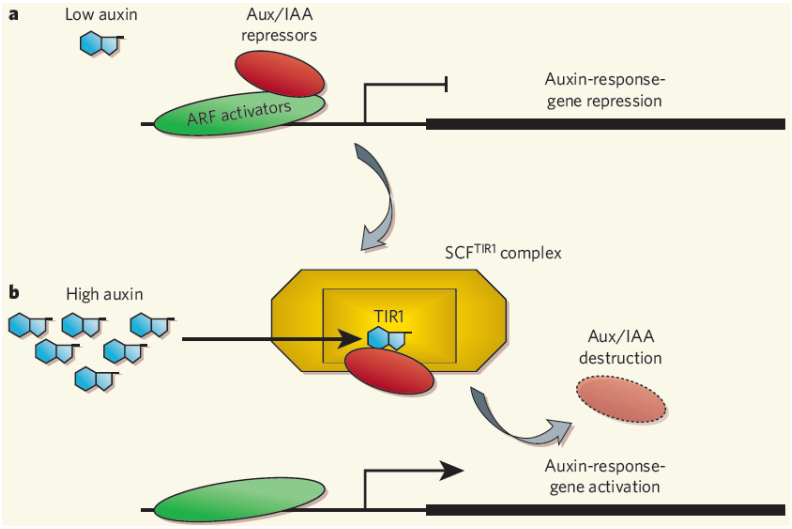


Figure 7: The auxin signalling pathway. From Guilfoyle (2007) [47]

The released ARFs starts to regulate its target auxin response gene [45][48], making it possible to indirectly deduce the presence of auxin by ARF activity. Several ARFs in the context of flowering have been demonstrated, especially in the model plant *Arabidopsis thaliana* L. Unfortunately the nomenclature of orthologue ARFs is not congruent in different plant species (e.g. *Arabidopsis* ARF1 is not *Solanum* ARF1), making the usage of the plant species initials necessary. The atARF8 has been shown to promote petal growth [49], additional to fertilization and fruit development [50] in *Arabidopsis thaliana*. Together with atARF6, the atARF8 transcription factor acts redundantly during flower maturation mediated by jasmonic acid [51][52]. Both genes are under control of microRNA167, and elevation of mi167 and subsequent down regulation of both gene leads to sterility, which has bee shown in *Arabidopsis* and tomato, highlighting the importance and the conservation (in rosids and asterids) of this mechanism [53][54]. Concerning the male reproductive part of the flower, the atARF16 has been shown to be active during stamen elongation in *Arabidopsis* [55] and just as atARF6 and atARF8, this transcription factor is targeted and most likely controlled by a microRNA, mi160 [56].

1.4.2 Auxin Response Factors in the context of flowering in *Salvia*

The first whole genome sequencing for the genus *Salvia* has been carried out 2016 (for *Salvia miltiorrhiza* Bunge) [57], and first investigations with the available data targeted the unraveling of the Auxin Response Factors in this East Asian *Salvia* species. The first transition to *Salvia* revealed a number of 25 different genes, controlling the genetic response of auxin in *Salvia miltiorrhiza*. The authors predicted the smARF1, smARF4, smARF6 and smARF25 genes to participate in flower development, with the strongest upregulation in floral tissue for smARF1 and smARF25. The homologues in the model organism *Arabidopsis thaliana* and the phylogenetic relation to *Salvia miltiorrhiza* genes were phylogenetically plotted. The smARF1 is homologue to atARF16, whereas smARF25 is a homologue to the atARF6, this fits to the spacial (flowers) and functional (floral development) mechanisms described earlier. The miRNA160 binding site in the *Arabidopsis* atARF16 sequence has been demonstrated in the smARF1 sequence of *Salvia miltiorrhiza* as well, and shows a match of 20 of 21 nucleotides compared to the distant related rosid *Arabidopsis thaliana*, proving the conserved mechanism in distant related angiosperms [57].

The points listed here demonstrate the special interest of smARF1 and smARF25 in context of flowering.

Auxin and the Auxin Response Factors play an important role in floral development, however, the genetic background of flowering, and its key regulators have been investigated intensively the last 30 years, leading to the well accepted ABC-model.

1.5 Flower development - genetic foundations of flowering (ABCDE-model)

1.5.1 General aspects and a short history of the flowering model

The floral organs are ordered in whorls and the today well-known ABC model of flowering has its roots in the analysis of mutants that displayed replacements of floral organs in different whorls [58][59]. These mutant-based changes in floral whorl architecture (called homoesis) has been observed in many species and fascinated biologist for a long time (reviewed [60]). In 1991, the ABC-model (see figure 8a) was formulated for the first time based on mutants in the model plants *Arabidopsis thaliana* (a rosid plant) and *Antirrhinum majus* (an asterid plant) [58][59] with the assignment of five different genes as homeoticly functional, namely APETALA1 (AP1) / SQUAMOSA (SQUA) (A-class gene), APETALA2 (AP2) / LIPLESS (LIP) (A-class gene), APETALA3 (AP3) / DEFICIENS (DEF) (B-class gene), PISTILLATA (PI) / GLOBOSA (GLO) (B-class gene) and AGAMOUS (AGA) / PLENA (PLE) (C-class gene), for *A. thaliana* / *A. majus*, respectively.

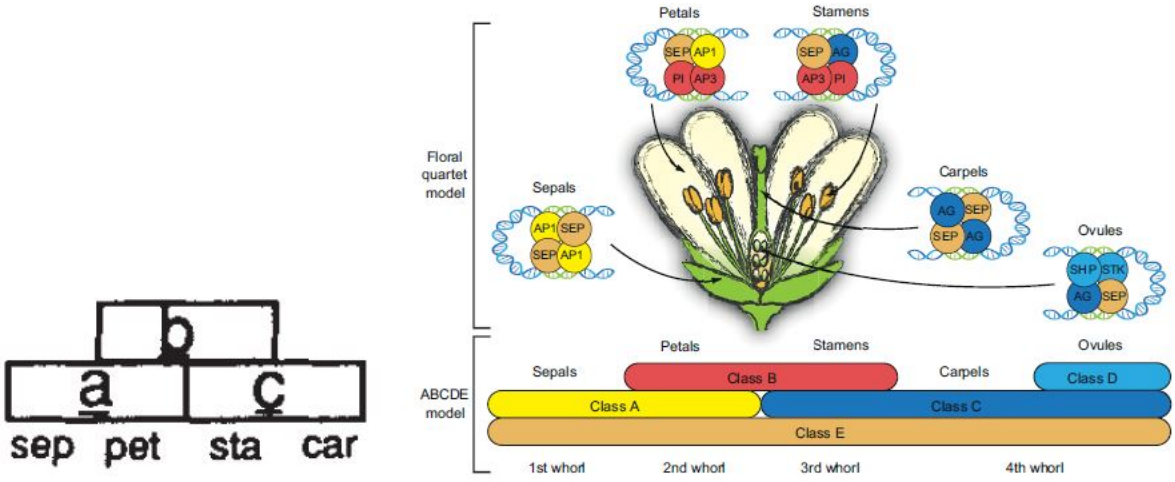


Figure 8: Evolution of the first ABC model to the current combined ABCDE / FQM model. From [58] and [61].

In the almost 30 years till today this model has been widely accepted, experimentally proved multiple times and refined based on accumulating data (for detailed review, see [61]). It has been shown that the ABC-genes are not sufficient to specify the identity of the floral organ alone, hinting to other homeotic genes [62].

A major refinement of the flowering model is the extension to the ABCDE model [63] (see figure 8b). A group of genes that were labeled as D-class genes (that specify the ovule identity) have been identified in *Petunia* [64], while the four SEPALLATA genes (E-class), whose organ-overspanning function were demonstrated in transgenic studies in *Arabidopis* [65], were taken together and subsequently summarized in the ABCDE model [63].

1.5.2 The working hypothesis for the ABCDE genes: the floral quartet model

In 2001, Theißen and Saedler suggested for the first time the working hypothesis for the transcription factors, which is based on the formation of floral quartet-like complexes (FQC) of four homeotic proteins that derive from the above presented genes (see figure 8b) [63]. The proteins encoded by the ABCDE flower developmental genes are composed of four different characteristic domains: the MADS domain (M), the intervening domain (I), the keratin-like domain (K) and the C-terminal domain (C), and therefore called MIKC-type MADS-domain proteins, that derive from MADS-box genes [61][66]. All genes shown in the further course of the text encode for MIKC-type MADS-domain proteins, with the exception of APETALA2 (see below).
The MADS domain is build up by around 60 amino acids, highly conserved and functions for DNA-binding of the transcription factors, as well as for the formation of dimers. The downstream intervening domain has dimerization functions as well, but compared to the MADS domain, this domain is composed of around 30 amino acids and displays higher variations of length in different plant species and consequently shows a relative weak conservation. The K domain comprises around 70 amino acids and functions for the formation of tetramers (see figure 9). The most variable domain, concerning length and conservation, of the MIKC-type MADS-domain transcription factors is the C domain with an approximate length of 40 amino acids. Though its function is mainly unknown, it most likely functions in transcriptional activation [61][66][67].

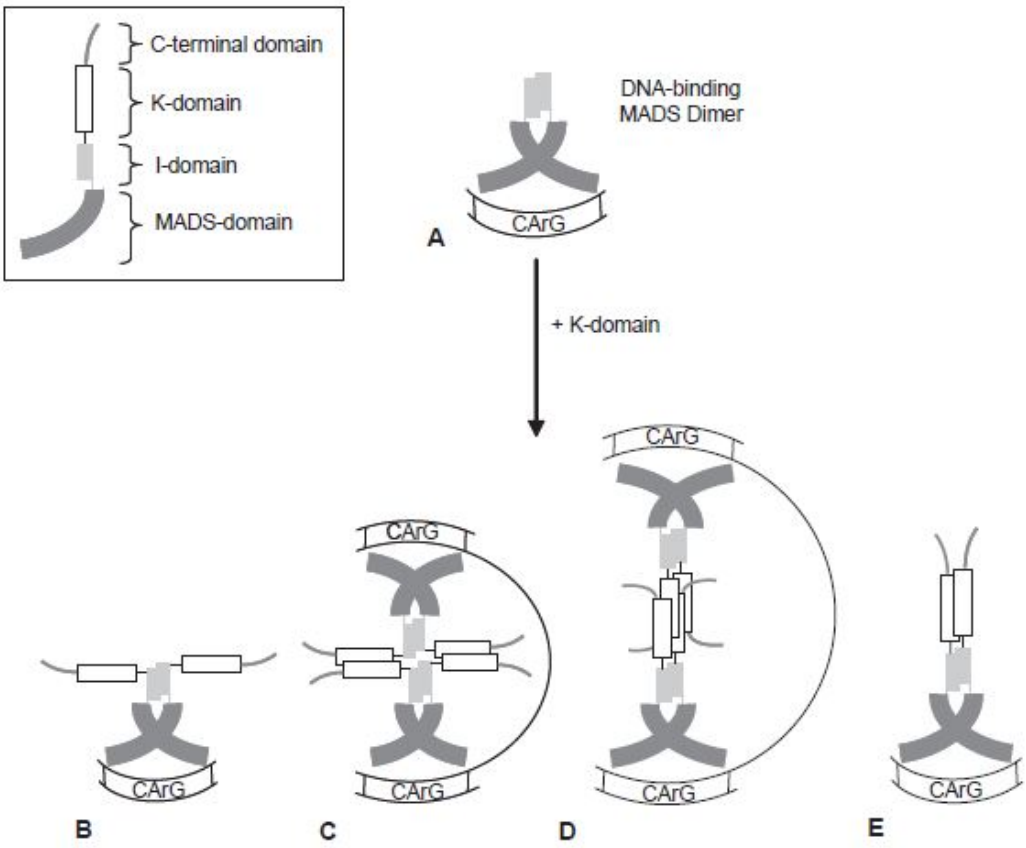


Figure 9: Model for tetramerization of MIKC-type MADS-domain transcription factors. From [67].

1.5.3 A-class genes - to start the alphabet and the development of a flower

The homeotic **APETALA1 (AP1)** gene encodes a MADS-MIKC transcription factor that belongs to the class A in the ABC-model of flowering [68]. In *Arabidopsis* it has been demonstrated to be expressed in the floral meristem and during later stages of development it determines the perianth of the flower [69]. The orthologue of A-class APETALA1 in *Antirrhinum* is SQUAMOSA (SQUA)(sequence identity of amino acids 68%), that likewise encodes for a MADS-box transcription factor [70]. AP1 is initially expressed in all whorls, but as development progresses it is restricted to the first two whorls by repressive activity of the C-class gene AGAMOUS [71]. Additionally, the proteins of the two B-class genes (APETALA3 and PISTILLATA, in *Arabidopsis*) bind to the AP1 promotor to repress its transcription [72]. Conversely, the AP1 gene has activation [73] and repression [74] potential for the expression of the B-class genes in *Arabidopsis* and does not show involvement in the repression of the C-class gene AGAMOUS (in *Arabidopsis*) activity [75]. Interaction with the E-class gene SEPALLATA3 has been demonstrated in *Arabidopsis* [74] pointing to the mode of operation of this A-class gene in building homodimers, that tetramerize with SEPALLATA homodimers to functionalize the development of the sepals [63] [61].

The **APETALA2 (AP2)** gene has been isolated in *Arabidopsis* and is the only gene in the ABC-model that does not possess a MADS-box domain in its amino acid sequence. Instead, two AP2 domains have been demonstrated to be highly conserved in different plant taxa [76]. Nevertheless, the restricted expression to the whorls 1 and 2 (the perianth) of APETALA2 meets the criteria for an A-class gene. APETALA2 is expressed in all parts of the flower, however, the translational repression in the two inner whorls by the microRNA172 has been shown, limiting the genes proteins to the perianth [77]. Further, in contrast to the A-class gene APETALA1, the APETALA2 gene represses the C-class gene in the two outer whorls, leading to the correct development of the sepal and the petal [77][62]. Together with the developmental gene SUPERMAN, APETALA2 activates the B-class gene PISTILLATA in *Arabidopsis* [78]. Two orthologues of APETALA2 in *Antirrhinum*, LIPLESS1 and LIPLESS2, have been isolated and the encoded proteins both display 65% similarity to APETALA2 [79].

1.5.4 B-class genes and a small history of gene duplications

DEFICIENS (DEF) as one of the first isolated transcription factors in the context of flowering was described in the model organism *Antirrhinum majus* [80]. Because of its early discovery, DEFICIENS lends its initial letter to the acronym for the MA**D**S-box genes. The homologue of the B-class gene DEFICIENS in *Arabidopsis* is APETALA3 (AP3) and likewise those genes are expressed in petals and stamens [81].

GLOBOSA (GLO), the other B-class gene, has been isolated by the same group that identified the DEFICIENS gene, and the close interaction with DEF has been proposed, based on similar phenotypes of mutants, suggesting overlapping functions of the genes [82]. The orthologue of this homeotic gene is called PISTILLATA in *Arabidopsis* and has been cloned in 1994 [78] and both genes have been demonstrated to facilitate the petal and stamen tissue identity from early developmental stages to floral maturity [82][83][78].

DEF and GLO, as well as AP3 and PI, regulate their own and the other genes expression, respectively [82][78]; their initial expression, however, is induced by the LEAFY gene (shown in *Arabidopsis* [75]). Additionally, the B-class genes are repressed in vegetative tissue by DELLA proteins. The inactivation of DELLA proteins which subsequently initiates flower development by expression of the AP3 and PI is mediated by gibberellic acid [84].
For the MIKC-type MADS proteins GLO (PI) and DEF (AP3) are an exception, because they exclusively dimerize with each other and are for most plant groups only functional as heterodimers [85][86].

This heterodimer was already present in the MRCA of the angiosperms and stayed functional highly conserved during plant evolution. However, the presence of GLO and DEF homodimers has been demonstrated, with an early loss of DEFICIENS homodimerization, whereas GLOBOSA homodimerization was lost later. DEF-DEF, GLO-GLO homodimers and DEF-GLO heterodimers were shown in *Gnetum gnemon* L., *Picea abies* L. (gymnosperms), *Amborella trichopoda* Baill. (a single species that is sister to all other flowering plants) and Magnoliids (basal angiosperms, e.g. *Liriodendron tulipifera* L.). This pattern can be seen in some higher eudicots as well, like *Medicago truncatula* Gaertn. for GLO-GLO homodimer and *Camellia japonica* L. for DEF-DEF homodimers, however, overall GLO or DEF homodimers are very rare in the core angiosperms and the clear predominant pattern is the GLO-DEF heterodimerization [86]. Additional to the broader appearance of homodimers of GLO and DEF, the gene expression of these

B-class genes is not limited to petals and stamen in basal angiosperms (e.g. Magnoliids) [87].
The contemplation of the genetic history of the B-class genes points to an early duplication event in a common ancestor of all extant angiosperms that resulted in the origination of the B-class genes DEF and GLO [88]. The history of genes and genomes is, especially in angiosperms, a history of gene or genome duplications [89]. With surprising congruence the duplication events of the B-class genes go hand in hand with the split of distinct groups of the spermatophytes [90].
From a genetic common ancestor the first duplication event resulted in the separation of B-class sensu stricto (the ancestor of AP3 and PI) and a B-sister clade during the diversification of the gymnosperms. A subsequent duplication event that yielded the APETALA3 and PISTILLATA genes from an initial B-class sensu stricto gene matches with the dawn of the angiosperms.

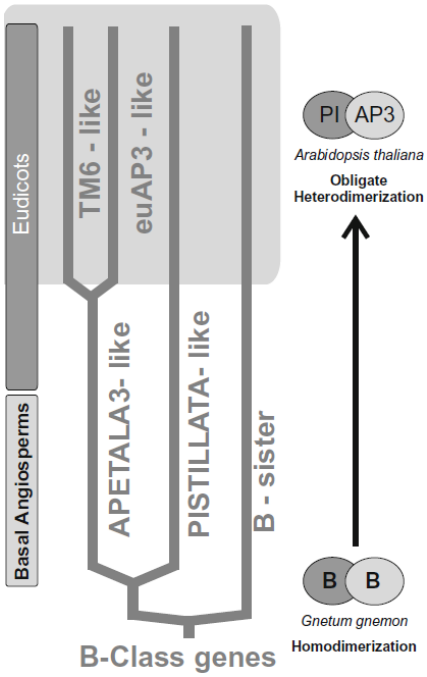


Figure 10: Model of B-class duplication during angiosperm evolution. Modified from Gioppato and Dornelas (2019) [90].
The split of ancient B-class into a B-sister lineage and an ancestor of arising APETALA3-like and GLOBOSA-like lineages in the basal Angiosperms as well as the duplication of the AP3-like lineage with (or as a trigger(?)) of Eudicot evolution are displayed. The switch from B-class homodimerization (evident in *Gnetum gnemon*) to obligate B-class heterodimerization in core eudicots (evident in *Arabidopsis thaliana*) is indicated.

The origin of the core eudicots overlaps with an APETALA3 duplication (leading to euAP3 and TM6 clades) [91], leading to at least three B-class genes in core eudicots, but often even at least three DEFICIENS-like genes [92]. It is therefore not surprising that DEF and GLO MIKC-type MADS genes display the fastest rate of evolvement compared to other MADS-box domain [93].

The general acceptance of duplication events in angiosperms together with evident rapid evolutionary rate of B-class genes, may either strengthen the present duplication models or additionally may lead to a view of additional duplication events, that just have been veiled by limited taxon sampling [91]. Contradicting B-class gene phylogenies and organismal phylogenies of known monophyletic groups point to this direction [91].

1.5.5 C-class genes - key player in the inner whorls

The C-class gene **AGAMOUS** was one of the first isolated MIKC-type MADS transcription factors [94] and is a naming member for the characteristic M**A**DS-domain of flower development genes. This homeotic gene determines stamen (together with B-class genes PI, AP3 and E-class gene SEP) and carpel (together with E-class gene SEP) identity and is functionally restricted to the two inner whorls [68][63]. The expression of AGAMOUS is repressed by APETALA2 in sepal and petal tissue [77], as pointed out earlier. AGAMOUS controls different parts of the stamen development, like filament morphogenesis and elongation, as well as anther formation [95]. The mode of action is facilitated by the regulation of different transcriptional signal cascades, including the induction of biosynthesis of the phytohormone jasmonic acid in late-stage stamen development and in order to promote anther dehiscence [95]. However, this mechanism has been elaborated in *Arabidopsis*, but studies in other model plants point to different mode of operations (reviewed in [95]), with non male-sterile JA biosynthesis defective mutants in tomato (*Solanum lycopersicum* L.) as one example [96]. In the floral model plant *Antirrhinum* two orthologues of

have been described, with PLENA (PLE) isolated from mutant lines [97] and FARINELLI (FAR), which shows expression during stamen development [98].

1.5.6 D-class genes - defining ovules

As pointed out earlier, the expansion of the ABC model to include D-class gene started with investigations of *Petunia*, where a class of genes that specifies the identity of the ovule has been demonstrated [64]. Without the activity of the D-class gene, ovules are replaced by structures that resemble carpels [62]. In the model organism *Arabidopsis thaliana* the genes SEEDSTICK, SHATTERPROOF1 and SHATTERPROOF2 have been identified to control ovule identity [99] and thus serve as representatives of the D-class in this species. However, as for all classes, the D-class genes need interaction partners in order to functionalize in the development process, and this is mainly done by the closely related C-class AGAMOUS gene in *Arabidopsis* [62][100].

1.5.7 E-class genes - the "glue" for floral quartets

The E-class was added to the updated version of the flowering model (ABC became ABCDE) and comprises the **SEPALLATA** genes, which are active in all whorls of the flower [65][63]. Hence, they have a high abundance in interactions with other MIKC-type MADS proteins, with an amount of 78 cases where at least one SEP protein appears in a tetramer of totally 106 proved interactions [101]. The interaction of A-class and E-class homodimers results in the development of the sepals [74]. Physical interaction of C-class gene AGAMOUS with SEP1, SEP2 and SEP3 (controlling carpel development) [102] as well as the B-class gene AP3-PI heterodimer interaction with SEP3 (controlling petal and stamen development, depending on the fourth interaction partner, AP1 for petal and AG for stamen, respectively) [103][61] have been demonstrated. Of all E-class genes, SEP3 appears to be the most dominant, and has because of that been ascribed as the "glue" for the formation of the floral quartet-like complexes [101]. In *Arabidopsis*, a shift from nuclear to cytoplasmic localization of SEP3 has been demonstrated in the developing petals and stamens [104].

The preceding chapters have described the development of flowers and the genetic background of this process, including the ABCDE model with its respective genes as well as hormone activity. However, depictions and illustration of these continuous process, like evolution itself, need to be framed for a certain extant to grasp / illustrate these processes. Especially for taxonomist, there is an unavoidable need to take snapshots of processes. And seemingly static things like definitions are crucial for biology and science in order to get a grip on the endless flow.

1.6 Species - a concept?

"It all comes, I believe, from trying to define the indefinable."
- Charles Darwin in a letter to his friend, the botanist Joseph Hooker.

It is not the claim of this work to bring the debate on what species *are* to an end, or to define the indefinable, however, some remarks on this subject are necessary, concerning the topics discussed before and the later presented results.

In the course of this PhD thesis the term "species" is used with the greatest of course, however, it can not be stressed enough that this term is far from satisfiable defined. The ambiguity of the term "species" is the first gateway for confusion, since there are two different objects of discussion, the species taxon (a concrete living biological object consisting of a population of single individuals, e.g. *Salva pratensis*) and the species category (indicating the rank in the Linnaean hierarchy) (reviewed in [105]). This leads to a two-stage problem on how to define species, with (1) what species concept to use, and (2) how to apply the chosen concept on delineating groups of individual from one another. The description and classification of single individuals and groups of individuals inhabiting this planet have undergone several anthropogenic treatments, yielding a vast amount of species concepts.

Because of that, "species" are unstable objects on multiple levels; the individual living entities summarized under the term "species" are to certain extant flexible on ecologically and genetically levels,

which is ultimately the driving force of natural selection. Meanwhile, they are unstable on the theoretically level as well, since discussions about and on "species" are part of the biological community [106]. The genetic and developmental constrain of an individual is more fixed compared to the "species" this organism "belongs to" [107]. But, these various organisms together form the "species", rather than solely be members that are allocated to an umbrella term "species" [106]. Species are held together by biological (reproductive isolation) and by theoretical borders (which are under anthropogenic influence).

Since humans (and thus taxonomists) are for the most part visually perceptive beings, the consideration of a morphological based characterization of individuals is rather obvious. The systematic application of the morphological species concept has been pushed forward by Carl von Linné in his *Species Plantarum* (1753) [108] and *Systema Naturae* (1758) [109] and nowadays there are morphology-based determination keys for almost all multicellular organisms. The morphoglogical species concept is based on visible, countable, measurable morphological traits, and from the similarities in highly specific traits, the affiliation of a group of individuals to a distinct morphospecies can be concluded. These morphospecies have a continuity of traits in the totality of their individuals and are distinguished from other morphospecies by a discontinuity of these traits. A species could therefore be described as a point in an n-dimensional system of n different describable visible characteristics. For the taxonomical practice dichotomous determination keys are used where those characteristics are listed and ultimately lead to a determination result, a species taxon (e.g. *Salvia pratensis*). However, there are some difficulties with this approach, because nature is subject to constant change, which means that differences in characteristics are not completely determined, but, like all of nature, are always in flux. Furthermore, the phenotype under consideration is not completely determined by the genotype, since abiotic factors can lead to phenotypic plasticity, which means different morphology with identical genotype [1]. In addition, morphological determination is often characterized by subjectivity, because in the description of species, vocabulary such as "frequent" or "slightly larger" inevitably appears, which each taxonomist may interpret differently.

The biological species concept extends the view from individuals to the populations they belong to, claiming that phenotypic variations within populations do not fall into account, concerning the definition of a species, as long as all the interacting individuals are able to produce fertile offspring. The reproductive isolation is the keystone of this concept [110].

There are a lot more species concepts, that are often theoretically based and philosophically driven (Ernst Mayr provocatively described the concepts and the authors that put them forward as "armchair taxonomists" [105]), however, a functional working basis when treating biological units should be the primary requirement of any species concept [111]. The species of interest in a study should be physically present (a plant in a botanical garden), in order to be identifiable by the working biologist [111], with a comprehensive description on the way of identification, to ensure the highest degree of objectivity.

"The species is the principal unit of evolution and it is impossible to write about evolution, and indeed about almost any aspect of the philosophy of biology, without having a sound understanding of the meaning of biological species." [105] - Ernst Mayr

1.7 Scope of the study

Since none of the proposed species concepts has fulfilled all requirement that are needed to grasp "species" in their entirety, it is only logical to *not* build a thesis on a single methodological pillar. Conversely, efforts should be made to build a biodiversity study on multiple pillars including different disciplines.

While morphological observations can miss a connection to the possible genetic background, comparatively especially genetic approaches often lack detailed morphological descriptions of the initial object of interest.
The scope of the study aimed therefore at a possible holistic approach on species of the genus *Salvia*.

Starting with the individual living object of interest, the species taxon, the morphological examination includes the statistical framing of characteristic floral traits. Flowers in their entirety represent the physical location of the reproductive isolation for flowering plants, making the framed illustration of this structure a necessary tool for biodiversity investigations. Consequently, the morphological view for an evo-devo approach needs to address the following question (concerning the genus *Salvia*):

Can the statistical evaluation of floral traits be utilized to discriminate closely and distant related species of the species-rich genus Salvia*?*

The formation of unique floral patterns that characterize different species taxa has an evolutionary developmental background driven by gene expression. As elaborated in chapter 1.5.1, the floral developmental genes have been extensively studied in model plants, resulting in the well-accepted ABCDE model. Yet the transmission of this model and its genes that orchestrate the construction of individual flowers has to be demonstrated for the Lamiaceae family ultimately leading to the question:

Can the expression patterns of the ABCDE model be transferred to the genus Salvia*?*

The evolutionary background of the genus *Salvia* and its representative species taxa has been worked out in-depth in the past, however, is still under continuous revision. This background has been inferred from molecular phylogenies, which are mainly based on trait-unrelated barcoding regions. Though phylogenomic approaches start to get used for the Lamiaceae family, a targeted usage of a trait-related marker, that is involved in floral development has not been applied for the Lamiaceae or *Salvia* in particular. To address this issue, the orthologues of model plant B-class genes are of special interest, since those genes shape the corolla and stamen, the crucial floral organs for the sophisticated plant-pollinator interaction that is evident in *Salvia*. The question arising from this issue is whether the documented genes of speciation can be utilized for constructing a relationship of *Salvia*, or:

Can the phylogenetic relationship of genus Salvia *be inferred using developmental genes?*

The issues raised in the preceding paragraphs were addressed in this thesis, using the mentioned and later further specified morphological and molecular methods. To the best of the obtained knowledge, the here unspoken question regarding the species issue shall be discussed at the end.

2 Material

2.1 Plant accessions

Table 1: *Salvia* species identified and used in this study for various experiments

accession ID	species taxon	origin	pollinator	subgenus according to [112]	clade / genus according to [36]
83	*Salvia officinalis* L.	Europe	Bee	*Salvia*	I-D *Salvia* s.s
478	*Salvia tomentosa* Mill.	Europe	Bee	*Salvia*	I-D *Salvia* s.s
1482	*Salvia miltiorrhiza* Bunge	Asia	Bee	*Glutinaria*	IV-B *Glutinaria*
4680	*Rosmarinus officinalis* L. * *Salvia rosmarinus*	Europe	Bee	*Rosmarinus*	*Rosmarinus*
4684	*Salvia lavandulifolia* Vahl	Europe	Bee	*Salvia*	I-D *Salvia* s.s
4686	*Salvia sclarea* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
5207	*Salvia elegans* Vahl	America	Bird	*Calosphace*	II-A *Lasemia*
5208	*Salvia patens* Cav.	America	Bird	*Calosphace*	II-A *Lasemia*
5209	*Salvia pratensis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
6565	*Salvia candelabrum* Boiss.	Europe	Bee	*Salvia*	I-D *Salvia* s.s.
6566	*Salvia canariensis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
7956	*Salvia verbenaca* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
8754	*Salvia hispanica* L.	America	Bee	*Calosphace*	II-A *Lasemia*
8827	*Salvia tiliifolia* Vahl	America	Bee	*Calosphace*	II-A *Lasemia*
8937	*Salvia columbariae* Benth	America	Bee	*Audibertia*	II-B/C *Ramona*
8982	*Salvia austriaca* Jacq.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
8983	*Salvia aethiopis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
8984	*Salvia argentea* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
8986	*Salvia judaica* Boiss	Europe	Bee	*Heterosphace*	*Salvia* s.s
9094	*Salvia jurisicii* Kosanin	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9095	*Salvia przewalskii* Maxim.	Asia	Bee	*Glutinaria*	IV-A *Glutinaria*
9096	*Salvia hians* Royle ex. Benth.	Asia	Bee	*Glutinaria*	IV-A *Glutinaria*
9127	*Salvia farinacea* cv. *VW* Benth.	America	Bee	*Calosphace*	II-A *Lasemia*
9128	*Salvia nemorosa* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9131	*Salvia verticillata* L.	Europe	Bee	*Heterosphace*	*Salvia* s.s
9132	*Salvia splendens* cv. *EB* Sellow ex. Schult.	America	Bird	*Calosphace*	II-A *Lasemia*
9133	*Salvia aethiopis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9134	*Salvia pratensis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9135	*Salvia argentea* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9136	*Salvia austriaca* Jacq.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9137	*Salvia farinacea* cv. *BB* Benth.	America	Bee	*Calosphace*	II-A *Lasemia*
9327	*Salvia ringens* Sm.	Europe	Bee	*Salvia*	I-D *Salvia* s.s.
9328	*Salvia hierosolymitana* Boiss.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9329	*Salvia virgata* Jacq.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9330	*Salvia nutans* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9333	*Salvia glutinosa* L.	Asia	Bee	*Glutinaria*	IV-A *Glutinaria*
9335	*Perovskia atriplicifolia* * *Salvia yangii*	Europe	Bee	*Perovskia*	*Perovskia*
9336	*Salvia pratensis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9345	*Salvia greggii* A.grey	America	Bird	*Calosphace*	II-A *Lasemia*
9346	*Salvia microphylla* Kunth	America	Bird	*Calosphace*	II-A *Lasemia*
9347	*Salvia x jamenensis*	America	Bird	*Calosphace*	II-A *Lasemia*
9349	*Salvia occidentalis* Sw.	America	Bee	*Calosphace*	II-A *Lasemia*

Asterisks indicate uncertainties in circumscription based on the dispute of Will and Bockhoff (2017) [36] and Drew *et al.* (2017) [35]. Describing authors according to Kriebel *et al.* [112]. Pollinators and subgenera according to Kriebel *et al.* [112]. Clades according to Will and Bockhoff (2017) [36]. See also figure 3 in the introduction for an overview of the different nomenclatures.

2.2 Machines

Table 2: Machines used

machine	type	provider
Balance	Analytical plus	Ohaus Corporation, USA
TissueLyser	TissueLyzer	QIAGEN GmbH, Hilden
Vortexer	VortexGenie	Bender and Hobein AG, Schweiz
Eppi Incubator	TSC THermo Shaker	Biometra GmbH, Göttingen
Shaker Incubator	GFL 3033	GFL Gesellschaft für Labortechnik GmbH
Incubator	Memmert	Memmert GmbH + Co. KG
Centrifuge	Heraeus Pico 17	Thermo scientific, Waltham, MA, USA
PCR-tube centrifuge	Mini-Zentrifuge Rotilabo	Carl Roth GmbH + Co. KG, Karlsruhe
96-well plate centrifuge	VWR® PCR Plate Spinners	VWR International GmbH, Darmstadt
Spectrophotometer	NanoDrop ND-1000	Peqlab Biotechnologie GmbH, Erlangen
Thermocycler	FlexCycler	Analytik Jena AG, Jena
	CFX96 Touch™ Real–Time	Bio - Rad Laboratories GmbH, München
AGE chamber	MupidOne	Advance, Mupid CO., Tokio, Japan
Gel scanner	Safe Imager	Invitrogen GmbH, Karlsruhe
Stereo microscope	Leica Stereolupe 420	Leica Microsystems GmbH (Wetzlar)
Camera	Leica DFC 500	Leica Microsystems GmbH (Wetzlar)
Keyence	VHX digital microscope	Keyence Deutschland GmbH
Scanner	Ecosys M4125idn	Kyocera
Digital camera	Cyber-shot DSC-WX350	Sony
Cleanbench	Sicherheitswerkbank HB2448	Heraeus Instruments GmbH, Hanau

2.3 Chemicals

Table 3: Chemicals used

purpose	chemical	provider
Dilutions	Nuclease-free H_2O	Biozym Scientific GmbH
PCR additive	Bovine serum albumine (BSA)	Carl Roth GmbH + Co. KG, Karlsruhe
PCR	dNTPs	New England Biolabs, Frankfurt am Main
cDNA synthesis	dNTPs	New England Biolabs, Frankfurt am Main
AGE	SYBRsafe	New England Biolabs, Frankfurt am Main
	Midori green Xtra	Nippon Genetics Europe GmbH
AGE	Agarose NEEO Ultra-quality	Carl Roth GmbH + Co. KG, Karlsruhe
AGE	100bp ladder	New England Biolabs, Frankfurt am Main
	1kb ladder	New England Biolabs, Frankfurt am Main
cDNA synthesis	Oligo-dT Primer	New England Biolabs, Frankfurt am Main
cDNA synthesis	RNase Inhibitor	New England Biolabs, Frankfurt am Main
qPCR	SYBRgreen	Thermo Fisher Scientific Inc., Waltham, MA, USA
qPCR	$MgCl_2$	USB Corporation, Cleveland, OH, USA

2.4 Buffers

Table 4: Buffers used

purpose	buffer	recipe / provider
AGE	Loading buffer	50% Glycerine / H_2O 0,05% Bromphenolblue 0,05% Xylencyanol
AGE	TAE-Buffer	2 M Tris 0,05 M EDTA 1 M acetic acid
PCR	TB Buffer	500 mM KCl 100 mM Tris-Cl 15 mM $MgCl_2$ [113]
cDNA synthesis	RT-Buffer	New England Biolabs, Frankfurt am Main
qPCR	GoTaq Buffer	Promega GmbH, Mannheim

2.5 Enzymes

Table 5: Enyzmes

purpose	buffer	provider
PCR	Taq DNA polymerase	New England Biolabs, Frankfurt am Main
A-tailing	Taq DNA polymerase	New England Biolabs, Frankfurt am Main
Cloning	T4 Ligase	New England Biolabs, Frankfurt am Main
cDNA synthesis	M-MuLV Reverse Transkriptase	New England Biolabs, Frankfurt am Main
RNA extraction	RNase freie DNase I	QIAGEN GmbH, Hilden
qPCR	GoTaq polymerase	Promega GmbH, Mannheim

2.6 Kits

Table 6: Kits used

purpose	kit	provider
DNA extraction	Invisorb Spin Plant Mini Kit	Stratec molecular, Birkenfeld
PCR purification	MSB Spin PCRapace	Stratec molecular, Birkenfeld
MiniPrep	Roti-Prep Plasmid Mini	Carl Roth GmbH + Co. KG, Karlsruhe
RNA extraction	Spectrum Plant Total RNA Kit	Sigma-Aldrich, Darmstadt

2.7 Cloning

Table 7: Agents for cloning

purpose	object	provider / recipe
Cultivation plates	LB-medium	Carl Roth GmbH + Co. KG, Karlsruhe
Resistance	ampicillin (100 g/ml)	Carl Roth GmbH + Co. KG, Karlsruhe
Amplification of plasmids	*E.coli* strain DH5α	Inoue method
Transformation	pGEM-T Easy Vector System	Promega GmbH, Mannheim
Blue-white screening	IPTG	Carl Roth GmbH + Co. KG, Karlsruhe
Blue-white screening	X-gal	Biomol GmbH, Hamburg

2.8 Software

Table 8: Software used

purpose	software	reference
Sequence quality evaluation	FinchTV version 1.4.0	[114]
Multiple sequence alignment	MEGA7 version 7.0.14	[115]
Evaluation of morphometric data	RStudio version 1.2.5033	[116]
Parameters for Bayersian Inferenece	BEAUti version 1.8.4	[117]
Bayesian inference calculation	BEAST version 1.8.4	[117]
Burn-in for BI calculation	TreeAnnotater version 1.8.4	[117]
Illustration of BEAST phylogenies	FigTree version 1.4.2	[118]
Evaluation of qPCR data	Bio-Rad CFX Manager version 3.1	Bio-Rad Laboratories GmbH

2.9 Databases

Table 9: Databases used

purpose	database	reference
BLASTing and sequence data	NCBI (National Center for Biotechnology Information)	[119]
Reference genome *S. miltiorrhiza*	herbal plant genomes	[120]
Reference genomes Lamiaceae	Mint Evolutionary Genomics Consortium	[31]

2.10 Primers for phylogeny

Table 10: Primers for phylogeny

name	purpose	sequence 5' - 3'	Tm	reference
GLO1b fw	Amplification of	GGGTAGAGGTAAGATTGAGATCAAG	61,6°C	This thesis
GLO2 rv	GLOBOSA gene	GAAACGCTCCTGCAGATTAGGC	64,8°C	This thesis
DEF1 fw	Amplification of	ATGGCTCGTGGGAAGATCCAGATC	67,1°C	This thesis
DEF2 rv	DEFICIENS gene	GCAAATGTAGTGAGGTCCGAGGC	66,3°C	This thesis
ARF1ex2 fw	Amplification of partly	TCAGGCACCCCCCTAGGTCCAA	70,4°C	This thesis
ARF1ex2 rv	2^{nd} exon of ARF1	GCTTGGTTTCATAAGAATAGTCAATC	60,5°C	This thesis
M13 fw	colony PCR	CGCCAGGGTTTTCCCAGTCACGAC	76,7°C	obtained
M13 rv	and sequencing	TCACACAGGAAACAGCTATGAC	61,8°C	from [121]

2.11 qPCR primers

Table 11: qPCR primers

name	purpose	sequence 5' - 3'	Tm	reference
18S fw	Amplification	GCGGAGTCCTAGAAGCAACA	64,4°C	[122]
18S rv	of 18S	CTTCGGGATCGGAGTAATGA	63,9°C	[122]
Actin fw	Amplification	AGGAACCACCGATCCAGACA	66,9°C	[122]
Actin rv	of actin	GGTGCCCTGAGGTCCTGTT	66,6°C	[122]
Ubiquitin fw	Amplification	GTTGATTTTTGCTGGGAAGC	63,4°C	[122]
Ubiquitin rv	of ubiquitin	GATCTTGGCCTTCACGTTGT	64,0°C	[122]
AP1 qPCR fw	Amplification	GAGGGAAAGTGGAATTGAAGAG	62,8°C	This thesis
AP1 qPCR rv	of APETALA1	CCTTTGCTGGAGAAGATGATGAG	66,0°C	This thesis
GLO1b fw	Amplification	GGGTAGAGGTAAGATTGAGATCAAG	61,6°C	This thesis
GLO1 qPCR rv	of GLOBOSA	TATCTTCTCCTTTCAGGTGCCT	62,6°C	This thesis
DEF qPCR Eu fw	Amplification	TGATGAATACCAGAAGACTGTTAGGG	65,0°C	This thesis
DEF v1 qPCR Eu rv	of DEFICIENS	CCGAGGTCTTTATACTTTCTATTGC	61,9°C	This thesis
DEF v2 qPCR Eu rv	copies Europe	CGCTCCATATCTTCGTCAGATC	65,8°C	This thesis
DEF v1 qPCR Am fw	Amplification	CGCGGTTCGCCTTCCGCCTA	76,9°C	This thesis
DEF v2 qPCR Am fw	of DEFICIENS	GCCCTTCACATGCCACCGCC	76,1°C	This thesis
DEF qPCR Am rv	copies America	GCAAATGTAGTGAGGTCCGAGGC	66,3°C	This thesis
AGA qPCR Eu fw	Amplification	AGAGCTGCACAATGCTAACATG	64,2°C	This thesis
AGA qPCR Am fw	of AGAMOUS	AGAGCTGCACAATGCAAACATG	67,3°C	This thesis
AGA qPCR rv		GTTTGATCCTGGCAAGAGTAGTG	63,7°C	This thesis
SHP qPCR fw	Amplification of	GGTGTGGAGAAATGAAATTGAGGA	66,9°C	This thesis
SHP qPCR rv	SHATTERPROOF	GCTGGCTTCTTGAAGCTCGTG	68,5°C	This thesis
SEP1 qPCR fw	Amplification	TCTTCAATCAAAGGAGAAGTTGATG	64,3°C	This thesis
SEP1 qPCR rv	of SEPALLATA1	TGGAGAGTAGAGTTGCAATCA	60,1°C	This thesis
SEP3 qPCR fw	Amplification	GAGTTGAGCTGAAGAGGATAGAGAAC	63,7°C	This thesis
SEP3 qPCR rv	of SEPALLATA3	GAGAAGATAATCAGAGCAACCTC	59,9°C	This thesis
ARF1 qPCR fw	Amplification	CAAGGTGTTCTTGGAGTCGGA	66,4°C	This thesis
ARF1ex2 rv	of ARF1	GCTTGGTTTCATAAGAATAGTCAATC	60,5°C	This thesis
ARF25 qPCR fw	Amplification	CTTCCAGCTACATGACTAGCTC	59,7°C	This thesis
ARF25 qPCR rv	of ARF25	CACAAATACAAGCTGCCAGCC	66,6°C	This thesis

3 Methods

3.1 Species authentication by floral traits

In order to obtain a reliable framework for all subsequent (morphological and molecular) investigations, various morphology-based determination keys were used to identify the reference plant material. Flowers are the control centers of sexual reproduction of plants and therefore the physical places where reproductive isolation actually takes place, meaning that differences in floral morphology are directly linked to barriers of gene flow. It is therefore not surprising that angiosperms can be distinguished and categorized by their flower structure. To authenticate the individuals that were part of this PhD thesis, the floral traits listed in the "Die Flora Deutschlands und der angrenzenden Länder" [11], the Flora of China [123] and the "Illustrierte Flora von Mittel-Europa" [39] have been used to confirm the identity of available *Salvia* individuals. Some *Salvia* material has been provided and determined by Dr. Max Seyfried, namely the individuals of *S. occidentalis*, *S. brousonetii*, *S. canariensis* and *S. somalensis*. Flowers have been documented and pictures have been deposited in the database of the Botanical Garden of the KIT (nicklab.de - [124]).

3.2 Floral morphometrics

In the Botanical Garden of KIT several of the reference plants developed flowers that could be used for a morphometric approach. The measurement parameters of Benitez-Vieyra *et al.* (2019) [125] were applied to 24 available *Salvia* species (see table 1). The lower lip length (lll), lower lip width (llw), upper lip length (ull), upper lip width (ulw), corolla tube length (ctl) and corolla tube width (ctw) were measured according to Benitez-Vieyra *et al.* (2019) [125] (and see figure 13), additionally the length of the stamen (connective + theca)(stamen length complete: slc) and the style (style length complete: stlc) were included into the dataset. For all species the above mentioned traits from in total 30 flowers (except *Salvia greggii*, *Salvia microphylla* and *Salvia x jamenensis*, for those species due to availability only 15 flowers were measured) of two different individual plants were measured. The measurements were analyzed using the software Rstudio (Version 1.2.5033) [116]. The dataset was evaluated in two different ways, firstly by principal component analysis (a two-dimensional projection of a multidimensional dataset, using the *prcomp* command), secondly as a dendrogram (a hierarchically based representation, using the *hclust* command). For both evaluations, the average means of all measurements for one species was calculated and used for the analysis.

3.3 DNA extraction

Genomic DNA from fresh leaves (using 60 mg of starting material) of the available *Salvia* individuals and outgroup plants from the Botanical Garden of KIT (see table 1) was isolated using the Invisorb Spin Plant Mini Kit (Stratec Biomedical AG), following the manufacturers instructions. The quality and quantity of isolated gDNA was evaluated by spectrophotometry (NanoDrop, Peqlab), and DNA concentration was diluted to 50 ng / μl to be used as template in polymerase-chain-reaction.

3.4 Primer design for B-class genes of genus *Salvia*

In order to specifically amplify B-class genes from genomic DNA of Lamiaceae (and *Salvia* species in particular) individual primers for the respective genes (GLOBOSA or DEFICIENS) were designed. B-class genes have been described in the early 90s (see introduction in chapter 1.5.1), and the sequences for model plants (like *Arabidopsis thaliana* or *Antirrhinum majus*) are available in the ncbi database, however, in one of the largest families of the angiosperms, the Lamiaceae, the information content for annotated floral developmental genes is surprisingly scarce. Apart from the model plant *Antirrhinum majus*, the developmental genes are satisfactorily annotated in sesame (*Sesamum indicum*), which belongs to the order Lamiales and therefore is closer related to *Salvia* than *Antirrhinum*. In 2016 Xu *et al.* published the annotated genome of *Salvia miltiorrhiza*, a medicinal plant, which is widely used in the Asian region for the treatment of cerebrovascular and cardiovascular diseases [57]. The B-class sequences of *Sesamum indicum* were used as BLAST template in the *Salvia miltiorrhiza* database [120], the top result was downloaded and based on both sequences that were aligned using the MEGA (version 7.0.14) software

[115] primers were designed. These primers flank the respective B-class genes from its start codon to its stop codon, including all exons and introns.

3.5 PCR (mastermix / setting / gel electrophoresis / PCR product cleaning)

A 30μl reaction volume containing 20.4μl nuclease free H_2O (Lonza, Biozym), onefold Thermopol Buffer (New England Biolabs), 1 mg / ml bovine serum albumin, 200 mM dNTPs (New England Biolabs), 0.2 mM of forward and reverse primer (see primer list, table 10), 100–150 ng gDNA template and three units of Taq polymerase (New England Biolabs) was used to amplify the B-class gene of interest. Thermal cycler conditions for the amplification included initial denaturation at 95°C for 2 min; following 35 cycles at 94°C for 45s, 58°C for 45s, 68°C for 2 min; ending with an extension of 68°C for 5 min. The PCR was subsequently evaluated by agarose gel electrophoresis using NEEO ultra-quality agarose (Carl Roth, Karlsruhe, Germany). DNA was visualized using SYBRsafe (Invitrogen, Thermo Fisher Scientific, Germany) or Midori green Xtra (Nippon Genetics Europe GmbH) and blue light excitation. The fragment size was determined using the 100 bp or 1 kb size standard (New England Biolabs). Amplified DNA was purified using the MSB Spin PCRapace kit (Stratec) according to the manufacturers instructions for subsequent cloning steps.

3.6 Cloning (A-tail, ligation, transformation, blue-white screening, miniprep)

For the A-tailing reaction 7μl of the cleaned PCR product was used in a total 10μl reaction volume containing additionally 1,7μl nuclease free H_2O (Lonza, Biozym), 1μl of 10x Thermopol Buffer (New England Biolabs), 0,2μl of 200 mM dNTPs (New England Biolabs) and one unit of Taq polymerase (New England Biolabs). This mixture was incubated at 68°C for 60 min and directly used for the subsequent ligation step. This step was performed overnight at 4°C with 2μl A-tailing product added to a mixture of 5μl ligation buffer, 1μl of pGEM-T Easy Vector System, 1μl of T4 ligase and 1μl of nuclease free H_2O. Ligation product was mixed with 40μl of chemo competent DH5α *E. coli* bacterial cells and after 30 min incubation on ice, transformation took place in a 42°C water bath for 50 sec. Samples were returned to ice directly for two minutes and after adding 950μl of LB medium, the transformed cells were incubated at 37°C with permanent shaking at 180 rpm for two hours. After two hours the cells were centrifuged at 3000 rpm for 2 min, then 850μl of medium removed, and the pelleted cells resuspended into the medium and then transferred to an LB-agar plate containing ampicillin (0,1%), IPTG (0,2%) and X-gal (0,8%). Plates were incubated overnight at 37°C and then stored at 4°C until the blue white screening. White (positive) colonies were picked sterile and transferred to single eppis containing 1 ml liquid LB medium and 0,1% ampicillin. The following overnight incubation at 37°C and shaking at 180 rpm was briefly interrupted after two hours to extract 1μl of medium (containing the transformed *E. coli*), which was used as a template for a colony PCR. Thermal cycler conditions corresponded to the earlier used program, with the difference that the universal primer pair M13 (fw/rv) (see table 10), that is nested in the vector, was used. Colonies with the correct sized insert were cultivated further, false positive colonies were discarded. As a last step the plasmids of the overnight incubated positive colonies, with the insert of desire, were isolated by mini-prep, using the Roti-Prep Plasmid Mini kit according to the manufacturers instructions. Sequencing was outsourced to different companies, the amount and concentration of samples sent was adjusted according to the company requirements. For both, GATC (which became eurofins in the course of this thesis) and macrogen, the universal M13 primers were used for sequencing.

3.7 Evaluation of sequencing results

The quality of the obtained sequences was examined with the software FinchTV version 1.4.0 [114]. Sequencing was done from two directions using the M13 forward and reverse primers (table 10). MEGA version 7.0.14 [115] was used to align the overlapping regions of the complete fragment to receive the full sequence of the B-class genes of different *Salvia* species. In order to detect the coding regions in the full sequence, the cDNA (which was obtained for another experiment, see section 3.8) of European *Salvia pratensis* and America *Salvia elegans* was used. The cDNA sequence was obtained by the same procedure as described in 3.5 and 3.6 (using cDNA instead of gDNA). MEGA7 was used to create a multiple sequence alignment of all gDNA and cDNA sequences of the respective gene. The sequences

were aligned using the MUSCLE algorithm [126] which is integrated in MEGA7. For investigations that relied on coding sequences the introns could be excised in-silico from the complete gene for the whole *Salvia* dataset.
The complete GLOBOSA and DEFICIENS sequences of *Salvia miltiorrhiza* were gathered from the herbal plant database [120], the sequences of *Sesamum indicum*, that functioned as an outgroup, from the ncbi database [119][127]. The full length multiple sequence alignment (including exons and introns) has been used for Bayesian inference. For Bayesian inference the BEAST package 1.8.4 was used [117]. Markov chains for 10 million generations, with sampling every thousandth generation, were performed to analyze the datasets. The calculation parameters were set using BEAUti, the calculation itself was performed with BEAST. The burn-in (10 percent; 1.000) was determined with TreeAnnotater, which is integrated in the BEAST package as well [117]. To illustrate the constructed trees the Java-based software FigTree version 1.4.2 has been used [118], for degree of significance the posterior probabilities (PP) option was chosen.

3.8 Sampling of floral organs / RNA extraction / cDNA synthesis

Closed and opened flowers of bee-pollinated purple-colored *Salvia pratensis* (indigenous in Central Europe) and hummingbird-pollinated red-colored *Salvia elegans* (originally from the New World) from plants in the Botanical Garden of the KIT were dissected into five parts (calyx/sepal, corolla/petal, ovary, stamen and style)(see figure 11 and 12), separated in five tubes respectively and immediately shock-frozen in liquid nitrogen. This procedure was done three times (biological replicates) for two developmental stages (closed / open) for both species, resulting ultimately in 60 samples in total. Closed flowers were chosen as: corolla visible, but closed, flower buds of total length 1 cm in *S. pratensis*, and total length of 2 cm in *S. elegans*. Opened flowers were chosen as: corolla open, releasing the reproductive organs, flower ready to pollinate, however, yellowish pollen still visible at the anthers to ensure recent opening. Frozen tissues (50 – 100 mg) were ground to a powder by a Tissue Lyzer prior to extraction of total RNA using the Spectrum™ Plant Total RNA Kit (Sigma-Aldrich, Deisenhofen, Germany) following the manufacturers instructions.

Figure 11: Example of closed flower of *Salvia pratensis* and the isolated floral organs used for RNA extraction. Calyx, corolla, ovary, stamen and style, from left to right.

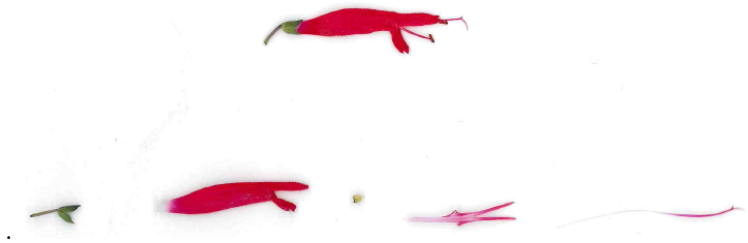


Figure 12: Example of opened flower of *Salvia elegans* and the isolated floral organs used for RNA extraction. Calyx, corolla, ovary, stamen and style, from left to right.

For cDNA synthesis, 1 μg of RNA was subjected to reverse transcription using the reverse transcriptase M-MuLV. For this reaction a primarily mastermix containing 0,4 μl of 10 mM dNTPs and 1 μl of oligo dT (100 μM) was mixed with the respective volume of RNA to reach the final RNA concentration of 1 μg. Depending on the concentration and the amount of RNA the reaction was adjusted to a total

volume of 16 μl with RNase-free H_2O. This reaction was incubated at 65°C for 5 min and subsequently stored on ice before adding 4 μl of the second mastermix which was composed of 2 μl of reverse transcriptase buffer, 0,25 μl reverse transcriptase (MuLV), 0,5 μl RNase inhibitor and 1,25 μl RNase-free H_2O. The reaction was continued with 42°C for 60 min, followed by 90°C for 10 min and a finalizing storage step at 12°C.
The transcript level of the floral developmental genes was determined by quantitative real-time PCR (qPCR), performed with the CFX96 Touch™ Real-Time PCR Detection System from Bio-Rad Laboratories GmbH (Munich), using the qPCR primers given in table 11 (see chapter 3.9). qPCR analysis was carried out in 20 μl reaction tubes containing final concentrations of 200 nM of each primer and each dNTP, 1x GoTaq Buffer, 2.5 mM $MgCl_2$, 0.5 units of GoTaq polymerase, 1x SYBR green and and 1 μl of 1:20 cDNA dilutions.
Thermal cycler conditions for the qPCR experiment included an initial denaturation at 95°C for 3 min; following 39 cycles at 95°C for 15 sec, 58°C for 40 sec and a plateread step; followed by a 10 sec denaturation step at 95°C for the melting curve. The melting curve was conducted by in total 60 repetitions of 5 sec steps that were started at 65°C with an increase of temperature by 0.5°C per step.
As an internal standard, house keeping genes 18S, Actin and Ubiquitin have been selected [122], and for each of the three biological replicates, three technical replicates in the same reaction have been conducted. The relative expression of the floral developmental genes was calculated by averaging the normalized c_t values (G) and deducting this value from the house keeping gene reference c_t value (H) that acts as an internal standard for the respective sample.

Formula: $\Delta c_t = c_t(G) - c_t(H)$

The resulting difference was applied to the $2^{-\Delta ct}$ formula to gather the relative gene expression value. This value has been used for illustrating the expression patterns of the floral developmental genes in the dissected floral tissues. To test the significance, standard deviation for every triple pair of biological replicates has been calculated. An additional test for significance was conducted by applying a two-tailed t-test, whose significance values were visualized by asterisks (*: p-value = 0.05, **: p-value = 0,01) in the results.

3.9 qPCR primer design utilizing the MINT database

The available cDNAs for the different floral organs were ought to be applied to the key players of floral development, therefore the representative genes of the ABCDE-model were selected for the analysis. Since the sequence information for GLOBOSA and DEFICIENS became available in the course of this thesis (see section 3.7), primers optimized for qPCR were designed referring to this dataset (see table 11). In 2018, the Mint consortium published the non-annotated genomes of 48 Lamiaceae species [31], including two *Salvia* species (namely *Salvia officinalis* and *Salvia hispanica*); or four, when sticking to the new nomenclature of *Salvia*, pushed forward by Drew *et al.* (2017) [35] and adopted by the Mint consortium as well. Taking into account this consideration including *Rosmarinus officinalis* (aka *Salvia rosmarinus*) and *Perovskia atriplicifolia* (aka *Salvia yangii*) would make in total four *Salvia* species. Anyways, apart from the (later discussed) recent taxonomic quarrel, the four mentioned species genome datasets have been used for identification of sequence regions suitable for qPCR primers for the respective floral developmental genes.

The mRNA data from ncbi dataset of *Sesamum indicum* has been used for BLASTing for homologues in the *Salvia miltiorrhiza* database. The top results for APETALA1 (AP1), AGAMOUS (AGA), SEPALLATA (SEP1), SEPALLATA3 (SEP3), SHATTERPROOF (SHP), Auxin Response Factor 1 (ARF1) and Auxin Response Factor 25 (ARF25) were used for BLASTing in the genomes of the two/four *Salvia* genomes. The received top results were again double-checked against the ncbi and the *Salvia miltiorrhiza* database to avoid false results. The obtained reliable sequences were aligned in a multiple sequence alignment for each gene of interest, respectively. In this alignments, conservative (and therefore semi-universal, which means universal for all *Salvia* species and most likely for some other Lamiaceae genera) sites have been chosen to function as qPCR primer binding sites. The amplicons of the designed primers ranged from 100 bp to 300 bp, with annealing temperatures between 59.7°C and 67,3°C. The primer pairs have been tested for their functionality in semi-qPCR experiments prior to applying to qPCR experiments. Thermal cycler conditions for the semi-qPCR included initial denaturation at 95°C for 2 min; following 35 cycles at 94°C for 30s, 58°C for 30s, 68°C for 30s; ending with an extension of 68°C for 5 min and a subsequent evaluation as in 3.5.

3.10 Evaluation of the ABCDE gene phylogeny in *Salvia*

As pointed out earlier, during this thesis a large amount of data on genomes of interest has become publicly available. Additionally to the annotated *Salvia miltiorrhiza* genome and the raw data from the MINT consortium (as described earlier), the short raw sequence reads for ten *Salvia* species (*S. bulleyana*, *S. chanryoenica*, *S. divinorum*, *S. hispanica*, *S. japonica*, *S. officinalis*, *S. pratensis*, *S. przewalskii*, *S. sclarea* and *S. splendens*), as well as *Rosmarinus officinalis* became lately available on ncbi [128]. B-class genes GLOBOSA and DEFICIENS have been evaluated in detail for various *Salvia* species in the context of phylogenetic background of the genus *Salvia* (as described in 3.6 and 3.7). For the gene expression experiment in developing flowers the focus was broadened on representatives genes from the ABCDE model (as described in 3.8 and 3.9). In order to have a merge of both approaches, part of the obtained sequence data from B-class GLOBOSA and DEFICIENS genes of different *Salvia* species has been used together with ACDE class sequences that were extracted from raw sequence data of whole genome sequence projects. The European bee-pollinated *Salvia pratensis* and the American hummingbird-pollinated *Salvia splendens* species were chosen to be included in a general overview analysis on *Salvia* ABCDE flowering gene phylogeny. The short reads of the coding regions from APETALA1 (A-class), AGAMOUS (C-class), SEEDSTICK (D-class) and SEPALLATA3 (E-class) genes were extracted from the bioprojects of *Salvia pratensis* (provided by the University of Life Sciences Prague [129]) and *Salvia splendens* [130], based on already obtained sequences of *Sesamum indicum* and *Salvia miltiorrhiza* for the design of qPCR primers (see 3.9). The respective coding regions for both species were assembled manually from the single sequencing reads using MEGA7 [115], resulting in a dataset of at least four species (*Sesamum indicum*, *Salvia miltiorrhiza*, *Salvia pratensis* and *Salvia splendens*) for all above mentioned ACDE genes. As additional outgroup for the B-, C- and E-class genes, sequences of *Arabidopsis* (a rosid plant) orthologues were included (received from ncbi). All evaluated genes belong the MIKC-type MADS gene class (see introduction 1.5.1), making a further outgroup necessary to put all data into a meaningful context, which was achieved through the introduction of a LEAFY gene (a functionally upstream (but not related) key floral developmental gene [131][132]) coding sequence. Prior to further analysis steps, all sequences were tested on their theoretical functionality by translating the coding sequences into amino acid sequences, in order to detect and discard sequences with possible unexpected stop codons. The final dataset consisting of four APETALA1 (A-class gene), eleven GLOBOSA (B-class gene), nine DEFICIENS (B-class gene), five AGAMOUS (C-class gene), three SEEDSTICK (D-class), six SEPALLATA3 (E-class) genes and the LEAFY outgroup was translated to amino acid sequences and subsequently aligned using MEGA7 [115] and the MUSCLE algorithm [126]. This dataset was used for BEAUti / BEAST analysis, Tree Annotator evaluation and illustration with FigTree (as described in 3.7) [117][118].

3.11 Utilizing the Auxin Response Factor 1 for phylogeny

In order to understand the mode of action of auxin, its transcriptional representatives (among others, the Auxin Response Factors) have been of special interest for a long time. For this reason *Salvia miltiorrhiza* has been in-depth investigated with regard to the Auxin Response Factors in 2016 [57]. The phlyogenetic analysis revealed close relations to homologue ARF genes of *Arabidopsis thaliana*. Plant tissue specific expression revealed the elevation of certain ARFs in distinct plant tissues. The smARF1 has shown an elevated expression in floral tissue of *Salvia miltiorrhiza*, for this reason it has been included to the expression study of this thesis (as mentioned above). Furthermore, a part of the second exon of this gene has been used for a phylogenetic approach as well, in order to compare *Salvia* auxin-based phylogeny to the B-class gene based phylogeny. The ARF1 primers listed in table 10 have been used to amplify the 2nd exon of the Auxin Response Factor 1 in *Salvia* according to 3.5. The second exon was chosen, because the first exon of smARF1 includes the highly conserved DNA-binding-domain, whereas the second exon with its less conserved regions is more promising in terms of sequence varieties. In contrast to the B-class genes, cloning was not necessary for sequencing, making the acquisition of sequences easier and the amount of sequenced individuals higher (for all accessions used for ARF1 exon 2 sequencing, see table 13 in the appendix).

4 Results

4.1 Short guideline to the result part

The first part (starting chapter 4.2) of the following results focuses on the morphological examination of the plants (the species taxa) of interest, the living entities that represent the backbone of all subsequent experiments. This investigations dealt with an in-depth morphological analysis of the *Salvia* species (methods used are elaborated in 3.1 and 3.2). The main interest hereby laid on the corolla and the stamen, because of their invaluable importance on attracting pollinators and attaching the pollen to the insects or birds body.

- Morphological description of species of interest (chapter 4.2 and 4.3.1 or pages 33 to 35)
- Statistical evaluation of floral measurements (chapter 4.3.2 and 4.3.3 or pages 36 and 37)

The second part (starting chapter 4.4) represents a merging of morphological and molecular approaches, marking a smooth transition between those two main topics within this thesis. The morphological sectioning of single flowers and subsequent molecular evaluation by quantitative PCR (methods used are elaborated in 3.8 and 3.9) of two distantly related *Salvia* species, *S. pratensis* and *S. elegans*, yields results for evo-devo statements.

- General results on the outcome of the qPCR experiment (chapter 4.4 to 4.4.4 or pages 38 and 39)
- Gene expression of ABCDE genes in the context of flowering in *Salvia* (chapter 4.4.5 to 4.4.15 or pages 39 to 46)
- Evaluation of Auxin Response Factors in developing *Salvia* flowers (chapter 4.4.16 or page 47)

The third part (starting chapter 4.5) displays the analysis of sequencing data and targets on an illustration of the genus *Salvia* phylogeny based on trait-related markers. The corolla and stamen shaping B-class genes GLOBOSA and DEFICIENS as well as the transcriptional representatives of plant hormone auxin were taking into account (methods used are elaborated in 3.4 to 3.7 and 3.10 and 3.11).

- ARF1 evaluation and phylogeny (chapter 4.5 or pages 48 to 50)
- B-class GLOBOSA gene evaluation and phylogeny (chapter 4.6 or pages 51 to 56)
- B-class DEFICIENS gene evaluation and phylogeny (chapter 4.7 or pages 57 to 59)
- On the duplication of B-class and a resulting model (chapter 4.8 or pages 60 and 61)
- ABCDE genes phylogeny including *Salvia* sequences (chapter 4.9 or page 62)

4.2 The spectrum of *Salvia* flowers

In order to obtain an scientific and aesthetic understanding of the diversity of flowers in a seemingly relatively narrow corridor of a single genus (*Salvia*), the images of *Salvia* inflorescences provided are intended as a visual excitation and morphological fundament for this thesis.

(a) *Salvia officinalis*

(b) *Salvia tomentosa*

(c) *Salvia lavandulifolia*

(a) *Salvia pratensis* **(b)** *Salvia nemorosa* **(c)** *Salvia sclarea*

(d) *Salvia verbenaca* **(e)** *Perovskia atriplicifolia* **(f)** *Rosmarinus officinalis*

(g) *Salvia glutinosa* **(h)** *Salvia miltiorrhiza* **(i)** *Salvia przewalskii (*)*

(j) *Salvia elegans* **(k)** *Salvia patens* **(l)** *Salvia microphylla*

Figure 13: Selection of *Salvia* species that were part of this study. With (a) to (i) European *Salvia* species, (j) to (l) East Asian *Salvia* species and (m) to (o) New World *Salvia* species.
(*) please note the insect in the flower of *S. przewalskii*, only visible by the backmost part of the abdomen and the tarsi.

4.3 Applying morphometrics of floral traits to genus *Salvia*

4.3.1 Statistical evaluation of visible floral traits reveal the floral diversity spectrum of genus *Salvia*

The size and shape of flowers from the genus *Salvia* vary remarkably, ranging from five centimeter big vividly blue colored blossoms (e.g. *Salvia patens*), to smaller than one centimeter inconspicuous inflorescences, that can remain closed all the time during development (e.g. *Salvia verbenaca*). The two above mentioned species were part of this study, as well as a wide range of other species, including a large selection of differently pollinated species (see table 1). In order to quantify and statistically illustrate the variety of floral diversity in the genus *Salvia* measurements of different traits from corolla (see figure 13), stamen and style of 24 available species was conducted and subsequently used for principal component analysis (see the contribution of principal components in figure 13) and plotting of a dendrogram, to elucidate the floral morphological relation of the species.

The raw data of lower lip length ranged from 0,15 centimeters in some floral individuals of European *Salvia verbenaca* and American *Salvia columbariae* to 3,2 centimeters in the American species *Salvia patens* flowers. The length of the upper lip varied from 0,2 centimeters in American *Salvia tiliifolia* flowers to 3,0 centimeters in *Salvia patens*. Measurements of the corolla tube length, which is a decisive feature for the way of pollination or visiting pollinators, revealed that again the American *Salvia columbariae* and *Salvia tiliifolia* and the European *Salvia verbenaca* have the shortest corolla tube of all investigated species with lengths of around 0,5 centimeters, respectively. The bright red colored corolla tube of the New World hummingbird-pollinated *Salvia splendens* showed the highest values with maximums of 3,6 centimeters. Differences in length of the stamen could be observed with the most elongated connectives in individual flowers of *Salvia patens* that had a complete length of 4,25 centimeters, which is connected to the length of the upper lip of the corolla. Accordingly, the shortest stamen were measured in *Salvia tiliifolia* and *Salvia columbariae* with lengths around 0,2 centimeters, respectively. The broad spectrum of floral colors of the different *Salvia* species was not included into the subsequent dataset and are therefore absent in the principal component analysis, however, the range from white (e.g. *Salvia aethiopis* or *Salvia argentea*), to pinkish- or purplish-white (e.g. *Salvia sclarea* or *Salvia verbenaca*), to purple (e.g. *Salvia officinalis*, *Salvia pratensis*, *Salvia miltiorrhiza*), yellow (e.g. *Salvia glutinosa*), blue (e.g. *Salvia patens*) and red (*Salvia splendens* or *Salvia elegans*) can be viewed in figure 13.

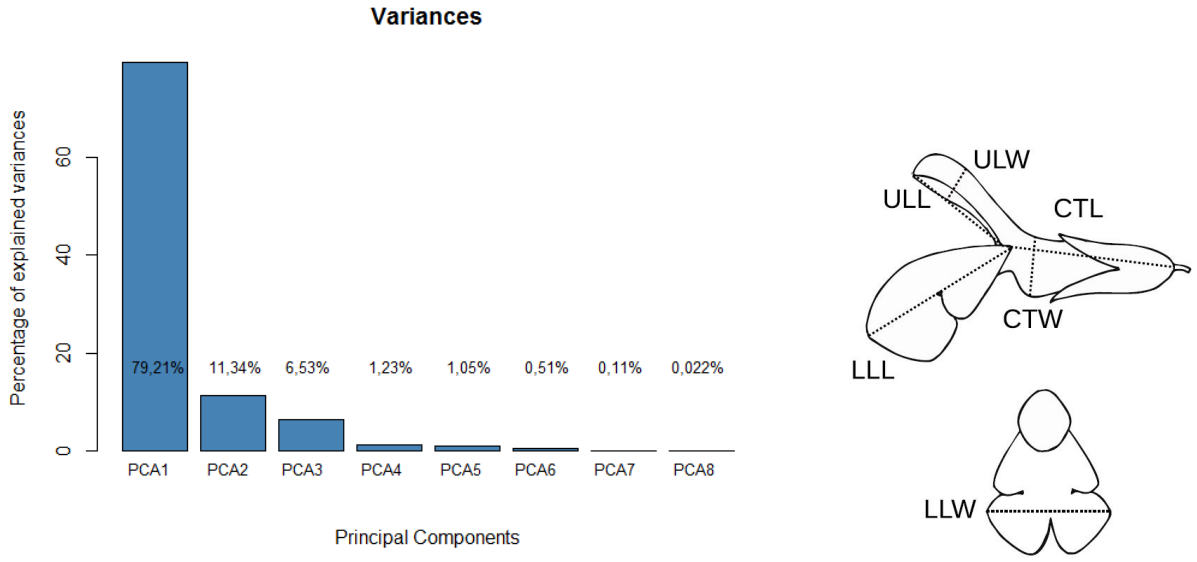


Figure 13: (A) Barplots of principal component relevance for the measured floral traits of different *Salvia* species, whereby PC1 describes size and PC2 the floral shape, (B) Floral measure sites of the single *Salvia* flowers, based on Benitez-Vieyra *et al.* 2019 [125]. lll: lower lip length, llw: lower lip width, ull: upper lip length, ulw: upper lip width, ctl: corolla tube length, ctw: corolla tube width, additionally stc: stamen length complete, stlc: stylus length complete

4.3.2 Morphology-based principal component analysis illustrates overlapping geometric floral patterns in geographically separated *Salvia* species

For the principal component analysis the combined measurements of the lower lip length (lll), lower lip width (llw), upper lip length (ull), upper lip width (ulw), corolla tube length (ctl), corolla tube width (ctw), stamen length complete (slc) and stylus length complete (stlc) was used, and resulted in a dataset of 660 single measurement of flowers from 24 species for the eight variables, respectively. The importance of components was calculated using the Rstudio software, which resulted in principal component 1 (PC1) explaining 79,21% and principal component 2 (PC2) explaining 11,34% of the evaluated data (see figure 13). This means that the two-dimensional matrix (see figure 14) displays 90,55% of all provided information.

In figure 14 the PC1 and PC2 the floral measurements of the 24 evaluated *Salvia* species are plotted in a two-dimensional matrix. The distribution is independent of the geographic appearance from the species in the *Salvia* genus. The difference in floral geometries that can be seen with the bare eye are transferred to the statistical evaluation. A general trend is the reduction of size along the x-axis (PC1) and the transformation of the corolla shape from tuberous to clear lip shape along the y-axis (PC2). The green labeled European species build are spread cluster that is localized in the upper right part of the graph, with two remarkable outliers, that are *Salvia argentea* and *Salvia sclarea* (both pollinated by carpenter bees). The closely related *Salvia officinalis* and *Salvia lavandulifolia* (which is often described as subspecies of *Salvia officinalis*, *S. officinalis* subsp. *lavandulifolia*) cluster closely together with a small shift in the PC1, meaning the overall slightly bigger flowers of *Salvia officinalis* are projected on the x-axis. The blue labeled American species are spread all over the matrix and do not form a visible cluster. The exception are the three closely related species *Salvia greggii*, *Salvia microphylla* and *Salvia x jamenensis* that cluster together with the Eurasian *Salvia hierosolymitana*.

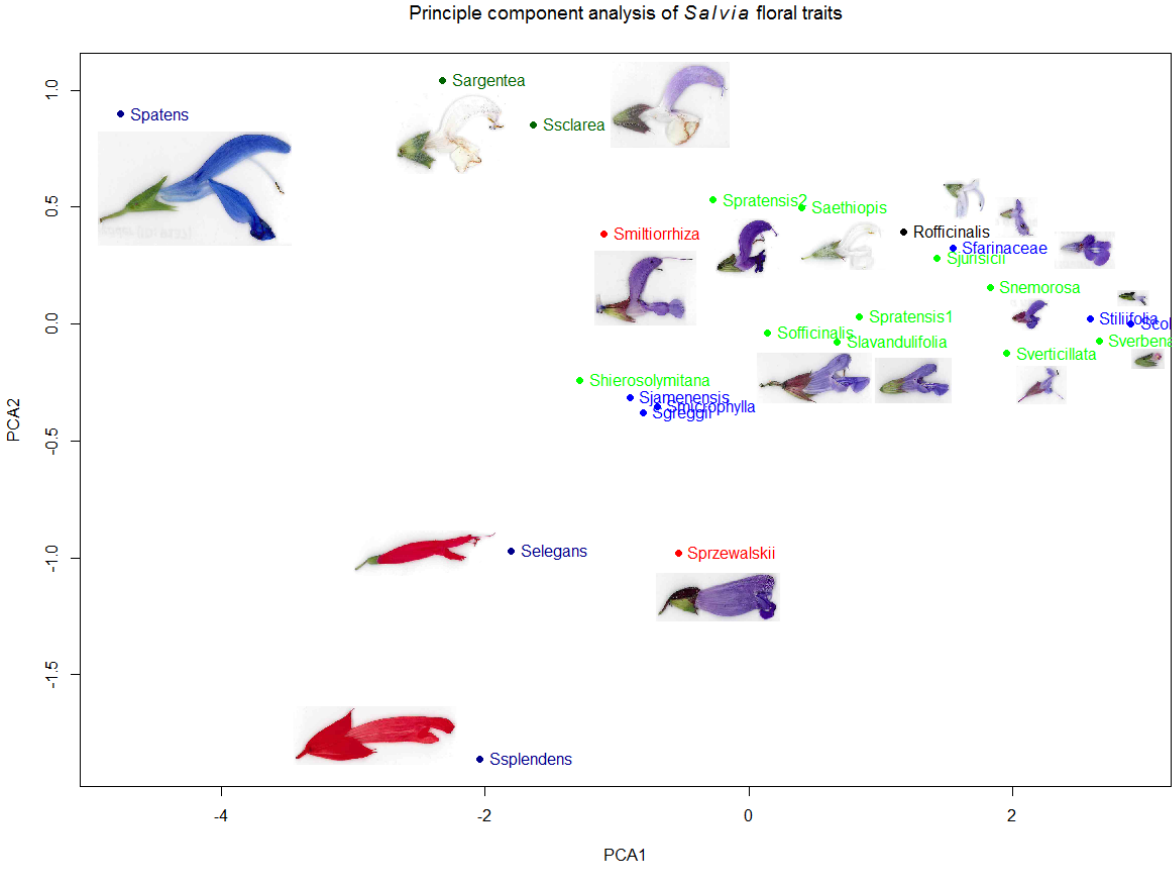


Figure 14: Principal component analysis (PCA) of *Salvia* floral traits. The measurements of the flowers (details see methods 3.2) of 24 different *Salvia* species were evaluated by Rstudio software. The average values of 30 flowers from every species have been plotted. The two plotted principal components project 90,55% of the original data. PCA1 covers 79,21% and PCA2 covers 11,34% of the observed information, respectively.
Flower pictures correspond to the actual size proportions.

The other New World species (*Salvia tiliifolia* and *Salvia columbariae*) cluster with the small-sized European *Salvia verbenaca*, while *Salvia patens* with its huge upper lip and *Salvia splendens* with its strongly elongated (and adapted to hummingbirds) corolla tube are outliers in this dataset. The latter forms a loosely group with the other two more tuberous-shaped flowers, namely hummingbird-pollinated *Salvia elegans* and *Salvia przewalskii*. *S. przewalskii* though is pollinated by bees, that crawl completely into the flowers (see figure 13i, red arrow), which is only possible because the widening of the corolla tube leaves enough space for the insects body. This, however, is not possible in *S. elegans* anymore (own observation), because the corolla tube is too narrow, an adaptation to the beak of hummingbirds.

4.3.3 Morphology-based dendrogram demonstrates unclassifiability of geographically distinct *Salvia* species by floral traits

The dendrogram is based on a distance matrix that was calculated in Rstudio using the same measurements as for the principal component analysis. The topology of the clusters reflects the results already indicated in the diagram of the principal component analysis (see figure 15). A grouping of geographically related species can not be observed, rather the species with flowers that resemble one another in size and shape cluster together, whereby size seems to be the most important factor. The species with the biggest-sized inflorescences (namely *S. splendens*, *S. elegans*, *S. sclarea*, *S. argentea*, *S. hierosolymitana* and *S. miltiorrhiza*) form one of the three main clusters, with the two New World species (intensive red-colored hummingbird-pollinated *S. elegans* and *S. splendens*), the two European (carpenter bee-pollinated *S. argentea* and *S. sclarea)* and Eurasian / East Asian species (bee-pollinated *S. hierosolymitana* and *S. miltiorrhiza*) forming sister taxa, respectively. This six-species cluster is sister to a bigger clade that includes among others the closely related European *S. officinalis* and *S. lavandulifolia* as well as the closely related American *S. greggii*, *S. microphylla* and *S. x jamenensis*. Species with smaller-sized flowers like the European *S. jurisicii* and *S. verbenaca* and also the American *S. columbariae* and *S. tiliifolia* form the third cluster, that includes *Rosmarinus officinalis* (or *Salvia rosmarinus*) as well. The species *S. patens* (that has huge blue flowers, with a pronounced lip shape) forms an outgroup that is separated from all other species. Taken together, as mentioned before, there is no interrelation in flower size and shape and geographic distribution; with an appearance of European and New World *Salvia* species in all clusters of the dendrogram.

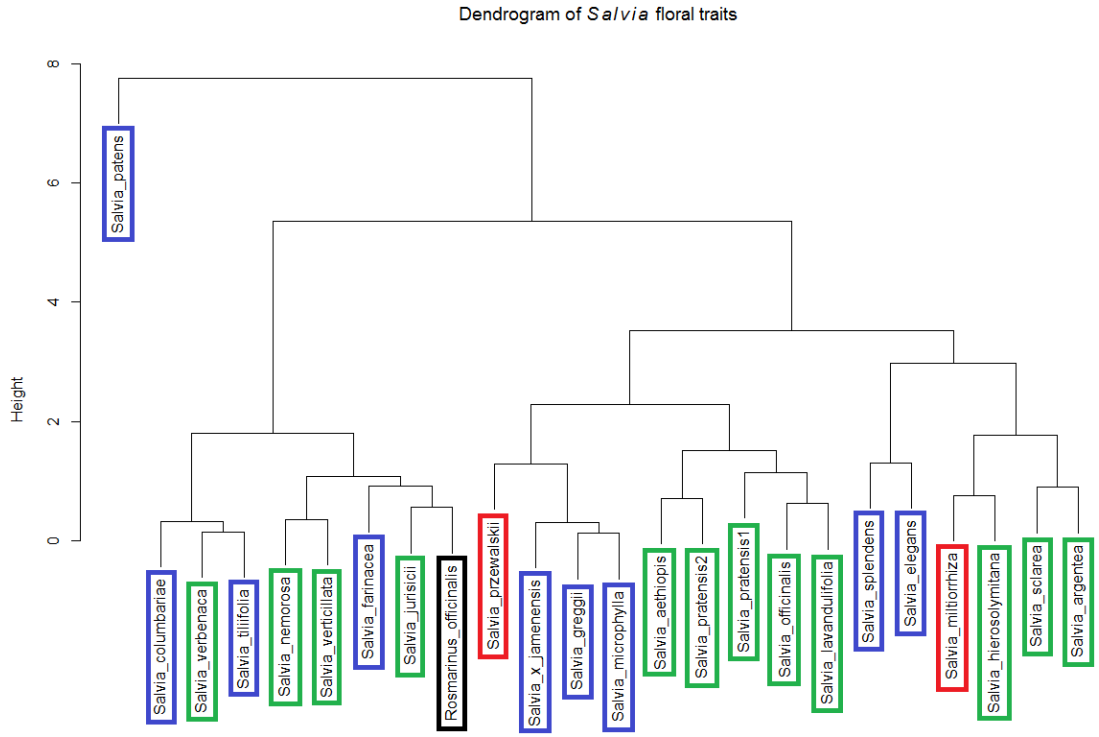


Figure 15: Dendrogram of the hierarchical clustering on principal components (HCPC) of *Salvia* floral trait measurements. European *Salvia* species are labeled in green, American *Salvia* species in blue and Asian *Salvia* species in red. *Rosmarinus officinalis* in black.

4.4 qPCR - On the tracks of the ABCDE-model in two *Salvia* species

The evaluation of developmental genes in developing *Salvia* flowers has been conducted in two ecological, geographical and morphological distinct species. *Salvia pratensis* is a purple colored bee-pollinated species from Middle Europe, whereas *Salvia elegans* represents an intense red colored hummingbird-pollinated species from Central America (see figure 13d and 13j). For both species the closed and opened floral tissue was dissected into five organs and all described qPCR experiments could be applied to all biological replicates, with the exception of the third replicate of opened flowers of *Salvia pratensis*, which did not yield satisfactorily result nd therefore was spared from the analysis.

4.4.1 For all classes of the ABCDE-model, except for D-class genes, representative genes could be evaluated satisfactorily in quantitative PCR experiments

For this study representative genes from the ABCDE classes of development have been investigated, namely APETALA1 (for A-class), GLOBOSA and DEFICIENS (for B-class), AGAMOUS (for C-class), SHATTERPROOF (for D-class) and SEPALLATA1 and SEPALLATA3 (for E-class). Furthermore, the putative auxin-based participants of flower development (as suggested in Xu *et al.*[57]), Auxin Response Factor 1 and 25, have been included into this study as well. Except for the D-class gene SHATTERPROOF, which did not yield evaluable results after qPCR, all the above mentioned genes could be amplified in qPCR reactions and subsequently evaluated as presented below.

4.4.2 In search for a house keeping gene

Three house keeping genes (see table 11) were tested in semi-qPCRs for their suitability as reference during flower development. The Ubiquitin gene amplification lead to double bands and the expression of Actin was unstable in different floral organs, solely the 18S gene appeared to be stable in all organs and in the different developmental stages. This observation of 18S in the semi-qPCR could be transferred to qPCR experiments, where this gene yielded stable c_q values in both species for both developmental stages and all tissues.

4.4.3 The special situation in the B-class gene DEFICIENS

During the process of primer design that was based on the coding sequences of *Salvia pratensis* and *Salvia elegans* DEFICIENS transcripts from different floral organs, the assumed single gene DEFICIENS turned out to have undergone gene duplication into two transcript variants (for a more detailed illustration of this result, see the phylogenetic trees in figure 16 and 41).

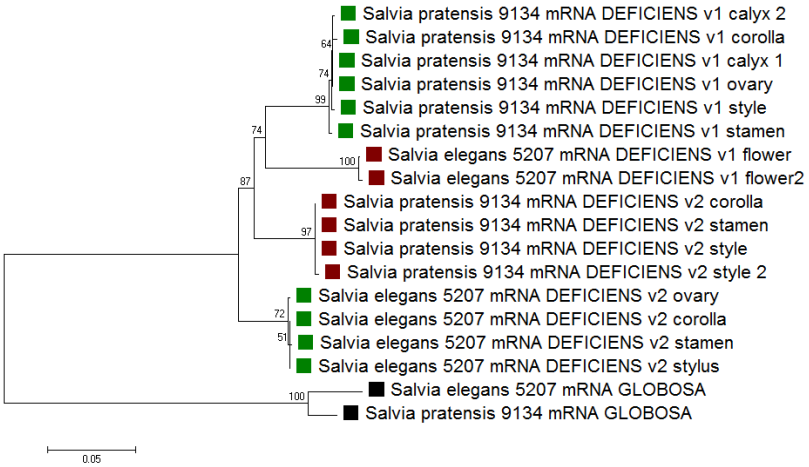


Figure 16: Depiction of the interrelation of the two different DEFICIENS transcripts. Related copies in *Salvia pratensis* and *Salvia elegans* are labeled numerically synonymous (variants: v1 and v2). Positive results for qPCR experiments are labeled green, negative in red. Transcripts of the other B-class gene GLOBOSA serve as outgroup. Neighbor-joining distant tree, bootstrap values (based on 1000 replicates) next to the branches.

Though very different (see figure 40), the variants are closer related to each other compared to the other B-class gene GLOBOSA (see figure 16). The distinct copies of DEFICIENS were circumscribed as DEFICIENS variant 1 and 2 for both *Salvia* species. Primers were designed for both copies in both species respectively to apply in qPCR (see table 11). Unfortunately for both *Salvia* species only one transcript respectively yielded robust qPCR results, while the other variant could not be further analyzed for different reasons (double peaks as well as highly varying technical replicates). The positive (green) and negative (red) results of the qPCR for the DEFICIENS are displayed in figure 16 as well. This figure clearly illustrates that the positive qPCR results of the different *Salvia* species originate from different copies of the DEFICIENS gene. The influence on the results are elaborated in the discussion part and should be kept in mind for the following qPCR results.

4.4.4 Evaluation of developmental gene expression overall confirms the ABCDE-model of flowering, with altering in expression comparing closed (developing) and opened (ready-to-pollinate) flowers

The expression pattern of all evaluated developmental genes are displayed for each floral organ (calyx, corolla, ovary, stamen and style) separately (figure 17 to figure 21 for *Salvia pratensis* and figure 22 to figure 26 for *Salvia elegans*). In these figures the expression of the genes in closed (still developing) and opened (ready to pollinate) are compared for all genes. Please note that the expression in organs slightly differed, for an overview with the focus on the genes (rather than the organ), please see appendix 8.2, and for a summary of both expression patterns see heatmap in figure 27.

4.4.5 Calyx – *Salvia pratensis* – The sepal envelope of the flower is enveloped by A and E-class expression

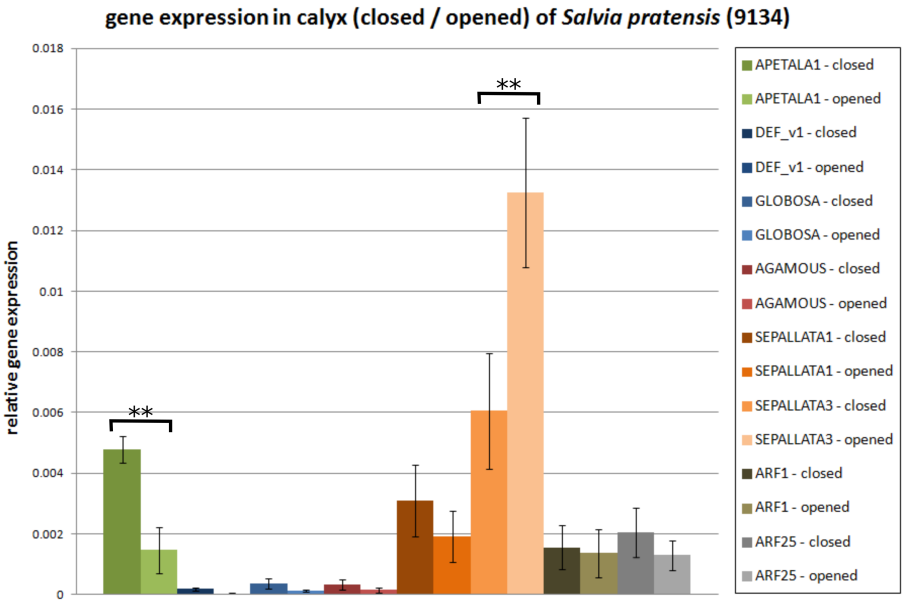


Figure 17: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ calyx in bee-pollinated European *Salvia pratensis*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

In the calyx of *Salvia pratensis*, the A-class gene APETALA1 shows a high expression. The level is significantly higher in closed flowers of *Salvia pratensis* compared to the expression of flowers that were opened and ready for pollination. B-class genes DEFICIENS variant 1 and GLOBOSA and C-class gene AGAMOUS are barely expressed. The two E-class genes SEPALLATA1 and SEPALLATA3 show the third highest (after AP1) and highest expression in this tissue during development, respectively. For the E-class SEP1 there is no significant difference in expression after opening, whereas SEP3 shows the highest expression in closed flowers and an even stronger expression in opened flowers, suggesting developmental activity after the opening of the flower of *Salvia pratensis*. The level of expression of Auxin Response Factor 1 and Auxin Response Factor 25 is comparable to the that of SEP1, and there is no significant variance in expression in the different developmental stages of flowering.

4.4.6 Corolla – *Salvia pratensis* – The rise of the Bs to attract the bees

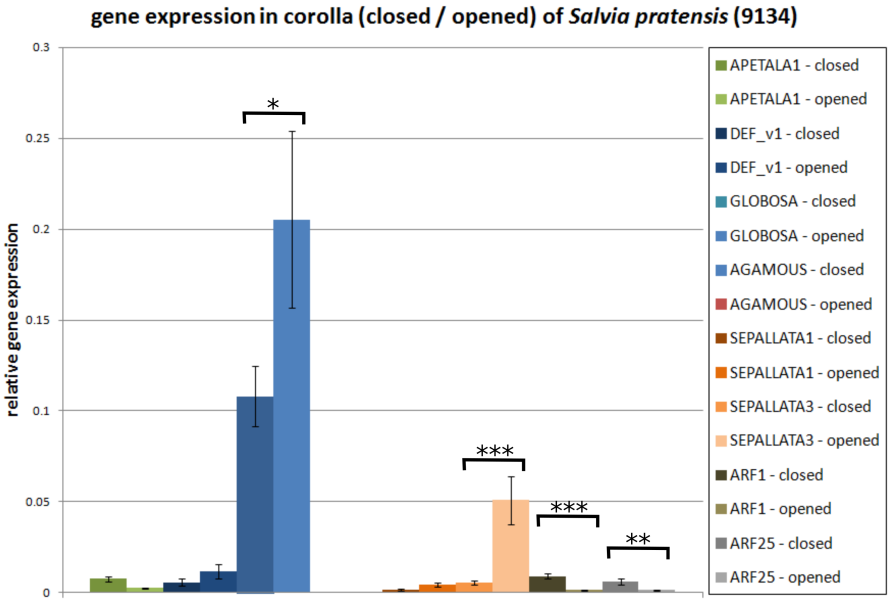


Figure 18: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ corolla in bee-pollinated European *Salvia pratensis*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

In contrast to the expression in the sepal tissue, there is an extraordinary rise of expression of B-class genes, especially in GLOBOSA in developing tissue. However, this strong expression is even elevated for the corolla tissue after the flower has opened. The expression level of A-class gene APETALA1 and B-class gene DEFICIENS variant 1 is similar, with a significant drop of AP1 after opening of the flower. The expression of DEFICIENS variant 1 does not show a significant difference in the stages of flowering. C-class gene AGAMOUS is not present at all. The E-class gene SEPALLATA1 has an increase of expression that is on a low level, however, still significant comparing the different stages of development. The elevation of SEPALLATA3 in opened flowers of *Salvia pratensis* is remarkable and shows the second highest peak of expression after GLOBOSA. The Auxin Response Factors 1 and 25 genes, though not so highly expressed, display a significant drop of expression after opening of the flower, indicating functions in corolla growth during development.

4.4.7 Ovary – *Salvia pratensis* – From A to E - ambiguity of expressitivity

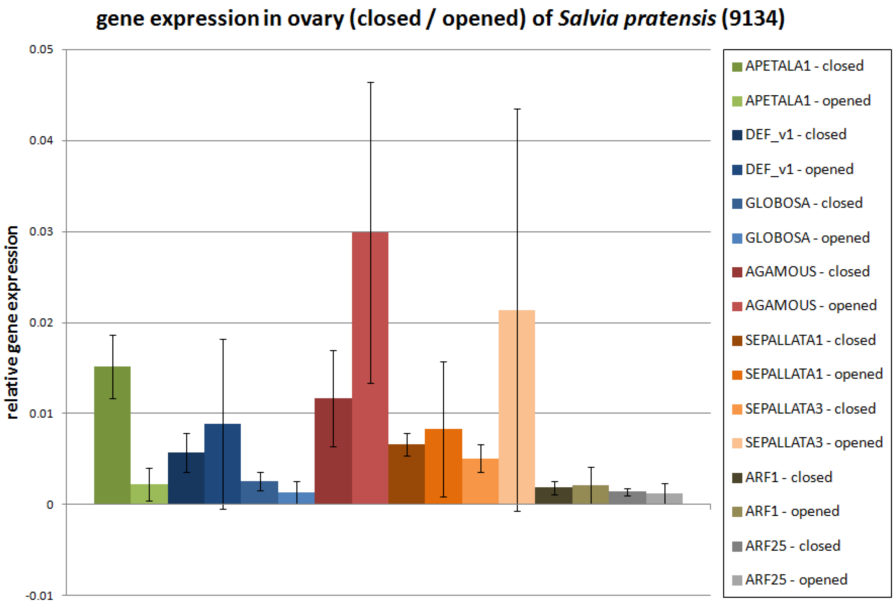


Figure 19: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ ovary in bee-pollinated European *Salvia pratensis*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

The overall expression in ovaries of *Salvia pratensis* as illustrated in figure 19 does not show a clear pattern, since all of the evaluated genes are expressed to some extant and the very high error bars prohibit well-founded statements. In general there is no gene that stands out in expression (like APETALA1 or SEPALLATA in calyx or GLOBOSA in corolla, as mentioned before). There is a surprising expression of the A-class gene APETALA1 in developing ovaries, that shows a drop of expression after the flower has opened. Comparing the B-class genes, the expression of DEFICIENS variant 1 is higher than GLOBOSA, however, there is no significant difference in expression in the developing stages, neither in DEF variant 1, nor in GLO. Expression of C-class gene AGAMOUS seems to be elevated, after opening of the flower, but the increase is not significant. Expression level of SEPALLATA1 is comparable to DEF variant 1, with a slight increase in opened flowers (that is not significant though). Expression of SEPALLATA3 is notably , because the expression in closed flowers can be compared to SEP1 or DEF v1, however, there is an strongly induced expression in the opened flowers. Expression levels of the Auxin Responsive Factors 1 and 25 is stable.

4.4.8 Stamen – *Salvia pratensis* – The construction of the staminal lever is facilitated by B and C-class genes

In the male reproductive organs of *Salvia pratensis* a high expression of the B-class gene GLOBOSA and the C-class gene AGAMOUS can be observed. GLOBOSA expression rises significantly after opening of the flower, just as in the petals. For the C-class gene AGAMOUS there is no significant difference in expression for the different developmental stages. This is true for the other B-class gene DEFICIENS-variant1 as well; this gene has overall the third highest expression in this floral organ. A-class gene APETALA1 and E-class SEPALLATA1 have overall low expressions and there is no significant difference in the developmental stages. SEPALLATA3, however, displays a significant increase of expression after opening of the *Salvia pratensis* flower. The two Auxin Response Factors show low (compared to the MIKC-type MADS gene expressions) but still significant differences in the developmental stages. In the closed flowers ARF1 and ARF25 show higher expressions, with a drop of expression after the flower has opened.

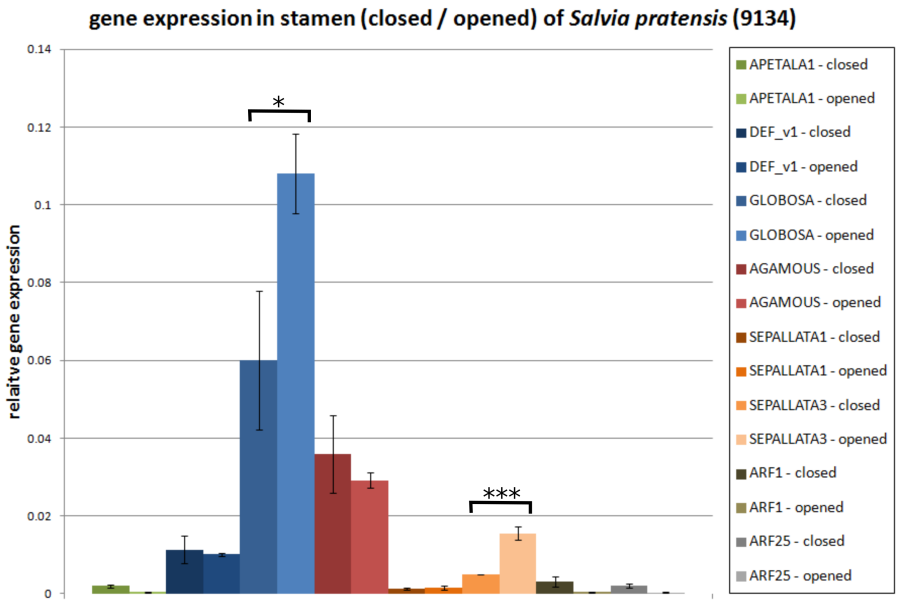


Figure 20: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ stamen in bee-pollinated European *Salvia pratensis*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

4.4.9 Style – *Salvia pratensis* – The road for the pollen tube is paved by C and E-class activity and assisted by ARF1

The highest expression in this organ can be observed for the C-class gene AGAMOUS, the B-class gene DEFICIENS variant 1 and the Auxin Response Factor 1. The gene expressions of A-class gene APETALA1, B-class gene GLOBOSA, C-class gene AGAMOUS and ARF1 drop significantly after the flowers of *Salvia pratensis* have opened. B-class gene DEFICIENS variant 1 and E-class gene SEPALLATA1 expressions drop as well, however, the differences are not significant. Solely the expression of SEPALLATA3 shows a significant elevation in the style tissue after the flower has opened. In opened flowers the expression of C-class AGAMOUS and E-class SEPALLATA3 is on a similar level.

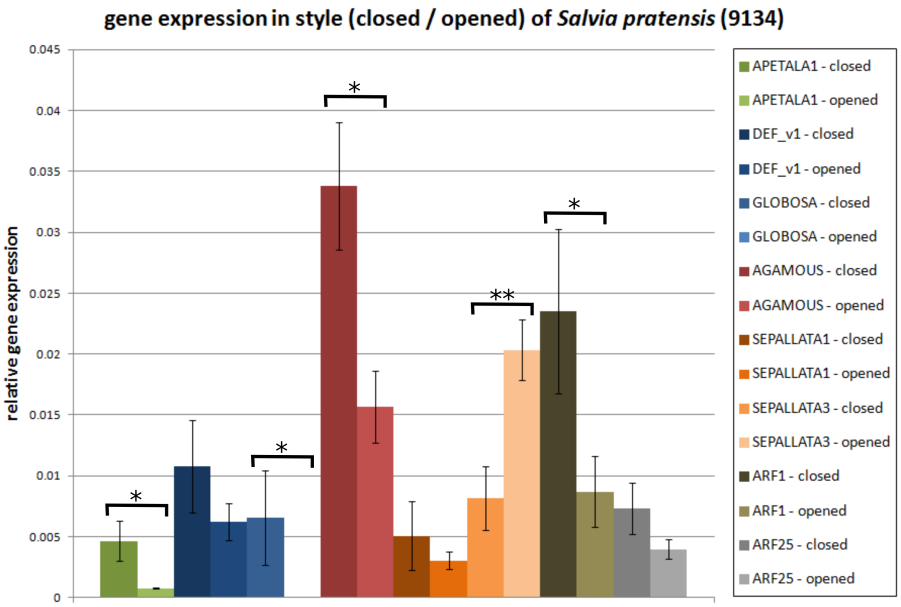


Figure 21: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ style in bee-pollinated European *Salvia pratensis*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

4.4.10 Calyx – *Salvia elegans* – A all alone – significant higher expression of APETALA1 over all other genes

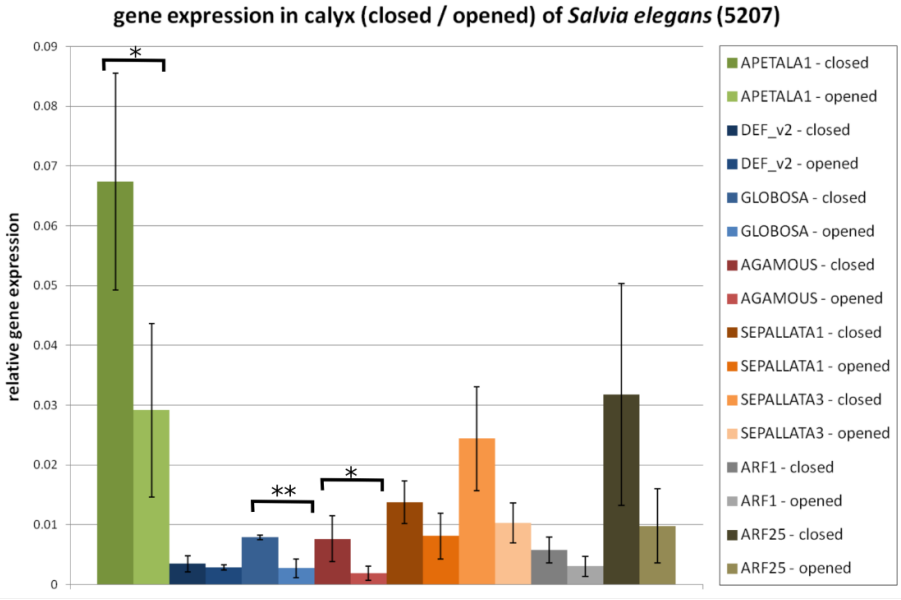


Figure 22: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ calyx in hummingbird-pollinated American *Salvia elegans*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

In developing sepal tissue of the closed flowers of *Salvia elegans* the A-class gene APETALA1 shows the highest expression; significantly higher compared to the expression of the same gene in opened flowers and significantly higher than all other genes that were part of the study. The drop in expression in different developing stages of flowering is significant, but still AP1 expression in opened flowers is higher than B- and C-class genes, that are only expressed on a low level. E-class gene expression seems to be higher, but there is no significant elevated expression of SEPALLATA1 compared to GLOBOSA or AGAMOUS. SEPALLATA3 again is higher expressed in developing tissue, with a drop after the flower has opened, a result that is contrary compared to *Salvia pratensis*, where SEP3 was strongly induced after opening of the flower. The expression of the Auxin Response Factor 1 is rather low in both opened and closed flowers, but the expression of ARF25 is the second highest after AP1 in developing tissue. The drop of expression for this gene is strong, however, not completely statistically significant.

4.4.11 Corolla – *Salvia elegans* – B for birds – the shaping for hummingbird-attraction

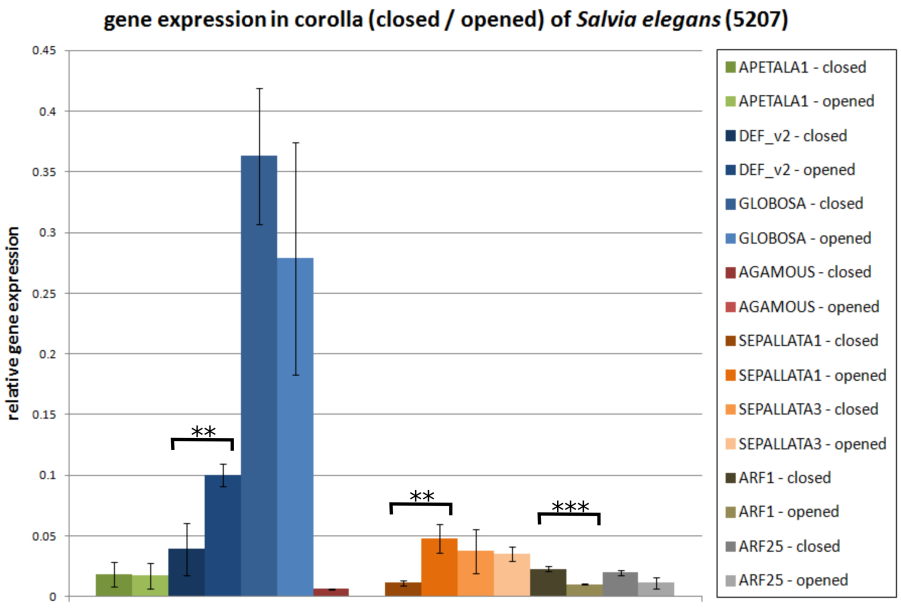


Figure 23: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ corolla in hummingbird-pollinated American *Salvia elegans*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

In corolla tissue of hummingbird-pollinated *Salvia elegans* the rise of B-class gene expression is remarkable, likewise in *Salvia pratensis* GLOBOSA shows a strong activation, however, unlike in *Salvia pratensis* there is no significant increase in expression of GLOBOSA but rather (not significant supported) a drop after opening of flower. The standalone of GLOBOSA expression that has been described in *Salvia pratensis* is not that prominent in *Salvia elegans*, since in this species the expression of DEFICIENS variant 2 is elevated as well, with a significant rise of expression after opening of the flower, suggesting post-opening controlled processes. The level of expression does not change significantly in the A-class gene APETALA1 and the E-class gene SEPALLATA3. In the E-class gene SEPALLATA1 there is a significant elevation of expression after the flower has opened its corolla. For both Auxin Response Factors the expression is overall lower than the B- and E-class genes and comparable to the A-class gene. In opened flowers a significant drop of ARF1 expression compared to developing flowers can be observed.

4.4.12 Ovary – *Salvia elegans* – Ovary development is strongly controlled by C-class

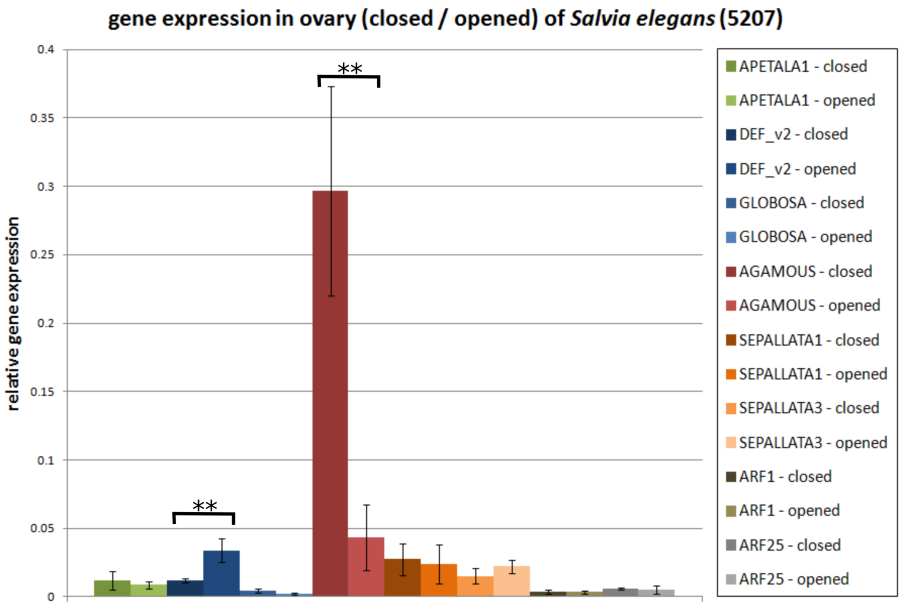


Figure 24: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ ovary in hummingbird-pollinated American *Salvia elegans*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

The overall expression pattern is superimposed by the very strong expression of the C-class gene AGAMOUS in the closed flowers of *Salvia elegans*. After opening of the flower this expression drops significantly, but remains compared to the other genes relatively high. The expression of the A-class gene APETALA1, the B-class gene GLOBOSA, the E-class genes SEPALLATA1 and SEPALLATA3, as well as the Auxin Response Factors 1 and 25 do not differ significantly in this organ for closed and opened flowers. The expression of the B-class gene DEFICIENS variant 2 is elevated after opening of the flower with an expression level that is comparable to the expression of the C-class gene AGAMOUS.

4.4.13 Stamen – *Salvia elegans* – B and C-class domination in male reproductive organs

The stamen of the hummingbird-pollinated *Salvia elegans* show high expressions of the B-class genes DEFICIENS variant 2 and GLOBOSA, and C-class gene AGAMOUS in closed flowers. This expression drops significantly in DEF variant 2 and AGAMOUS, compared to an increase of expression in GLOBOSA, which is, however, not significantly supported. A-class gene APETALA1 and the Auxin Response Factors 1 and 25 show low stable expressions, with no changes in the different stages of development. The E-class genes SEPALLATA1 and SEPALLATA3 display an elevation of expression in opened flowers of *Salvia elegans*, with an expression level that is comparable to the expression of B-class DEFICIENS v2 and C-class AGAMOUS.

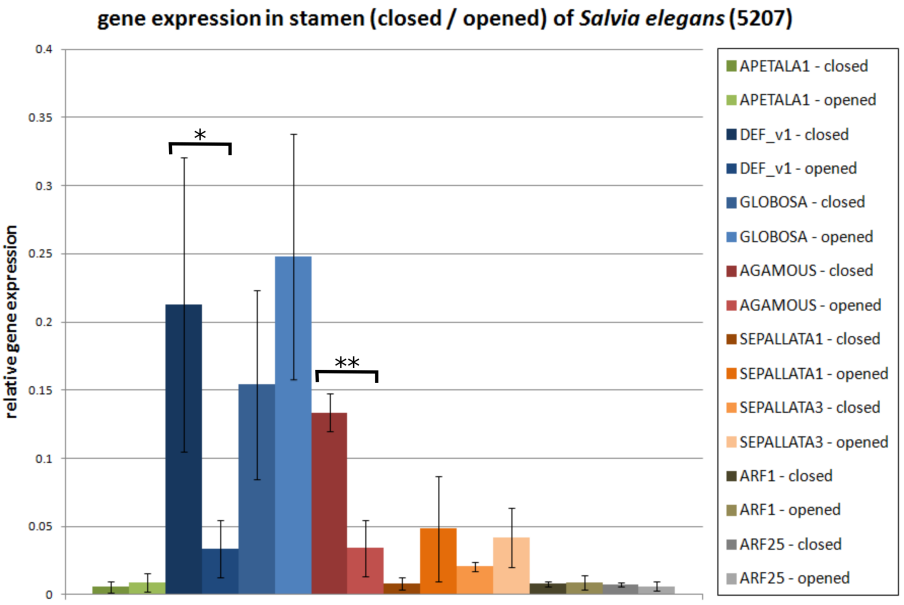


Figure 25: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ stamen in hummingbird-pollinated American *Salvia elegans*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

4.4.14 Style – *Salvia elegans* – Surprising support of B-class to the pollen tube road additional to the C-class and ARF1 facilitators

In the style tissue of closed flowers of *Salvia elegans* the B-class gene DEFICIENS variant 2, the C-class gene AGAMOUS and the Auxin Response Factor 1 show the highest expressions, dominated clearly by the C-class gene. The expression drops significantly in DEF variant 2 and AGAMOUS, but is stable in ARF1 after the flower has opened. The expression of the A-class gene APETALA1, the B-class gene GLOBOSA, the E-class gene SEPALLATA3 and ARF25 is stable in the different developmental stages as well, but with an overall lower expression compared to DEF variant 2, AGAMOUS and ARF1. The expression of E-class gene SEPALLATA1 rises significantly after the flower of *Salvia elegans* has opened.

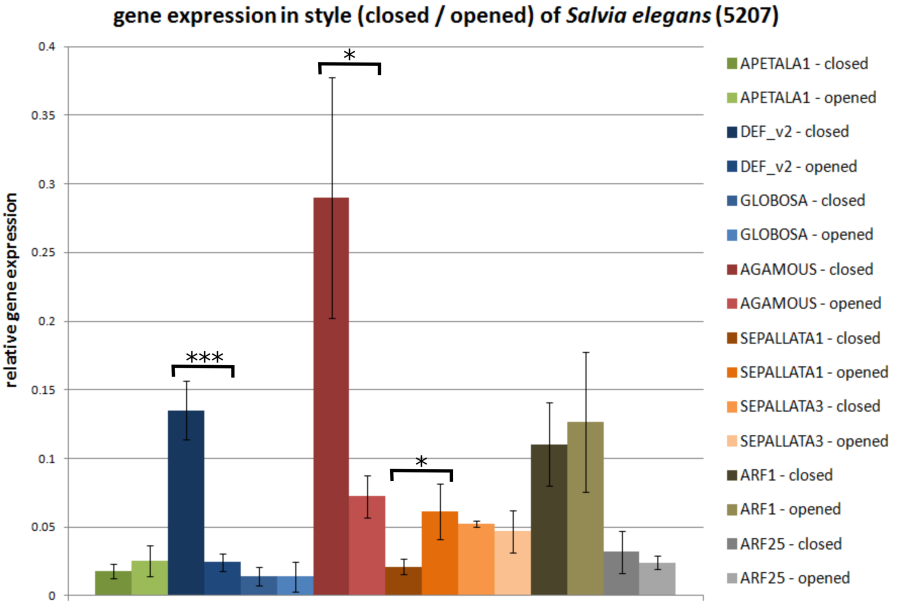


Figure 26: Comparative gene expressions in closed (flower buds; darker shade of color) and opened (ready-to-pollinate; lighter shade of color) flowers in the floral organ style in hummingbird-pollinated American *Salvia elegans*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

4.4.15 Overview and comparison in gene expression of *Salvia pratensis* and *Salvia elegans*

For better illustration, all above described results have been plotted into two heatmaps, that display all qPCR results (all genes, both developmental stages) for *Salvia pratensis* and *Salvia elegans*, respectively (see figure 27). Since the results for the individual species have been described above, in this paragraph the commonalities and differences of expression in the European and American *Salvia* are highlighted. The expression pattern of the B-class gene GLOBOSA is very similar in *Salvia pratensis* and *Salvia elegans*, with an almost exclusive expression in corolla and stamen tissue, whereas in the other three organs the gene is close to absent. C-class gene AGAMOUS expression is conserved as well, since this gene is not abundant in calyx and corolla in both investigated species, while there is a strong presence in ovary, stamen and style. The Auxin Response Factor 1 pattern is quite similar, with the main expression in the style of *S. pratensis* and *S. elegans*, respectively. However, in the American *Salvia* species the level of ARF1 is elevated after the flower has opened, whereas the expression in the closed flower is more dominant in the European *Salvia* species. In both species, the expression is enhanced in developing corolla and overall low expressions in ovary and stamen. The A-class gene APETALA1 is quite similar as well, because the expression is low in stamen and style and the highest in the developing calyx tissue. In *S. pratensis* there is an exception, that is due to an unexpected high expression in developing ovary tissue. The other representative of the Auxin Response Factor group, ARF25, shows low expressions in ovary and stamen as well. The highest expression can be observed in calyx and style tissue, additionally this gene is present in developing petals. The second B-class gene DEFICIENS is quite variable in *S. pratensis*, with an elevation of expression in corolla and ovary and a decrease of expression in stamen and style after the opening of the flower. This elevation and demotion in the same organs can be observed in *S. elegans* as well, however, in this species the differences are more pronounced.

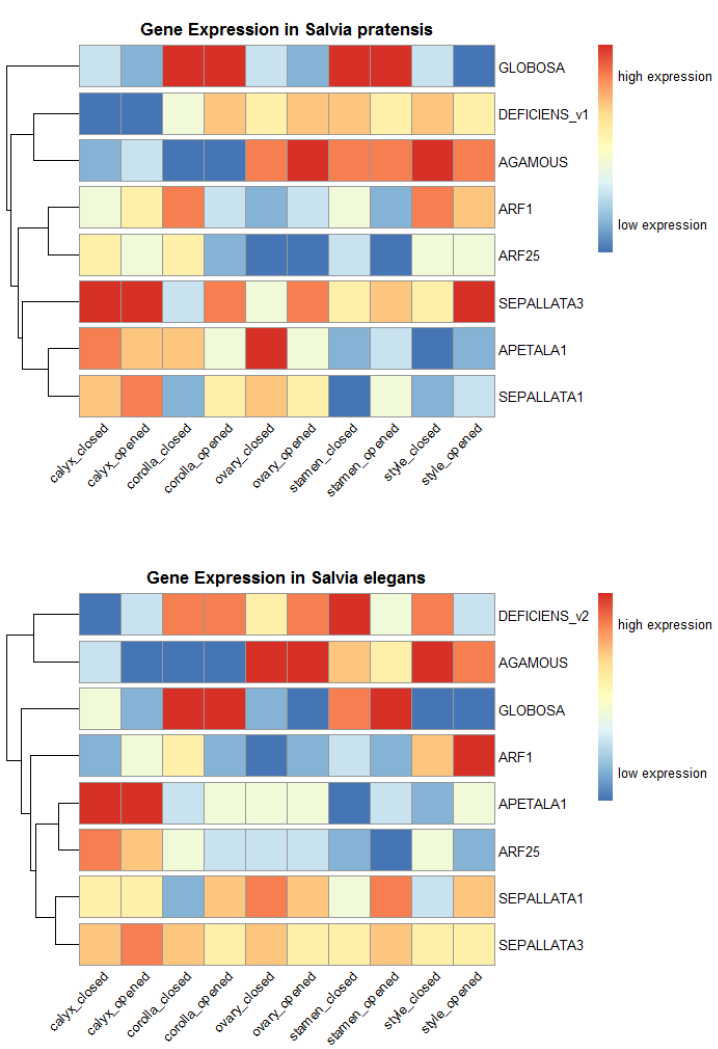


Figure 27: Gene expression overviews illustrated by heatmaps with expression patterns of *Salvia pratensis* (upper) and *Salvia elegans* (lower). Red color indicates strong gene expression, blue color indicates low gene expression.

4.4.16 Expression patterns of Auxin Response Factors 1 and 25

The gene expression patterns of two candidate Auxin Response Factors have been displayed for each floral organ in the two different species *Salvia pratensis* and *Salvia elegans* (figure 17 to figure 21 for *S. pratensis* and figure 22 to figure 26 for *S. elegans*). For a better illustration concerning the spatial temporary expression of these two transcription factors, the expressions of ARF1 and ARF25 in closed and opened flowers of the two species are displayed in figure 28 and figure 29.
In *Salvia pratensis* there is no change in expression for the floral organs calyx and ovary for both Auxin Response Factors. Compared to that, in corolla and stamen there is an initial high expression for both of the genes during development, with a significant drop of expression after the organs have formed and the flowers have opened. The highest expression though can be observed in the style, especially for ARF1, which shows the highest expression by far, followed by a significant drop, resulting in an expression activity that is still higher compared to expression in closed corolla for instance. ARF25 is expressed in the style on a high level as well, with a subsequent decrease of expression in the opened flower.
Looking at the ARF expression in *Salvia elegans* a slightly different pattern can be observed. The ARF1 has stable and low expressions in calyx, ovary and stamens. The expression in the corolla tissue is somewhat higher and there is a significant drop after the flowers are opened in ARF1. In the style the expression of ARF1 is strongly elevated compared to all other organs. The high expression does not drop after the flower has opened and remains almost at the same level. For ARF25 a significant drop of expression during development can be observed in corolla tissue, for all other organs (ovary, stamen and style) there is no significant difference in expression. However, likewise in ARF1, the expression in the style is the highest, but there is no drop of expression in this tissue after the flower has opened. The second highest expression (after style) is displayed in the calyx tissue, with a down regulation of expression that is close to significant.

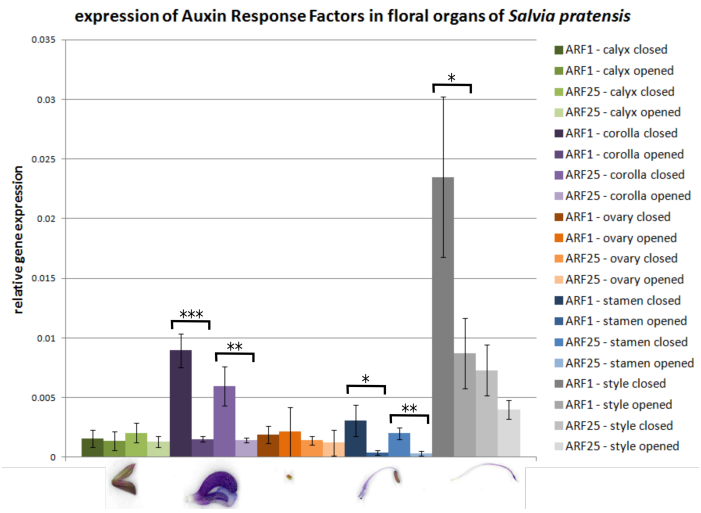


Figure 28: Comparison of gene expression of Auxin response factor 1 and 25, in closed and opened floral organs of *Salvia pratensis*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

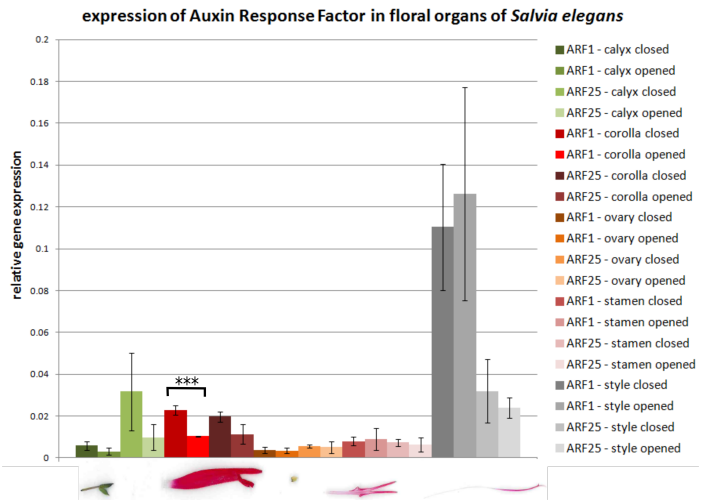


Figure 29: Comparison of gene expression of Auxin response factor 1 and 25, in closed and opened floral organs of *Salvia elegans*. The expression is relative to the housekeeping gene 18S. Error bars indicate standard deviation of the three biological replicates. Asterisks indicate t-test significance.

4.5 Auxin Response Factor 1 analysis in *Salvia*

4.5.1 Sequence evaluation reveals conservation of the miRNA160 target site in ARF1

PCR with primers for the second exon of the Auxin Response Factor 1 yielded intense bands on an agarose gel (see figure 30), with fragments lengths around 900 basepairs. The purified PCR products could be used for sequencing directly and resulted in sequences of good quality for most of the evaluated species. Trimming of the sequences at the 5' and 3'-end, resulted in sequence lengths between 743 basepairs (for *Salvia patens*) and 769 basepairs (for *Salvia dominica*). These differences in length arise from varying lengths of the short second intron of the Auxin Response Factor 1, that was part of the amplified fragment. The homologue of smARF1 in *Arabidopsis thaliana*, atARF16, is under post-transcriptional control of miRNA160 (as elaborated in the introduction). The conserved target sequence for miRNA160 has been shown in the ARF1 sequence of *Salvia miltiorrhiza* as well [57]. The amplified ARF1 fragments for the *Salvia* species of this study included the region for the putative miRNA160 target site. Multiple sequence alignment analysis of the available sequence data using MEGA7 lead to a dataset of 46 different species (in total 75 different accessions and thus including multiple individuals for some species) with an aligned length of 795 nucleotides. For all evaluated species (including the genera *Salvia*, *Rosmarinus*, *Perovskia*, *Dracocephalum*, *Mentha* and *Sesamum*) the target sequences site for atmiRNA160 (which is from *Arabidopsis thaliana*, a very distant related rosid) is conserved by 20 of 21 nucleotides (see figure 31 and 32), for all asterids the conservation is 100%. This results in parsimony-uninformative sites, however, pointing to an evolutionary highly conserved control mechanism of ARF1 by miRNA160.

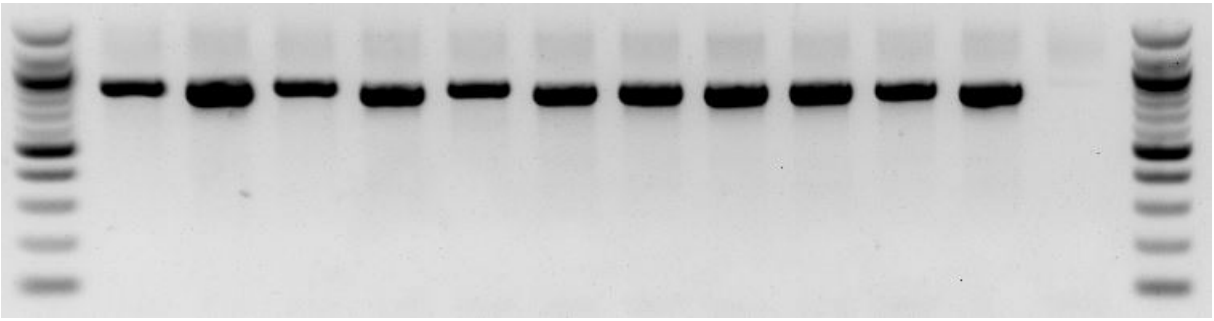


Figure 30: Amplification of the second exon of the ARF1 gene for different *Salvia* species by PCR from genomic DNA and evaluation by AGE. 100 bp DNA ladder.

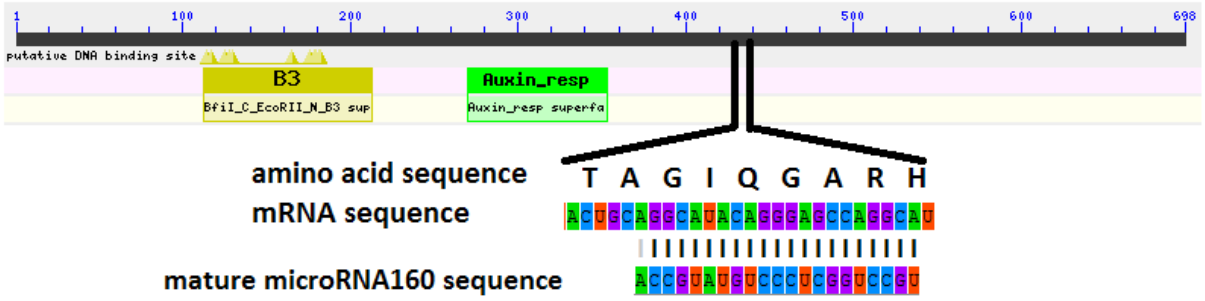


Figure 31: BLAST result of the smARF1 protein, with the conserved B3-DNA-binding-domain (yellow) and the conserved region of auxin-responsive transcription factors (green). The mRNA sequence and the binding sequence of the atmiRNA160, as well as their matching nucleotides are displayed.

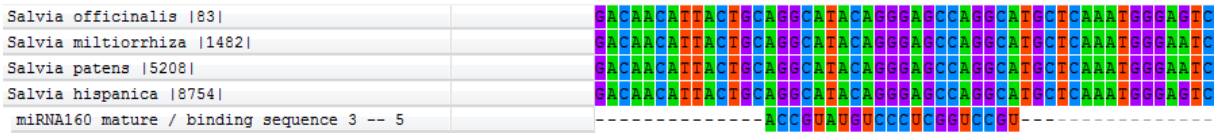


Figure 32: Examples of *Salvia* sequences and the miRNA160 binding sequence.

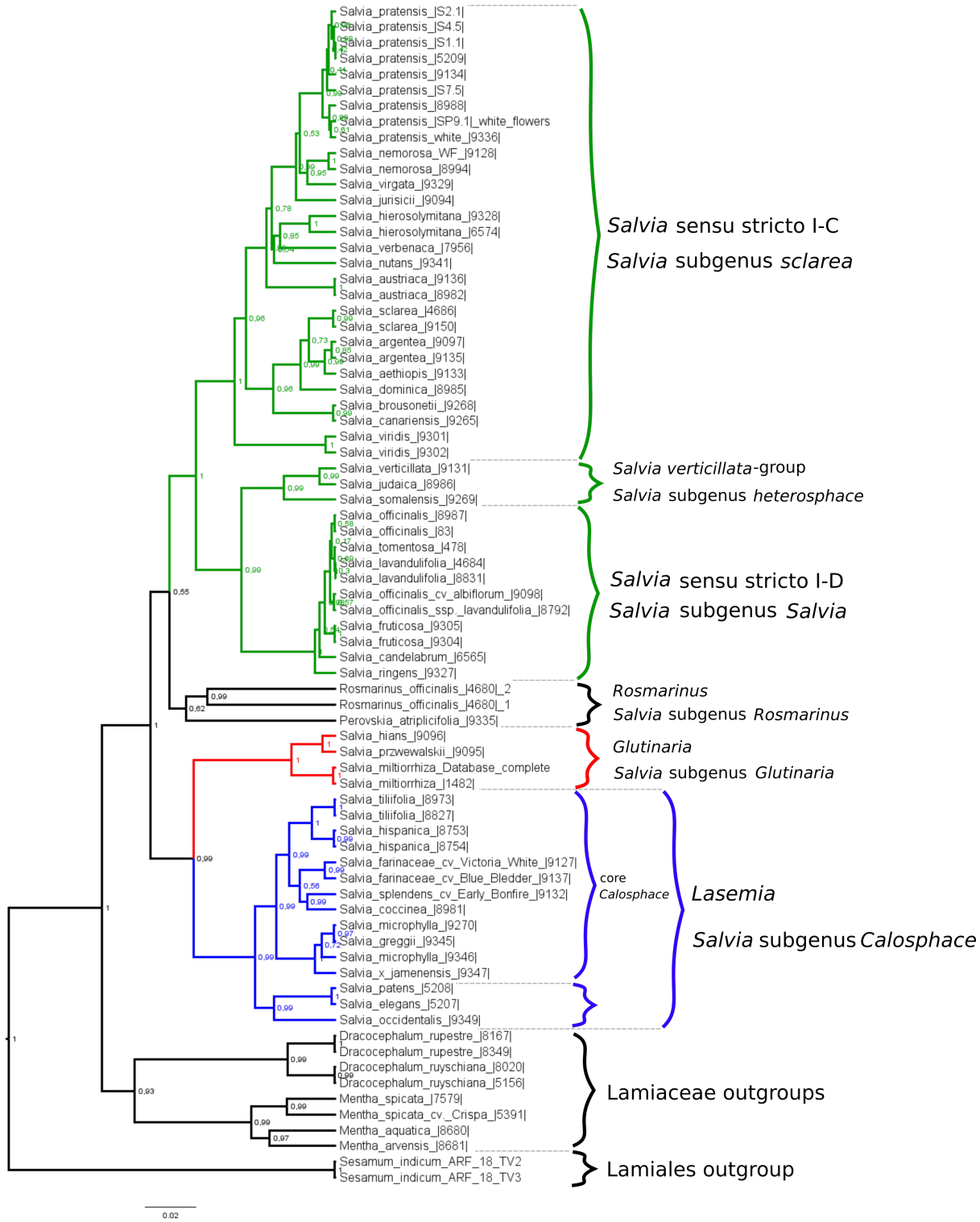


Figure 33: ARF1 gene based phylogenetic tree. European *Salvia* species labeled green, New World *Salvia* species labeled blue and East Asian *Salvia* species red. Lamiaceae outgroups are labeled in black. Bayesian inference was used and posterior probability values are labelled next to the node. The two current classifications of genus *Salvia* are given for each clade. The upper circumscription is by Will and Claßen-Bockhoff (2017) [36] and the lower one by Drew *et al.* (2017) [35], respectively.

4.5.2 ARF1 phylogeny endorses the topology of trait-unrelated barcoding marker based phylogenetic trees

The multiple sequence alignment dataset of the ARF1 fragment was evaluated by BEAUti and BEAST and illustrated by Figtree. Apart from the Lamiales outgroup (*Sesamum indicum*) and the Lamiaceae outgroups (*Dracocephalum* and *Mentha*) four clusters can be observed, that represent the geographic regions of European, the New World and East Asian *Salvia*, and the *Rosmarinus*-cluster.
Using this marker, the genus *Salvia* showed its paraphyletic nature again, with *Rosmarinus officinalis* and *Perovskia atriplicifolia* disrupting the monophyly (see figure 33). New World and East Asian *Salvia* form sister clades, as well as the European *Salvia* and the *Rosmarinus* clade. The separation of these two clusters is strongly statistically supported, which applies to the separation of the Asian and the American, as well as and also within the American Cluster. The rough separation of species that belong to the core *Calosphace* (e.g. *Salvia greggii* or *Salvia hispanica*) and species that are excluded from this clade, like *Salvia elegans*, *Salvia occidentalis* (both belonging to the Uliginosae clade) and *Salvia patens* (belonging to the Hastatae clade) can be shown with significantly values. Though involved in plant sexual organ growth (shown in 4.3.14) and thus being one of the triggers of reproductive isolation, the pollination syndrome can not be estimate by this gene, because the phylogenetic mixing of melittophilous and ornithophilous is evident in ARF1 based phylogenies as well. Insect-pollinated *Salvia farinacea* is sister to a clade that comprises the two hummingbird-pollinated species *Salvia splendens* and *Salvia coccinea*. The same thing is true for *Salvia elegans* (ornithophilous) and *Salvia occidentalis* (melittophilous), both belonging to the Uliginosae clade. The small dataset of East Asian species, shows the close and supported relation of *Salvia przewalskii* and *Salvia hians*, which are sister to a *Salvia miltiorrhiza* clade that consists of two sequences, one obtained from the herbal genome database and one from the *Salvia miltiorrhiza* reference plant from the botanical garden of the KIT. While the significance values are high and strongly support the provided gene-tree for New World and East Asian *Salvia* species, this is not that evident for the European *Salvia* / *Rosmarinus* clades. The separation of the New World / East Asian-clade is supported significantly, however, within the European distributed species, including *Rosmarinus*, there are statistical uncertainties. The node that represents the genetic ancestor of *Rosmarinus* and *Salvia officinalis* is supported by 48%, showing a non supported differentiation of *Rosmarinus* from all other European *Salvia* species. Regarding the European (green labeled) *Salvia* species, a strongly supported general split of the subgenus '*Sclarea*' / I-C clade and the combined subgenus '*Salvia*' / I-D clade, subgenus '*Salvia*' / I-A clade, subgenus '*Heterosphace*' / *Salvia verticillata*-clade can be observed, as well as a supported separation of I-D and I-A clade / *S. verticillata* clade, however, an unexpected sister taxa gathering of I-A (*Salvia somalensis*) and *S. verticillata* (*S. verticillata* and *Salvia judaica*) can been seen in the ARF1 gene tree as well.

4.6 Phylogenetics of the B-class gene GLOBOSA in *Salvia*

4.6.1 Designed primers for the GLOBOSA gene yield strong bands

The amplification of the GLOBOSA gene from start to stop codon in a PCR using the primers from table 10 yielded fragment lengths between 1,5 kb and 2,0 kb, which have been visualized in an agarose gel (figure 34). The primers lead to successful amplification in all investigated *Salvia* species and additionally in other genera of the Lamiaceae like *Rosmarinus*, *Perovskia*, *Ocimum*, *Mentha*, *Lavandula* and *Dracocephalum*.

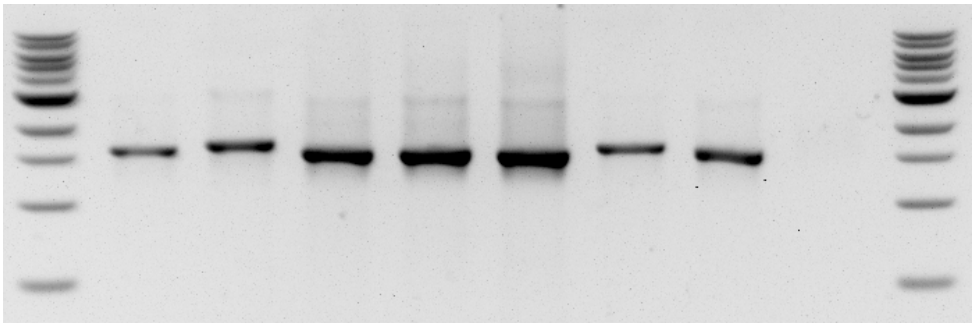


Figure 34: Amplification of the complete B-class GLOBOSA gene for different *Salvia* species by PCR from genomic DNA and evaluation by AGE. 1 kb DNA ladder. Fragment size of the GLOBOSA gene is approximately 1700 bp.

Direct sequencing of GLOBOSA PCR products resulted in unusable data, since the chromatogram of the sequenced data showed double peaks with base calls (Q-values) lower than the significance value of Q20. In order to avoid inconclusive double peaks, the PCR products were cloned as described in method part 3.6 and the plasmids of positive colonies were used for sequencing analysis. This approach delivered high-qualitive sequences that could be used for downstream investigations.

For GLOBOSA in total 38 species were sequenced, of which 19 were *Salvia* individuals from Europe (including eponymic species *Salvia offcinalis* from the subgenus '*Salvia*' / I-D cluster, but also several individuals from the subgenus '*Sclarea*' / I-C cluster, including *Salvia pratensis*), ten from the New World (including bee-pollinated *Salvia farinacea* and hummingbird-pollinated *Salvia patens* and *Salvia elegans*) and three from East Asia (including *Salvia przewalskii*). The remaining six investigated species are from other Lamiaceae genera (*Rosmarinus officinalis*, *Perovskia atriplicifolia*, *Lavandula angustifolia*, *Ocimum tenuiflorum* and two species from the genus *Dracocephalum*).

4.6.2 Conserved differences of exon length in geographically separated *Salvia* groups

The complete (exons and introns) length of the GLOBOSA *Salvia* sequences ranges from 1638 nucleotides in one copy of American *Salvia greggii* to 1790 basepairs in a copy of East Asian *Salvia glutinosa*. The total length of the Lamiaceae outgroups is slightly larger, with for example 1851 basepairs in *Perovskia atriplicifolia* and 1883 basepairs in *Dracocephalum rupestre*.

The combined length of the seven exons ranges from 633 basepairs in copies of *Salvia elegans* to 648 basepairs in different European *Salvia* species (see figure 35).

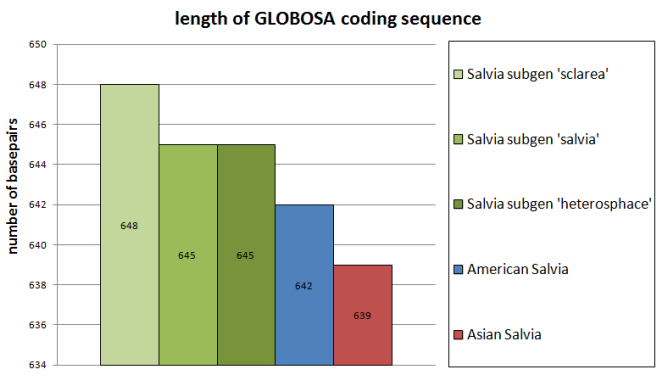


Figure 35: Comparison of the coding region in the GLOBOSA gene of *Salvia* species with different geographical background revealed conserved exon lengths. New World *Salvia* species are labeled in blue, East Asian *Salvia* in red, and different European *Salvia* subgenera in different shades of green.

When looking at the length of the coding regions only, there is pattern that supports the geographical background of the genus. The translation to amino acid sequence was done in-silico and yielded the primary structure of the proteins.

European *Salvia* from the subgenus '*Sclarea*' / I-C cluster (*S. aethiopis*, *S. argentea*, *S. austriaca*, *S. hierosolymitana*, *S. jurisicii*, *S. nemorosa*, *S. nutans*, *S. pratensis*, *S. sclarea*, *S. verbenaca*, *S. virgata*) have a combined exon length of 648 basepairs (translated to 216 amino acids).

European *Salvia* from the subgenus '*Salvia*' / I-D cluster (*Salvia canariensis*, *Salvia candelabrum*, *Salvia judaica*, *Salvia lavandulifolia*, *Salvia officinalis*, *Salvia ringens*, *Salvia tomentosa*, *Salvia verticillata*) have a coding length of 645 basepairs (translated to 215 amino acids). This pattern can be seen in *Rosmarinus officinalis* and *Perovskia atriplicifolia* as well.

All evaluated American *Salvia* species have a coding length of 642 basepairs (translated to 214 amino acids), with exception of *Salvia elegans*, for this species two copies of 633 basepairs (translated to 211 amino acids) have been found. The outgroups from the genera *Dracocephalum* and *Ocimum* show the same length. The length of 639 basepairs (translated to 213 amino acids) can be observed in all evaluated East Asian species, likewise in *Lavandula angustifolia* and *Sesamum indicum* (which was obtained from the ncbi database).

4.6.3 Variations in coding length affect the C-terminal domain of the GLOBOSA protein

BLASTing the primary amino acid sequence revealed that the characteristic MIKC-type MADS protein domains of the flowering ABC genes (see introduction) could be traced in all *Salvia* species as well. In all amino acid sequences the first 60 amino acids refer to the MADS-domain (which has DNA-binding functions), the downstream amino acids at sites 61 to 83 refer to the intervening domain (responsible for dimerisation of MIKC-type MADS proteins), the amino acids at sites 84 to 164 in the alignment can be assigned to the K-box domain (important for tetramerisation). The function of the C-terminal domain is mainly unknown; in *Salvia*, the amino acids from sites 165 to 211/216 refer to this domain.

The above described differences in coding sequence length (and visualized in figure 35) affect the number of amino acids in the C-terminal domains in the primary protein structure. In the amino acid multiple sequence alignment at total position 188 (which is located in the C-terminal domain) there is highly conserved valine (V) which is present in all investigated species.

The preceding glycine (G) at position 187 is present in European and East Asian *Salvia* species as well as in the outgroups (substituted to alanine (A) in *Rosmarinus officinalis* and threonine (T) in *Ocimum*). This glycine is absent, or more precisely, deleted in all American *Salvia* species, which explains partly the reduced primary amino acid sequence length in the American *Salvia* species.

Starting at site 193 of the alignment there is an AAQ-motif, that is only present in European subgenus '*Sclarea*' / I-C cluster, partly deleted to -AQ in European subgenus '*Salvia*' / I-D cluster and all American species and completely deleted in all East Asian species, which explains the differences in length between subgenus '*Sclarea*' and subgenus '*Salvia*' and the shorter length of East Asian *Salvia* amino acid sequence length.

4.6.4 GLOBOSA phylogeny reveals the highly supported paraphyly of *Salvia* by a trait-related marker

For the phylogenetic analysis the complete GLOBOSA region (including introns) of in total 87 sequences (deriving from 38 species and their multiple clones + *Salvia miltiorrhiza* from the herbal genome database + *Sesamum indicum* from the ncbi database) was evaluated. The alignment of all sequences resulted in a total length of 2517 basepairs. This dataset was used for BEAUti Analysis, evaluated with BEAST and TreeAnnotater and illustrated by FigTree (as described in the method part). The resulting phylogenetic tree is displayed in figure 36. The species taxa and the accession ID are labeled at the tips of the tree, the geographic background is colored by following code: European *Salvia* in green, New World *Salvia* in blue, East Asian *Salvia* in red, Lamiaceae outgroups in black.

The formerly proved paraphyly of *Salvia* could be shown based on the GLOBOSA gene phylogeny as well. *Sesamum indicum*, which belongs to the same order (Lamiales) but to a different family (Pedaliaceae) than *Salvia* (Lamiaceae), functions as outgroup for the Lamiaceae cluster. Outgroups within the Lamiaceae (labeled in black at the bottom of figure 36) are *Ocimum*, *Lavandula* and *Dracocephalum*, with the latter as sister clade to the paraphyletic *Salvia*.

Within this clade there is a highly significant separation of New Worlds and East Asian *Salvia* on the one hand and European *Salvia* and *R. officinalis* / *P. atriplicifolia* on the other. These subclades are in turn again separated resulting in a monophyletic New World, East Asian, European and *Rosmarinus*-cluster, that are supported significantly by high posterior probability values.

4.6.5 The monophyletic European *Salvia* clade falls into two main clusters

The European cluster (labeled green) consists of totally 19 different species (including multiple copies of single species) and is subdivided into two subclades, that refer to the nomenclature subgenus '*Salvia*' and subgenus '*Sclarea*' of Kriebel *et al.* (2019) [112] or likewise to two different clades of "Clade I" (*Salvia* sensu stricto), I-C and I-D, of Claßen-Bockhoff and Will (2017) [36]. The subgenus '*Heterosphace*' (Kriebel *et al.*) / *S. verticillata* group (Claßen-Bockhoff and Will) is nested within the subgenus '*Salvia* / I-D cluster.

Salvia officinalis shows the closest relation to *Salvia lavandulifolia*, that is assigned as subspecies of *Salvia officinalis (S. officinalis* subspec. *lavandulifolia*) as well [37]. The '*Salvia*' / I-D cluster of this study is composed of the eponymous species *Salvia officinalis* and morphological related species like *S. lavandulifolia* and *S. tomentosa* (see figure 36 and 13b) and highly supported by posterior probability values.

The *Salvia verticillata* group is represented by the species *S. verticillata* and *S. judaica* in this study. Those are nested within the '*Salvia*' / I-D cluster and depict the sister clade of the *S. officinalis* cluster. The '*Sclarea*' / I-C cluster of this study comprises eleven species, whereby two bigger clusters can be observed. *Salvia sclarea*, *Salvia argentea* and *Salvia aethiopis* form a significantly supported cluster, that is sister to the other eight species (including *S. pratensis*, *S. nemorosa*, *S. virgata*, *S. hierosolymitana*, *S. jurisicii*, *S. nutans*, *S. verbenaca*, *S. austriaca*). For all above mentioned species several sequences were available, originating from colonies of the same individuals that form "intraspecific" clusters (e.g. *Salvia verbenaca* 7956 K2 / K3), that are separated by high significance values from "intraspecific" cluster of other individuals (see next paragraph for details).

4.6.6 Differences in intraspecific sequences are significantly lower compared to sequences of the closest related species

For many *Salvia* species different colonies of the GLOBOSA gene resulting from different positive colonies of the cloning experiment were obtained. To illustrate the intraspecific variances compared to differences in closest related species, three examples are shown in detail (see figure 37).

The morphological and concerning pollinator syndrome quite different species *Salvia austriaca* and *Salvia verbenaca* differ in 99 of the in total 1724 nucleotides (19 of these variances occur in coding regions). The alignment of the two sequences of *S. austriaca* results in a length of 1702 nucleotides, of which 21 are variable (with five of them located in coding regions). Likewise, the aligned length of the two *S. verbenaca* colonies is 1695, with four variable sites (two, however, in the coding regions).

The morphologically similar species *Salvia nutans* and *Salvia jurisicii* have an aligned length of 1744 nucleotides, with 28 variable sites, of which seven are located in coding regions.

The alignment of the two sequences of *S. jurisicii* results in a length of 1726 nucleotides, of which six are variable (with two of them located in the coding regions). Likewise, the aligned length of the two *S. nutans* colonies is 1732, but the number of differences is lower with three variable sites (two, however, in the coding regions) (see figure 35). These two examples show that the overall intraspecific differences in GLOBOSA sequences are much lower, compared to interspecific variation.

The two species mentioned above are sister to a *Salvia pratensis*-clade, that was circumscribed like that, because all species included (*S. hierosolymitana*, *S. nemorosa* and *S. virgata*) have been considered as subspecies of *Salvia pratensis* (that is part of the clade obviously) in some determination keys. The aligned length of the GLOBOSA gene of *Salvia pratensis* and its closest relative *Salvia nemorosa* is 1721 basepairs. Comparing these two closely related species, 30 variable basepairs can be observed. Compared to that, the intraspecific number of variable sites is lower, with six differences in *Salvia pratensis* and fourteen in *Salvia nemorosa* (see figure 37).

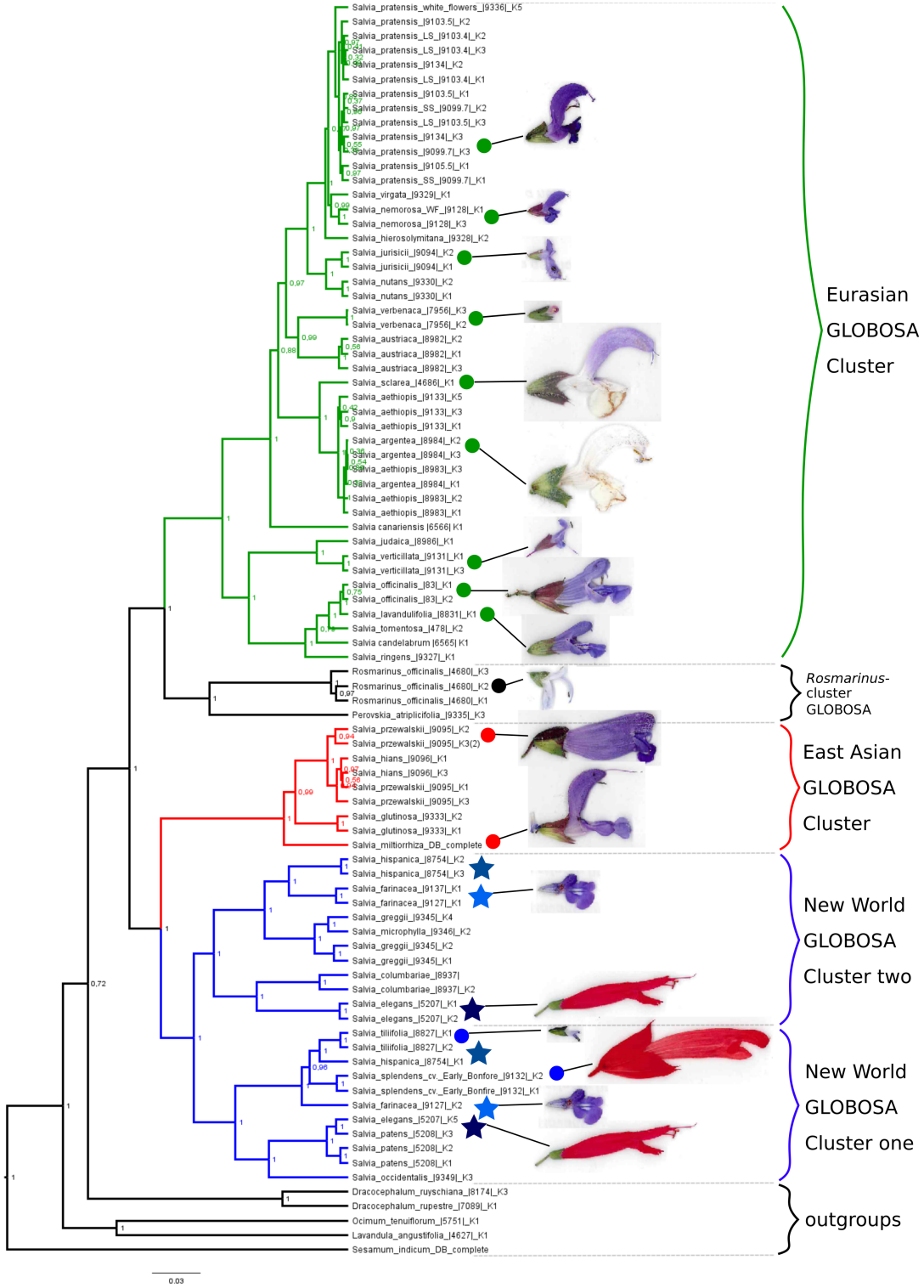


Figure 36: GLOBOSA gene based tree. European *Salvia* species labeled green, New World *Salvia* species labeled blue and East Asian *Salvia* species red. Lamiaceae outgroups are labeled in black. Bayesian inference was used and posterior probability significance values are written next to the nodes. The flowers of selected species positioned in the tree were illustrated. Stars indicate gene duplications in the monophyletic New World GLOBOSA cluster.

4.6.7 Polyphyly of several American *Salvia* species in the GLOBOSA gene-cluster

The New World cluster (labeled blue) consists of totally ten different species (with multiple copies for most species) and is subdivided into two subclades. The first thing that stands out is that GLOBOSA

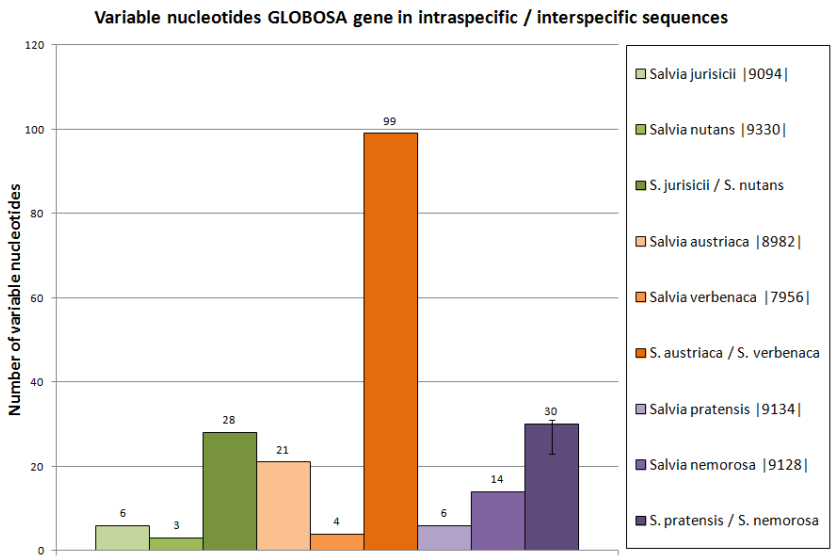


Figure 37: Comparison of intra- and interspecific differences in GLOBOSA sequences in European *Salvia* species

sequences from the same individual can be found in both of those separated clusters (*Salvia hispanica*, *Salvia farinacea* and *Salvia elegans*), labeled with stars in figure 36. All posterior probability values in the American cluster have significance values of 100 percent.
The following designations for the two American clusters are not intended as a proposal for a nomenclature, but are introduced for the sake of clarity in this thesis. In the displayed phylogenetic tree, the lower American cluster is assigned as New World GLOBOSA cluster one, whereas the upper American cluster is assigned as New World GLOBOSA cluster two (see label figure 36).

The New World GLOBOSA cluster one (lower) is subdivided into two smaller clusters, whereby one cluster can be assigned to species of the core *Calosphace* (according to Kriebel *et al.* (2019) the *Calosphace* is a subgenus of *Salvia*) [112], including *Salvia tiliifolia*, *Salvia splendens*, *S. hispanica* and *S. farinacea*. The other clade allocates species that are not part of the core *Calosphace*, with *Salvia occidentalis* and *S. elegans* (both belonging to the Uliginosae clade) and *Salvia patens* (that belongs to the Hastatae clade).

In the core *Calosphace* clade, a close relation of melittophilous *S. tiliifolia* and *S. hispanica* can be observed, however, *S. splendens*, which is ornithophilous, is closer related to the before mentioned species, compared to the melittophilous *S. farinaceae*. There is no relevant clustering of species that are melitto- or ornithophilous in the core *Calosphace* cluster. The same thing is true for the other cluster, that includes *S. elegans*, *S. patens* (both ornithophilous) and *S. occidentalis* (melittophilous).

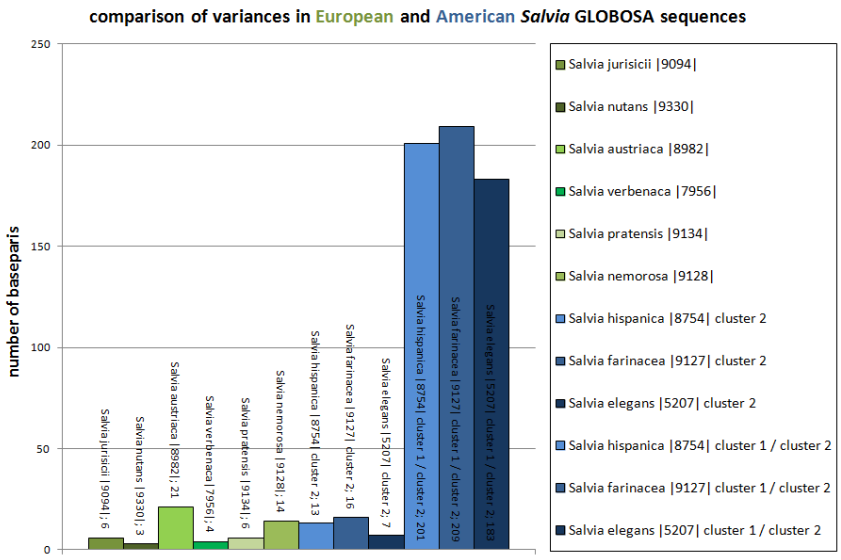


Figure 38: Data from figure 35 is expanded by intra- and interspecific differences in GLOBOSA sequences in American *Salvia* species, with focus on sequences from New World GLOBOSA cluster two and comparison of those sequence to New World GLOBOSA cluster one

The New World GLOBOSA cluster two (upper) is subdivided into two smaller clusters as well, and as in cluster one a separation of sequences of species from the core *Calosphace* (here *S. hispanica*, *S. farinacea*, *S. greggii* and *S. microphylla*) and species from the Uliginosae clade (*S. elagans*) and the subgenus *Audibertia* (*Salvia columbariae*) can be observed.

S. hispanica and *S. farinacea* (part of the Angulatae clade within the core *Calosphace*) are closely related and sister to a clade that comprises the species *S. greggii* and *S. microphylla* (part of the Fulgentes clade within the core *Calosphace*), that are morphologically very similar (bright red color) and can form hybrids (*Salvia x jamenensis*). The non-core *Calosphace* cluster includes the species *S. columbariae* and *S. elegans*, two species that differ morphologically in floral size, color and pollinator (see figure 14).

In the two New World GLOBOSA clusters one and two a separation of core *Calosphace* species compared to species not belonging to this group can be observed. In this study the genetic polyphyly of the GLOBOSA gene in *Salvia hispancia*, *Salvia farinacea* and *Salvia elegans* has been shown, which means, that this phenomena exists in species that differ significantly in floral traits (see figure 14), including size and color and thus vary in the pollinator–interaction partner (bird and insect).

The varying phylogenetic classification is attributable to the amount of variable basepairs in the sequences of New World GLOBOSA cluster one and two, which has been illustrated in figure 38. For *Salvia hispanica* the colonies K2 and K3 were grouped in cluster 2, whereas K1 was categorized in cluster 1. The total length of the three-sequence alignment totals 1795 nucleotides. The amount of variable sites for cluster 2 (K2/K3) is 13 nucleotides, compared to the significantly higher number of 201 variations when comparing cluster 1 and 2 (K1/K2/K3)(37 in coding, resulting in 9 / 214 aa nonsynonymous variable sites).

A very similar result was observed in *Salvia farinacea*, the colonies K1 of 9127 and K1 9137 were grouped in cluster 2, whereas K2 9127 was categorized in cluster 1. The total length of the three-sequence alignment totals 1776 basepairs. The amount of variable sites for cluster 2 (K1/K1) is 16 nucleotides, compared to the significantly higher number of 209 variations when comparing cluster 1 and 2 (K1/K1/K2) (39 in coding, resulting in 9 / 214 aa varieties).

The hummingbird-pollinated *Salvia elegans* GLOBOSA sequences K1 and K2 are grouped into cluster 2, whereas K5 was categorized in cluster 1. The total length of the three-sequence alignment totals 1777 basepairs. The amount of variable sites for cluster 2 (K1/K2) is 7 nucleotides, compared to the significantly higher number of 183 variations when comparing cluster 1 and 2 (K1/K2/K5) (44 in coding + nine additional basepairs in K5 resulting in 7/214 aa variables + 3 aa in K5).

Taken together, the huge differences in the two distinct clades that both contain sequences from the same individuals, while the total New World clade maintains its monophyly clearly points to a GLOBOSA gene duplication event in the New World for genus *Salvia*.

4.7 Phylogenetics of the B-class gene DEFICIENS in *Salvia*

4.7.1 Designed primers for the DEFICIENS gene yield strong (sometimes two) bands

The amplification of the DEFICIENS gene from start to stop codon in a PCR yielded fragment lengths between 1,5 kb and 2,0 kb, which has been visualized in an agarose gel 39. The primers that were used (see table 10) lead to successful amplification in all investigated *Salvia* species and additionally in other genera of the Lamiaceae like *Rosmarinus*, *Perovskia*, *Ocimum* and *Lavandula*. In contrast to the amplification of the GLOBOSA gene in some of the samples the amplification of DEFICIENS clearly lead to two bands (figure 39).

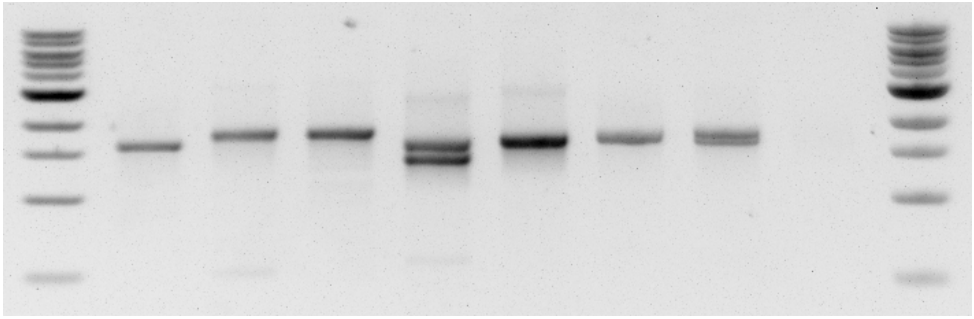


Figure 39: Amplification of the complete B-class DEFICIENS gene for different *Salvia* species by PCR from genomic DNA and evaluation by AGE. 1 kb DNA ladder. Variances in fragment size are visible, as well as the amplification of two distinct bands in some species.

Direct sequencing of DEFICIENS PCR products resulted in unusable data, since the chromatogram of the sequenced data showed double peaks with base calls (Q-values) lower than the significance value of Q20. In order to avoid inconclusive double peaks, the PCR products were cloned, just like GLOBOSA, and the plasmids of positive colonies were used for sequencing analysis. This approach delivered high-qualitative sequences that could be used for downstream investigations. In contrast to GLOBOSA, the subsequent evaluation of DEFICIENS turned out to be more complex. This was due to difficulties in alignment caused by huge length differences (showed and discussed below), also the amount of positive clones (correct insert fragment size, visualized by colony PCR) was lower compared to GLOBOSA. As a consequence of both, the datasets of GLOBOSA and DEFICIENS are not completely congruent in terms of species number.

For DEFICIENS in total 27 species were sequenced, of which 15 were individuals from Europe (including eponymic species *Salvia offcinalis* from the subgenus '*Salvia*' / I-D clade, but also several individuals from the subgenus '*Sclarea*' / I-C clade, including *Salvia pratensis*), six from the New World (including bee-pollinated *Salvia farinacea* and hummingbird-pollinated *Salvia patens* and *Salvia elegans*) and two from East Asia (including *Salvia przewalskii*). The remaining four species are from other Lamiaceae genera (*Rosmarinus officinalis*, *Perovskia atriplicifolia*, *Lavandula angustifolia* and *Ocimum tenuiflorum*).

4.7.2 Huge differences in complete and coding lengths in DEFICIENS gene(s)

The complete (exons and introns) length of the DEFICIENS *Salvia* sequences ranges from 1512 nucleotides in one copy of *Salvia patens* to 2003 basepairs in a copy of *Salvia glutinosa*.

The combined length of the seven exons in *Salvia* sensu lato ranges from 675 basepairs in *Rosmarinus officinalis* (resulting in 225 amino acids) to 723 basepairs (resulting in 241 amino acids) in a copy of *Salvia patens*.

4.7.3 Variations in coding lengths affect the C-terminal domain of the DEFICIENS protein

BLASTing the primary amino acid sequence revealed that the characteristic MIKC-type MADS protein domains of the flowering ABC genes (see introduction) can be traced in the DEFICIENS amino acid sequence of all *Salvia* species as well. In all amino acid sequences the first 60 amino acids refer to the MADS-domain (which has DNA-binding functions), the downstream amino acids 61 - 83 to the intervening domain (responsible for dimerisation of MIKC-type MADS proteins) and the amino acids

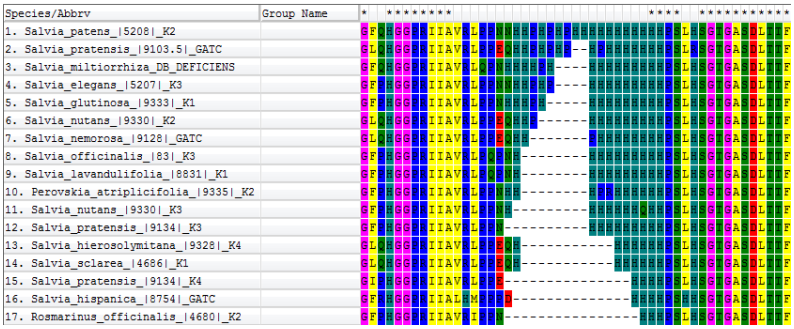


Figure 40: C-terminal ends of selected *Salvia* DEFICIENS proteins. The length of the amino acid sequence varies, because of the unequal length of histidine repeats.

84 - 164 can be assigned to the K-box domain (important for tetramerisation). The function of the C-terminal domain is mainly unknown; in DEFICIENS of *Salvia* the amino acids from site 168 to 225 / 241 refer to this domain. The above described differences in coding sequence length strongly affect the number of amino acids in the C-terminal domains of the primary protein structure. The observed disparity of sixteen amino acids can be traced back to a long histidine-rich region starting at site 213 in the alignment. This region varies from three histidines (in *Rosmarinus officinalis*) to nineteen (78,9% or 15 / 19 histidines, in one copy of *Salvia patens* (see figure 40).

4.7.4 DEFICIENS phylogeny reveals two remarkable clusters

It must be stated in advance that all of the posterior probability values in the nodes of the calculated and described DEFICIENS gene-tree have significance values of 100 percent and that the sequences that were fed to the dataset have been of highly reliable quality.

The lower cluster (DEFICIENS variant 2) comprises European, New World and East Asian *Salvia* as well as the *Rosmarinus* / *Perovskia* group (see figure 41), with a split of the European *Salvia* into two major clusters, namely '*Salvia*' / I-D and '*Sclarea*' / I-C clusters, that could be shown before in the GLOBOSA gene-tree. Looking closer at the European (gree) cluster, one can see the African species *Salvia somalensis* (that is unfortunately not present in the GLOBOSA dataset and belongs to the I-A cluster), forms a standalone lineage, which is sister to the '*Salvia*' / I-D clade. This clade contains two sequences of *Salvia officinalis* and *Salvia lavandulifolia*, whereby the two copies of *Salvia officinalis* do not cluster together, because the *Salvia officinalis* K1 shows a closer relation to the *Salvia lavandulifolia* cluster, pointing to a not so clear species delineation of this gene for the mentioned species. *Salvia tomentosa* is sister to the *S. officinalis* / *S. lavandulifolia* cluster, while *S. ringens* is the outgroup of the '*Salvia*' / I-D cluster. In the '*Sclarea*' / I-C cluster the delineation of species is clearer, with a distinct *Salvia pratensis* and *Salvia jurisicii* cluster. The latter has the closest relation to *Salvia nutans* (as in the GLOBOSA gene-tree). Together with *Salvia austriaca*, those two species form a three-species cluster that is sister to *Salvia pratensis*. In contrast to the GLOBOSA gene-tree *Salvia hierosolymitana* does not show a close relation to *Salvia pratensis*, but rather is the outgroup of the '*Sclarea*' / I-C cluster. The complete European cluster is sister to a clade that comprises the New World (blue) and East Asian (red) *Salvia* species, as well as the *Rosmarinus* clade. The latter is sister to the New World *Salvia*, whereas the Asian *Salvia* clade is more distantly related.

The upper cluster (DEFICIENS variant 1) comprises European and New World *Salvia* species, that are separated into two distinct clades (in figure 41). The American cluster is divided in three species of the core *Calosphace*, *Salvia farinacea*, *Salvia hispanica* and *Salvia tiliifolia*, whereby the latter two display the closest relation. One sequence of *Salvia elegans*, a hummingbird-pollinated species, that belongs to the Uligionosae clade, which is not within the core *Calosphace*, forms the outgroup of this American cluster. The European clade is subdivided into three clusters, with a representative of the subgenus *Heterosphace Salvia verticillata* forming an outgroup to the remaining European clusters, that are split up into two clusters, that both consists of species from the '*Sclarea*' / I-C group. One of them comprises the species *S. aethiopis*, *S. argentea* and *S. pratensis*, whereas the other clade includes sequences of *S. argentea*, *S. hierosolymitana*, *S. nemorosa*, *S. nutans*, *S. pratensis*, *S. sclarea*, *S. verbenaca* and *S. virgata*. The latter cluster contains all members of the *Salvia pratensis*-cluster (see GLOBOSA), but their relation in DEFICIENS is not as evident as in the gene-tree of the other B-class gene. The appearance of *Salvia*

argentea in this cluster and the simultaneous appearance of a sequence of *Salvia pratensis* in the *Salvia aethiopis* / *Salvia argentea* cluster, shows both of the separated clades include sequences that derived from the same individual (namely *Salvia argentea* and *Salvia pratensis*, labelled with stars in figure 41), which ultimately shows an additional genetic polyphyly within this European cluster and therefore a putative additional duplication event of DEFICIENS in this geographic region.

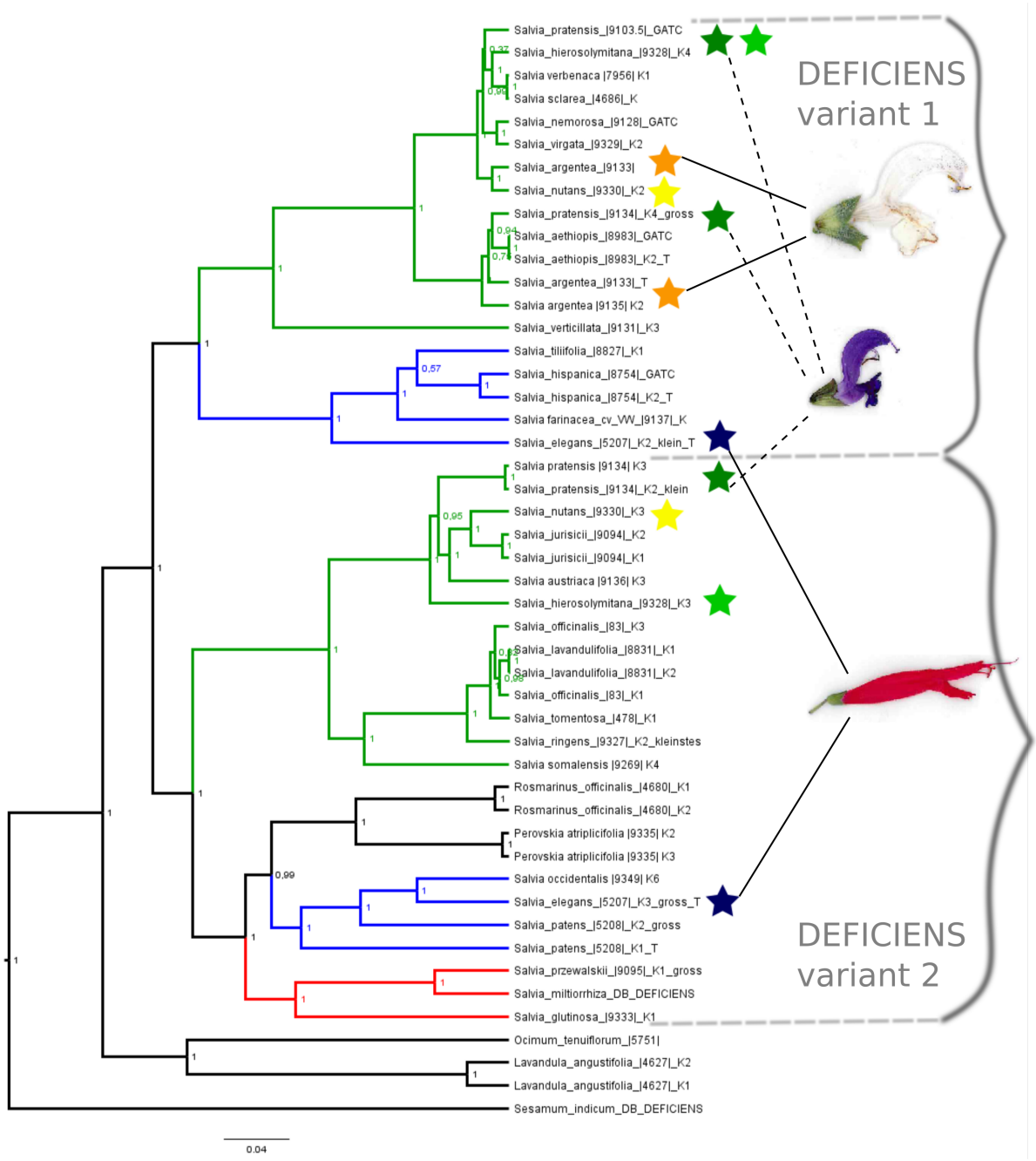


Figure 41: DEFICIENS gene based phylogenetic tree. European *Salvia* species labeled green, New World *Salvia* species labeled blue and East Asian *Salvia* species labeled red. Lamiaceae outgroups are labeled in black. Bayesian inference was used and posterior probability significance values are labelled next to the nodes. Stars indicate gene duplications in the DEFICIENS gene and the visualized flowers are examples for species in which gene duplications could be traced. The two distinct clusters were circumscribed as DEFICIENS variant 1 and 2 according to the transcripts of New World *Salvia elegans* (ID: 5207) described in the qPCR experiment (see figure 16).

4.8 Putative B-class gene duplication events in *Salvia*

For both B-class gene-trees phylogenies polyphyly could be shown for a certain extant. Table 12 displays the commonalities and differences of the two B-class genes GLOBOSA and DEFICIENS concerning this issue. Commonalities of both gene-trees are the very high posterior probabilities values, emphasizing the significance of the displayed data. Furthermore, the genus *Salvia* sensu lato (that includes *Rosmarinus* and *Perovskia*) forms a monophyletic cluster for GLOBOSA and DEFICIENS. The two trees show differences when the geographic background is taken into account. While the GLOBOSA tree is integer concerning the regions, resulting in monophyletic clades for European, New World and East Asian *Salvia* species, this pattern can not be demonstrated for the DEFICIENS gene. For this B-class gene there are two statistically significant separated DEFICIENS clades, that both include *Salvia* species from the different geographic regions. From the dataset four species could be identified that occur in both clusters, with three European *Salvia* species (*Salvia pratensis*, *Salvia hierosolymitana* and *Salvia nutans*) and one species that originates from America (*Salvia elegans*). However, going into detail of the GLOBOSA gene tree, one can observe a similar phenomena on a small scale as well, since the monophyletic GLOBOSA New World *Salvia* cluster falls into two subclades that both comprise sequences from the same species (e.g. *Salvia elegans* or *Salvia farinacea*, see figure 36).

Table 12: Comparison of GLOBOSA and DEFICIENS phylogeny

	GLOBOSA	DEFICIENS
Overall high posterior probability values?	yes	yes
Salvia sensu lato forms a monophyletic cluster?	yes	yes
Geographic regions form monophyletic clusters?	yes	no
Sequences of the same species taxon cluster together?	European: yes New World: no East Asian: yes	European: no New World: no East Asian: yes(*)

(*) in total only three sequences of DEFICIENS from East Asian *Salvia* species have been evaluated, most likely because of the narrow dataset, duplication could not be observed here.

Based on the results for both B-class genes and their putative duplication events, a model has been elaborated and illustrated that simultaneously suggests a history of B-class genes in *Salvia*, as well as a possible explanation for the elevated floral diversity in the New World for this genus (see figure 42).
The explanation rests upon the assumption that in the angiosperms with very few exceptions GLOBOSA and DEFICIENS proteins only form heterodimers for their mode of action (as described in [86]). With the duplication of GLOBOSA in the New World the amount of possible GLO-DEF heterodimers doubles from two to four, which subsequently enhances the amount of floral quartet complexes that can be build with the respective GLO-DEF heterodimers and therefore ultimately leads to an expansion of possible transcription factor interactions.
This might be a driving force behind the extension of floral diversity and hence to richer species diversity of *Salvia* in the New World.

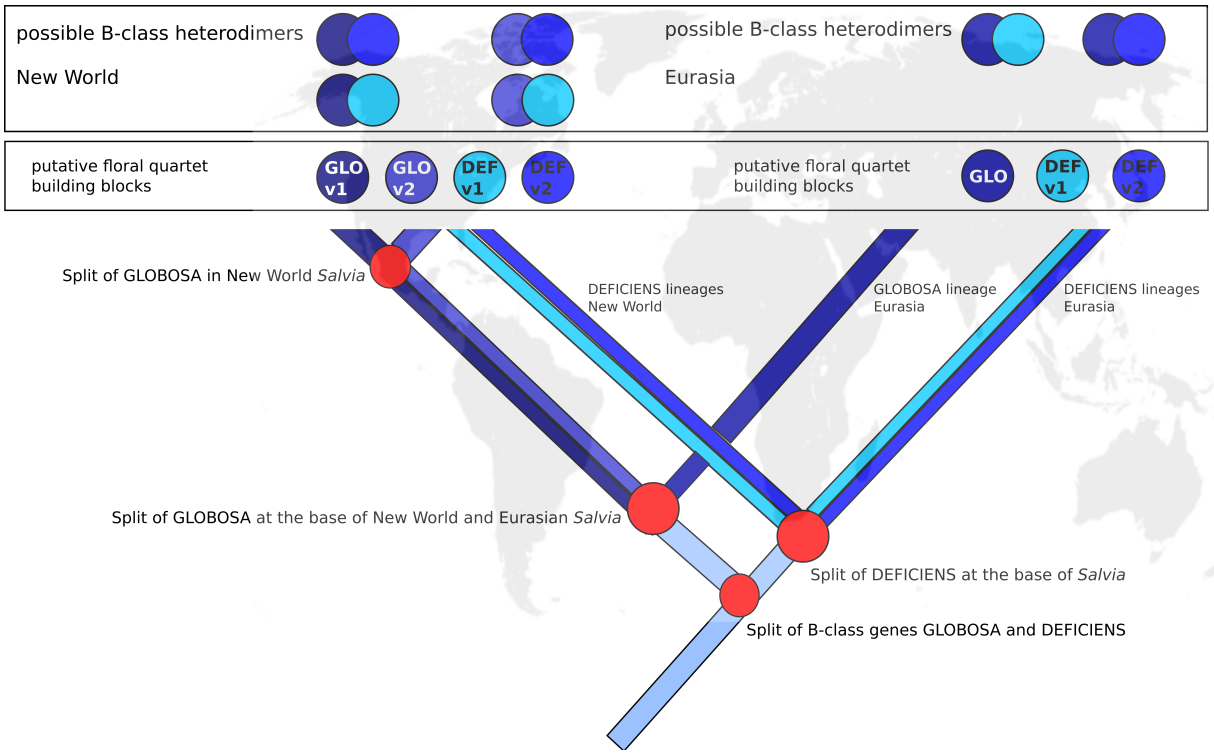


Figure 42: Proposed model on the history of B-class genes in *Salvia*, together with a working hypothesis explaining the extended floral diversity in the New World. Red circles display putative duplication events. Blue lines display the different B-class gene lineages. Blue circles display the B-class MIKC-type MADS-domain proteins. Assembling of those proteins to heterodimers and the ensuing number of possible interactions are illustrated as well.

4.9 ABCDE genes from different realms of the flowering plants cluster together

The phylogenetic gene tree of the ABCDE flowering gene amino acid sequences of different *Salvia* species, as well as the *Arabidopsis thaliana* rosid outgroup yields a statistically highly supported tree the clearly separates the different classes of flowering genes (see figure 43). The MIKC-type MADS genes form a supercluster, that is separated from the *Solenostemon* LEAFY gene outgroup by 100%. Within the ABCDE cluster there is a clear separation of the B-class genes (GLOBOSA and DEFICIENS, labeled in dark and light blue) forming one cluster, and the ACDE genes, forming another cluster. The B-class cluster is split in GLOBOSA and DEFICIENS clades, which contain all *Salvia* species of the dataset as well as the rosid outgroup *Arabidopsis*, showing a clear angiosperm over spanning delineation of those functional closely related B-class genes. Sequences of *Salvia* species that were separated in the GLOBOSA and DEFICIENS gene tree which point to possible duplication events (New World GLOBOSA species, see 4.4.1 and DEFICIENS in general 4.4.2) stay within the respective *Salvia* clusters and are separated from rosid *Arabidopsis*, supporting the hypothesis of an independent duplication event in the Lamiaceae. The sister clade of the B-class MIKC-type MADS genes comprises all other classes of the ABCDE model. This cluster is split into an A-class / E-class (labeled green / orange) and a C-class / D-class (labeled red / turquoiseish) cluster, respectively, and all separations depicted are highly (close to 100%) supported. The functional connection of the ABCDE genes within the homonymous model can be read of this data, since the key players of sepal (A and E) and carpel (C and D) development form distinct clusters. Furthermore, the functional restriction of B-class genes to form heterodimers exclusively with themselves is congruent with the displayed data, since the B-class gene form a separated monophyletic genetic lineage that is sister to the ACDE gene clade.

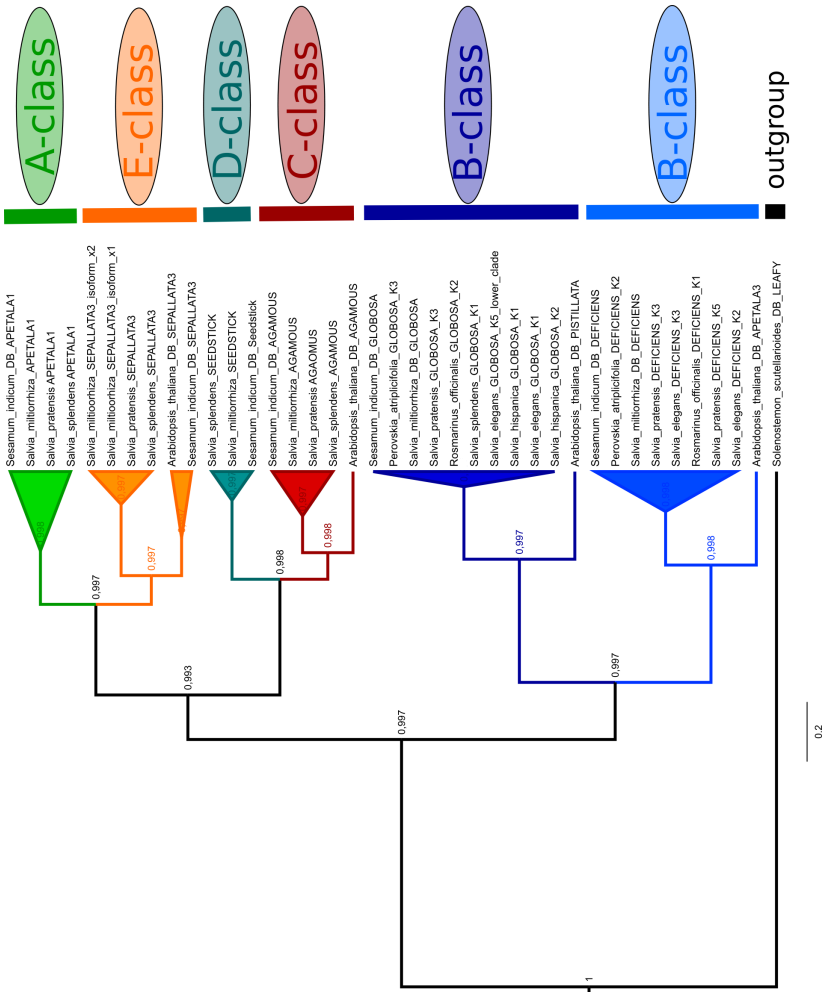


Figure 43: Illustration of the genetic relation of the ABCDE genes in *Salvia* with reference to distant related rosid *Arabidopsis thaliana* sequences. A-class: green labeled; B-class: blue labeled (light and dark for DEFICIENS and GLOBOSA, respectively); C-class: red labeled; D-class: turquoiseish labeled; E-class: orange labeled.

4.10 Short summary of the results

Morphological analysis and evaluation of *Salvia* flowers

- The variety of *Salvia* species is represented by a unique pollination mechanism and variformity in flowers, including size, shape and color.
- The statistical evaluation of floral measurements by principal component analysis reveals gradients of flower size along the x-axis (PC1) and the transformation of the corolla shape from tuberous to clear lip shape along the y-axis (PC2).
- New World *Salvia* species display a greater floral diversity leading to a spread all over the PCA graph, whereas European *Salvia* species show a more limited spread.
- Nevertheless, there are no clear geographical clusters visible, since there are overlaps of European, New World and East Asian *Salvia* species.
- This pattern is endorsed by a dendrogram depiction of the same dataset, that reveals different clusters, that comprise European and American *Salvia* species, respectively.

qPCR based evo-devo evaluation of bee- and hummingbird pollinated *Salvia* species

- The general gene expression pattern proposed by the ABCDE model could be demonstrated in two distant related *Salvia* species for the first time.
- The five isolated floral tissues of *Salvia pratensis* and *Salvia elagens* all display characteristic expression patterns, that correspond with the ABCDE model.
- The significant differences in gene expression comparing closed and opened flowers endorses the crucial role of ABCDE genes during flower development.
- The transcriptional representatives of auxin, the Auxin Response Factors 1 and 25, are involved during development of the flower, especially in corolla and style growth.

Phylogeny of auxin-based transcription factor and flower developmental B-class genes

- *Salvia* ARF1 multiple sequence alignment reveals a highly conserved control mechanism of ARF1 by miRNA160.
- The ARF1 based phylogeny endorses the GLOBOSA gene tree topology as well as the paraphyletic nature of the genus *Salvia*.
- In-depth analysis of B-class genes for a large number of *Salvia* species yielded a dataset for a significant illustration of the phylogenetic relation for the genus *Salvia* by a trait-related marker.
- The phylogenetic analysis of the GLOBOSA gene resulted in a gene tree that is statistically strongly supported and leads to clear geographic clusters, on the one hand, and demonstrates the paraphyly of the *Salvia*, on the other hand.
- The monophyletic New World *Salvia* cluster of GLOBOSA shows signs of gene duplications.
- The DEFICIENS gene shows a clear duplication, since the phylogenetic tree demonstrates two clear clusters, that both include sequences originating from the same European and American *Salvia* individuals.
- Evaluation of ABCDE genes of *Salvia* in a wider context reveals clusters of A- and E-class genes, C- and D-class genes and a B-class gene cluster, respectively.

5 Discussion

5.1 Morphology - *Salvia* species taxa inhabit distinct statistical niches

The amount of *Salvia* species evaluated in the course of this thesis displays only a small fraction of the worldwide distributed 1000 different species of this genus. Nevertheless, care was taken to ensure a broadly diversified choice of species with respect to origin, flower characteristics or pollinator, in order to establish a robust morphological backbone for the thesis.
The question put forward in the scope of the study on the viability of the statistical evaluation of floral traits for the classification of *Salvia* shall be discussed against the geographical background of the genus.

Species taxa are represented by single dots in the PCA graph (see figure 14). The reason for the spread of the American *Salvia* species over the whole PCA graph as well as in the dendrogram lies in the extensive variformity of *Salvia* flowers in the New World. With an approximate amount of 500 different species, this region is clearly underrepresented in this study (with nine species), however, the inclusion of morphological very different representatives allows for a sound statement. Even from the narrow corridor of evaluated New World *Salvia* species the vast diversity of American *Salvia* may be carefully guessed, since the spread among the whole PCA plot points to an overall higher variation concerning size and shape for *Salvia* species on this continents compared to by number better represented Eurasian *Salvia* species.
The statistically deducted floral geometric similarity of e.g. Eurasian *Salvia hierosolymitana* and American *Salvia greggii* or European *Salvia jurisicii* and American *Salvia farinacea* illustrates the geographical independent formation of similar floral geometries in the Old World and in the New World. Although there are similarities in flower shape, the European limitation of floral characteristics becomes evident in the statistic evaluation, because the European *Salvia* species are more restricted in almost all geometric dimensions. The floral expansion of American *Salvia* species applies to size (larger and smaller) and to shape, ranging from absolute lip form (e.g. *Salvia patens*) to perfect tuberous form (e.g. *Salvia elegans*), meaning that the development and the stretch in dimensions of floral structures correlates with the geographic migration of an ancestral *Salvia* species to the New World and the conditions encountered in this new environment.

The genus *Salvia* is thought to have originated in the Mediterranean region (according to [36]) or Southwest Asia (according to [112]) in the early Oligocene around 32 million years ago [36][112]. The first migration of *Salvia* to the New World most likely occurred in the late Oligocene, 24 to 22 million years ago, from Southwest Asia through the Bering land bridge (which could be colonized by deciduous plants at that time) to the North American continent [36][112]. This dispersal event unleashed the rise of subgenus *Audibertia* / genus *Ramona* and subgenus *Calosphace* / genus *Lasemia* [112] (nomenclature according to [35] and [36], respectively) and is strongly supported by a pollen fossil finding of an *Audibertia*-type *Salvia* in Alaska that is dated to the upper Miocene (around 20 million years ago) [133]. The first transition to bird pollination based on the presence of hummingbirds in the New World opened a window for another type of pollination and thus a co-evolutive adaptive radiation driven species diversification. Though this transition marks a cornerstone for *Salvia* species burst on the American continents, its the multiple reversals back to bee pollination in the hummingbird clade that displays another major facilitator of species richness in this genus [112].
In the evaluated morphological dataset *Salvia elegans*, *Salvia patens* and *Salvia splendens* are hummingbird-pollinated species, while the reversals to bee-pollination are represented by significant smaller flowers with differently shaped corollas in *Salvia hispanica* and *Salvia farinacea* that both belong to the "hummingbird-cluster" (see figure 14).

Morphometric analyses of *Salvia* flowers with focus on the corolla geometry have been conducted in the past for South African ornithophilous species [134] or Middle East melittophilous species [135]. However, measurements carried out in this thesis were orientated on Benitez-Vieyra *et al.* (2019) [125] concerning petal measure sites as well as evaluation and illustration of data sets. The authors analyzed in total 18 melittophilous and ornithophilous *Salvia* species from the New World, with *Salvia elegans* as the only species that was also evaluated in this study. Although different *Salvia* species from different continents were analyzed, the overall resemblance of results is astounding. The first principal component (79,21% explained variations) in the present data set was connected to flower size, which is congruent with the study of Benitez-Vierya *et al.* where the corolla size (and here the corolla tube length) provided

84,50% to the first principal component. Furthermore, the second principal component in their study separated *Salvia* species with tuberous floral forms from *Salvia* species with clearly bilabiate flowers [125], which also overlaps with the present results. The original data sets of both studies were composed by other species from different geographical regions, leading to the assumption, that first size and then shape are the main characterizing features of *Salvia* flowers, or to be more precise the corollas of *Salvia* flowers.

Narrowing the corolla tube width with a simultaneous extension of the corolla tube length prohibits insects to go for the nectar, reserving the pollination reward for approaching hummingbirds.
There is an ongoing debate whether it is possible to predict the pollinator from floral structures. So called pollination syndromes are "reproductive characters that match flowers to functional pollinator groups" [136] and describe floral traits that pinpoint to a specific pollinator. In a large-scale study, which comprised the evaluation of flowers from different communities like tropical rain forests or savanna on three different continents, for two thirds of the flowers the "most common pollinator" could not be predicted [137], whereas a more recent study of 550 angiosperms from the Andean region hinted for possible pollinator prediction and thus positive pollination syndromes [136].
With regard to the data collected in this study, syndromes for hummingbird pollination seems to be size of the flower on the one hand, but on a more fine scaled level, the relation of corolla length and width. For corollas with an enlarged length and the maintenance of a sufficient width, the insect (observed for bees) is still able to crawl into the flower (e.g. in *Salvia przewalskii*), though it needs to increase the effort and has to squeeze its body completely into the corolla of the flower and only the backmost part of the abdomen is still visible (own observation displayed in figure 13i). New World *S. elegans* has a corolla length that is comparable to East Asian *S. przewalskii*. However, the insect is not able to crawl into the flower of *S. elegans*, because the corolla width limits the possibility to enter and to access to nectar [own observation]. The flowers of hummingbird pollinated *Salvia* species are restricted for insects, but not vice versa, since hummingbirds should be able to raid the nectar of smaller flowers as well. A personal observation of hummingbirds could not be done in this study, since those birds are nowadays restricted to the New World. But it could be observed that bees approach the flowers of the hummingbird-pollinated species *S. elegans* in the botanical garden of the KIT and raid the nectar by successive destroying the back part of the corolla, to avoid the too narrow corolla tube and get "illegally" access to "hummingbird-nectar" [own observation].

Hybridization is common in closely related *Salvia* species. As an example, hybrids of *Salvia pratensis* and *Salvia nemorosa* were already described in the 19^{th} century [138], with possible hybridization events of small *S. pratensis* flowers and large *S. nemorosa* flowers [34]. The illustration of the very similar shape of those two species, and only a shift of the flower size has been demonstrated in the PCA graph (figure 14).
Another example is the eponymous species *Salvia officinalis* that has been shown to form hybrids with *Salvia fruticosa*, whereby hybrid individuals form floral morphological intermediates, which still could be clearly separated [139]. Though *Salvia fruticosa* was not part in the present study concerning morphology, another candidate for hybridization, with even unclear taxonomical classification or status, namely *Salvia lavandulifolia* (also described as *S. officinalis* ssp. *lavandulifolia*), was analyzed by floral measurements in the present study. The principal component analysis revealed completely similar corolla geometry with a small shift along the x-axis because of the floral size. This result strengthens the suitability of morphometry based principal component analysis for separation of closely (and distant) related *Salvia* species taxa into distinct statistical niches. With regard to these results, the importance of morphological measurements in addition to molecular approaches should be emphasized.
Differentiation of closely related species is possible, although the very significant structural differences of stamen geometry, which have been elaborated in detail by the group of Claßen-Bockhoff over the years (reviewed in [38]) have not even considered to full extant. This suggests that more intense evaluation of floral traits might lead to an even clearer differentiation of closely related species, that might ultimately be exalted above phenomena like phenotypic plasticity.
It should be noted though that the statistical evaluation of floral traits is not suitable for drawing a morphological based phylogeny that fits to the previously mentioned dispersal events, since geographically independent homologous floral development, as described in *S. hierosolymitana* / *S. greggii* and *S. jurisicii* / *S. farinacea* is a too predominant feature of the *Salvia* genus.
Anyways, the assumption that each of the species taxa carefully identified by morphological traits inhabits a statistical niche based on the deduced two-dimensional illustration of multidimensional measurements strengthens all downstream applications and therefore could be the prerequisite for biodiversity studies.

5.2 qPCR - demonstration of temporal spatialization in development

Despite the astonishing number of Lamiaceae species and their co-evolution with insects, which emphasizes the importance of floral structures, there are little to no studies on the gene expression of flower developmental genes in the context of flowering for this plant family. In the course of this thesis expression patterns for developing and opened flowers were performed in two distant related *Salvia* species, European bee-pollinated *Salvia pratensis* and American hummingbird-pollinated *Salvia elegans*. The approach of this thesis was bottom up, meaning that the orthologues of the most prominent transcription factors from the ABCDE model were browsed in the available genome data of *Salvia miltiorrhiza* and subsequently evaluated in the European and America *Salvia* species based on designed conserved primers. The objective that was put forward in the scope of the study, to transfer the ABCDE model to *Salvia* could be could be demonstrated for the two *Salvia* species, which marks the first transmission of the flowering model to this large genus. The exception is the D-class because here was no representative gene available.

The expression analysis for genes of interest depends on the comparison to an internal standard that is provided by a constitutively expressed house keeping gene. Generally finding stable house keeping genes for the respective experiment is not a trivial task. Different house keeping genes have been tested in the context of flowering and two of the "classics" in plants, namely actin and ubiquitin, turned out to be unusable for different reasons. As described in results 4.4.2, the amplification of by ubiquitin primer pair that was proposed in [122] yielded double bands for both *Salvia pratensis* and *Salvia elegans* in semi-qPCR experiments that would have been naturally transferred to the qPCR results as well. The amplification of the actin gene lead to one band, however, the intensity of fragments in semi-qPCR was inconsistent in different floral tissue, making this house keeping gene as well unusable as internal reference. The differing expressions of actin might point to a reduction or elevation of actin in different floral tissue, because of rearrangement of the actin skeleton during floral development. Since it is not feasible for usage as house keeping gene this track was not pursued further, might be worthwhile in the future though. In fact, the 18S gene was the only house keeping gene that was stable during flower development, and therefore was used as reference.
However, the third biological replicate of opened flowers of *S. pratensis* displayed an unstable expression for the 18S gene, leading to distorted results, that were spared from the analysis. Because of that, for future approaches further house keeping genes should be considered and tested in the context of flowering.

The duplication of DEFICIENS has been clearly demonstrated by evaluation of transcript sequences as displayed in figure 16. Both transcripts resulting from this duplication are apparently functional, since no early stop codon could be traced in the sequence alignments, however, whether this functionality is evident *in planta* as well, needs to be evaluated. This has been partly done in the course of this thesis, for one of the two copies in qPCR experiments for *Salvia pratensis* and *Salvia elegans* respectively, because unfortunately the respective other transcript yielded unusable data (often double peaks in the qPCR curves) in both species. As illustrated in figure 16 the different transcripts are related since all of them belong to a DEFICIENS cluster that is clearly separated from B-class GLOBOSA transcript sequences, which is the closest related gene of DEFICIENS (see ABCDE phylogeny in figure 43). However, the positive results of the different DEFICIENS transcripts for *S. pratensis* and *S. elegens* originate from distant related copies of this gene.
This means DEFICIENS does not equal DEFICIENS for the data of the two *Salvia* species and consequently the evaluation and comparison should be handled with care.
In the following paragraphs the individual floral organs and the collected gene expression data will be discussed in the context of the organs development and function.

5.2.1 Calyx - enclosing and ensuring save floral development

According to the ABCDE model, A-class and E-class homodimers tetramerize, in order to orchestrate the development of the calyx. In both *Salvia* species the expression of A-class and E-class show the highest values (see figure 17 and 22) compared to the other flowering genes. Although for the examined closed flowers a relatively late stage of development has been chosen, especially concerning size of the calyx, which almost had the size that it has in the opened flower, there is still a significant higher expression of A-class gene APETALA1 in closed flowers of both species. This indicates that this MADS-gene controls calyx development throughout the whole developmental process and not only functions for the initial

formation of this organ (as suggested in [69]). The significant drop after opening of the flower (in both species again) furthermore indicates that this expression is to a great part limited to the development of the calyx until the flower has opened. However, though AP1 expression level drops after opening of the flower, in *Salvia elegans* this gene remains active and shows the highest expression of all genes in this organ and stage. This pattern can not be observed in *Salvia pratensis*; in this species, the decrease of AP1 gene expression falls below the expression of other flowering genes, e.g. SEPALLATA1 and SEPALLATA3.
During seed production, which takes place covered within the calyx, this floral organ undergoes a further increase in size, while the rest of the flower, especially the corolla and the male reproductive organs are dismantled by senescence processes. Another elevation of APETALA1 is likely during the seed-producing process, since it has been demonstrated to function in calyx growth. This assumption needs to be evaluated though, by analysing the sepal transcripts of APETALA1 during the stage of seed development.
The expression of E-class genes SEP1 and SEP3 in both *Salvia* species in both developmental stages of the calyx strengthens the assumption that this gene class acts together with the A-class on the protein level in the most outer whorl. In both *Salvia* species the SEP1 expression is stable and shows no differences in the evaluated developmental stages. The SEP3 expression, however, shows a big difference between the two species, since there is a significant upregulation in *S. pratensis* and a down regulation in *S. elegans* suggesting varying functions in this organ.
The higher expression of SEP3 indicates a more dominating part in the floral quartets compared to the SEP1 protein, which goes hand in hand with earlier findings in other plant species and underlines the gluing function of SEP3 in floral quartets [101].
The B- and C-class genes display a very low expression that is congruent with the ABCDE model, since those genes have been shown to be downregulated by APETALA1 (true for B-class genes [74]) and by APETALA2 (which applies to AGAMOUS [77]) in *Arabidopsis*. Though this downregulation has not been explicitly shown for the *Salvia* species, it can be assumed from the available data that this conserved control mechanism is also evident in the Lamiaceae, represented here by genus *Salvia*.

5.2.2 Corolla - birds and bees attracted by B-class induced beauty

The gene expression pattern in the corolla tissues of both *Salvia* species is dominated by the B-class gene GLOBOSA, especially in *Salvia pratensis* where the expression is vastly elevated compared to all other genes. The expression of those genes in the corolla organ was expected and the positive results therefore again support the conservation of the ABCDE model in *Salvia*. However, the extreme domination of the B-class gene GLOBOSA is still surprising. Considering the floral quartet model, the B-class gene GLOBOSA makes up only one forth of floral quartets that control corolla and stamen development, enhancing the surprise caused by the vast expression, especially with regard to the other B-class gene, DEFICIENS. This gene is, compared to GLOBOSA, only weakly expressed, although those transcription factors only function as heterodimers in almost all angiosperms [86]. Explanation attempts for the highly elevated expression (especially in *Salvia pratensis*; in *Salvia elegans* GLOBOSA is highly expressed as well, but there is a significant rise of elevation in DEFICIENS expression as well) could be that *Salvia* is one of the very few exception where the formation of functional GLOBOSA homodimers occurs. While this is very unlikely, the possibility should still be mentioned. Another explanation could be the autoloop feedback mechanism of B-class genes for the promotion of their own expression [78][82]. The high expression of B-class GLOBOSA might be due to initial initiation of this feedback loop by transcripts of this gene.
In this thesis the duplication of the B-class GLOBOSA gene in New World *Salvia* species has been demonstrated (see results figures 36 and 42 and the discussion on this issue in 5.4), while this duplication could not be shown for European species (based on the missing duplication in *Salvia pratensis*, see also figure 36). This duplication event was traced by sequencing and analyzing the complete gene of B-class GLOBOSA, using the primers displayed in table 10. Those primer cover the whole gene, while for the qPCR experiment a new pair of primer has been designed (see table 11) to have an optimal length for the qPCR experiment.
While the "phylogenetic primers" (table 10) amplified both American GLOBOSA copies, because the whole gene was covered, the qPCR primers (table 11) might have unintendedly only amplified one of the American GLOBOSA copies. This would explain on the one hand the single peak for the qPCR results for *Salvia elegans* (a New World species) GLOBOSA (while there are clearly at least two copies of this gene), as well as the relative weaker expression of American GLOBOSA compared to the vast expression of GLOBOSA in the European *Salvia pratensis*. In this case the two GLOBOSA genes of the New World

would gone through subfunctionalization or neofunctionalization. The complete absence of C-class gene AGAMOUS in the sterile part of the flower (the perianth) underlines its exclusive functioning in the third and fourth whorl.

5.2.3 Ovary - hidden development might blur clear gene expression patterns

The interpretation of ovary gene expression results in *Salvia pratensis* is difficult, because the expression varies strongly, resulting in very high error bars, prohibiting sound statements. In addition, there is no gene that really stands out, complicating an interpretation. Coming from the ABCDE model an elevated expression of C-class gene AGAMOUS together with D- and E-class genes was expected. This pattern could be shown in part, with the highest expression of AGAMOUS and SEPALLATA3 after the flower has opened which represents a significant increase of these genes compared to closed flowers in *Salvia pratensis*. This can be explained with the developmental process namely fertilization going on in the ovary after the flower has opened. Though care was taken to have similar / comparable developmental stages, this has been tricky for the ovary, since it could not be observed whether a fertilization event has taken place. This might serve as an explanation why the expression results in the ovary varies so strong as it does. The expected D-class gene expression could not be demonstrated, because the representative of the D-class, SHATTERPROOF, could not be successfully amplified during a qPCR experiment.
For this thesis, the expression evaluation in distinct floral organs of *Salvia* has been conducted for the first time; for future attempts ovary and style could be left together, in order to explore the expression by floral whorls.

5.2.4 Stamen - Maintaining the expression to maintain the flexibility of the tissue

The expression pattern of the flowering genes in the male reproductive organs of *Salvia* is for the most part congruent with the expectation based on the ABCDE model. Relative B-class gene GLOBOSA expression is similar in both species, with a high initial expression in developing tissue, and a subsequent elevation of this gene in opened flowers.
It has been demonstrated that the staminal lever mechanism of *Salvia* can be triggered multiple times, up to 17 times in *Salvia pratensis* [34], resulting in multiple chances of pollen spread for the plant, until the tissues shows signs of fatigue. The continuous expression, with even an elevation of expression for the B-class genes, might maintain the integrity of the staminal tissue to ensure all of the pollen enclosed in the anthers can be dispersed.
There is a big difference in the expression of the B-class gene DEFICIENS when comparing the European and American *Salvia* species. While the DEFICIENS expression remains stable in the stamen through the different development processes in *Salvia pratensis*, there is high expression in closed flowers in *Salvia elegans*, followed by a significant drop after the flower has opened. As demonstrated in the phylogenetic experiments there are (at least) two different copies of the DEFICIENS gene that result from a duplication event at the base of the *Salvia* genus (see figure 42). For both of these copies, qPCR primers have been designed and tested, however, for both *Salvia* species only one of the copies yielded usable qPCR data (see figure 16 and discussion 5.2), DEFICIENS variant 1 for *Salvia pratensis* and DEFICIENS variant 2 for *Salvia elegans*. In the respective other variant of both species the occasional occurrence of double peaks lead to unusable qPCR results. The different copies might have undergone different subfunctionalizations, which could be an explanation for the differing expression patterns in the two species.

Commonalities in staminal gene expression are the increase of expression in both species of the E-class genes SEPALLATA1 and SEPALLATA3, pointing to an conserved mechanism of upholding the staminal integrity. In *Salvia elegans* both representatives of the E-class display a similar level of expression, while in *Salvia pratensis* SEP3 is more predominant, which could be an indicator for varying composition of the floral quartet complexes in the two different species.
The C-class AGAMOUS gene transcription factor that is, according to the ABCDE model, as well part of the floral quartet complex during stamen development, shows a corresponding expression in both *Salvia* species. This supports the model and further underlines its projected control mechanism in repressing A-class gene in the inner whorl, that can be deduced from the close to absent expression of APELATA1 in staminal tissue.

5.2.5 Style - Indications for ARF1 mediated style elongation

The pollen tube road that is represented by the style displays slightly different gene expressions comparing *Salvia pratensis* and *Salvia elegans*. To start with the commonalities, in both species the highest expression for this tissue is the C-class gene AGAMOUS during the development of the flower, confirming the expectation for the inner whorl from the ABCDE model. The significant drop of AGAMOUS expression in both species underlines the main activity of this gene during development when the style is not yet exposed. The ABCDE model as well as the floral quartet model proposes the joint expression of C-class and E-class genes in carpels. While the strong activation of C-class gene AGAMOUS could be demonstrated, the expression of the E-class genes SEP1 and SEP3 is surprisingly low, with the exception in opened flowers of *Salvia pratensis*. Possible explanations could be that the relatively low expression of E-class genes compared to its tetramerization partner AGAMOUS is due to the fact, that the C-class comprises only one gene, whereas there are more E-class genes (four in *Arabidopsis*) and the single representatives could be complementary expressed. For the model plant in total four SEPALLATA gene have been described and in this thesis only two have been examined, it is therefore possible that different SEPALLATA genes interact in different tissues with the more specific ABCD gene classes. Thus, lacking SEP2 or SEP4 genes in this study, might be the interaction partners for carpel identity. The upregulation of SEP3 in *Salvia pratensis* after the flower has opened, suggests for processes that take place after pollen has attached to the stigma.
The strongly elevated expression of ARF1 compared to all other investigated floral tissues provides a strong indication that the elongation of the style within the closed flower is under control of this Auxin Response Factor, proving the elevated presence of auxin. The main growth within the flower is displayed by significant drop of ARF1 after the flower has opened in *Salvia pratensis*. However, the growth of the style is important after the flower has opened as well, to reach and pick up the pollen that arrives on the abdomen of the insects (see figure 5). In contrast to this, the expression of ARF1 in *Salvia elegans* is stable in closed and opened flowers, suggesting a more permanent growth or a possible down regulation by miRNA160.

5.3 Auxin Response Factor 1 - between conserved and variable

5.3.1 The conserved interaction with microRNA160

For all evaluated species (including the genera *Salvia*, *Rosmarinus*, *Perovskia*, *Dracocephalum*, *Mentha* and *Sesamum*) the target sequences site for miRNA160 is conserved by 100% (or 21 of 21 nucleotides), which results in parsimony-uninformative sites for phylogenetic analyses, however, pointing to a highly evolutionary conserved control mechanism of ARF1 by miRNA160 for the whole Lamiaceae family.
The conserved interaction of smARF1 (atARF16) and miRNA160 (please note that the numeric labelling of ARFs varies, while labelling of microRNA is more "conserved"; the biological conservation is kind of reflected in the anthropogenic nomenclature) has been proposed in different realms of the flowering plants (from rosids (*Arabidopsis*) to asterids (*Salvia*)). The results of this study clearly indicates that the conservation of this post-transcriptional control mechanism is evident throughout the major clades of the genus *Salvia* sensu lato. Since this has been shown for other Lamiaceae genera as well, the assumption that miRNA160 regulates the expression of ARF1 throughout the whole family is funded. However, this is based on the presence of the binding site of miRNA160 in the ARF1 sequences in the Lamiaceae, the *in planta* interaction is still to be shown. The provided floral ARF1 expression patterns in flowers may serve as background for subsequent miRNA160 expression experiments to find potential antiproportional overlappings.
Having this interaction in mind, all interpretations on gene expression experiments of ARF1 have to be handled with care, because seemingly high abundance of transcripts can possibly degraded before translation into proteins. Simultaneous expression levels of microRNA precursor genes need to be evaluated in order to deduce a possible down regulation of the target transcript (ARF1), however, so far the precursor genes of miRNA160 remain mostly unknown, making a said evaluation an outlook for the future.

5.3.2 The variable possibilities for phylogeny

The full Auxin Response Factor 1 gene of *Salvia miltiorrhiza* has a complete length of 2260 basepairs. The phylogeny of the ARF1 in this thesis was based on sequences of the second exon and second intron of this gene, that had a fragment length of about 800 basepairs (see results 4.5.1).

The idea was to access an easy-to-sequence part of a gene that takes part in the speciation process. The participation was deduced from the putative involvement in floral development suggested by Xu *et al* (2016) [57] and confirmed in this thesis, especially for the facilitation of style elongation in developing *Salvia* flowers. The arrangement of the ARF1 protein domains comprises a highly conserved DNA-binding domain as well as the characteristic auxin-response domain followed by more variable regions (see results, figure 31). The first two domains are conserved and therefore appeared to be phylogenetically uninteresting; consequently the focus laid on the second part of the gene. The designed ARF1 primers (see table 10) flanked the second exon and intron over a region of around 800 basepairs. This length was chosen to easily amplify the fragments with current commercial approaches that can cover lengths up to around 1100 basepairs. The single-copy nature of this gene made amplification and accession of sequences easy and no laborious cloning was necessary. The resulting sequences were gathered, aligned and evaluated as described in the method part 3.11. The topology resembles earlier plastidic based phylogenies and underlines paraphyletic nature of *Salvia*, with *Salvia* sensu lato forming a monophyletic cluster. The split into separated geographic clusters is strongly supported and the two main clusters are European *Salvia* including a *Rosmarinus* cluster and a New World / East Asian cluster. The monophyletic New World *Salvia* cluster falls into the core *Calosphace* (including bee-pollinated *Salvia hispanica* and hummingbird-pollinated *Salvia splendens*) and the non-core members of subgenus *Calosphace* (namely Uliginosae member *Salvia occidentalis* (melittophilous) and Hastatae member *Salvia patens* (ornithophilous)), that both include differently pollinated species and extreme differences in floral size and shape.

ARF1 most likely has additional functions in *Salvia* plants next to the style elongation during flower development, which might be the reason that the gene-tree topology is not redrawn in a developmental sense, but resemble trait-unrelated plastidic markers.
Still, the phylogenetic analysis of fragments from ARF1 provides statistically supported results that are comparable to combined psbA-trnH and ITS data and has, unlike the aforementioned, a demonstrated background in floral development, a key factor for plant speciation.

5.4 B-class gene GLOBOSA - evo-devo candidate for phylogeny and prime example for duplication studies

The designed primers for the amplification of GLOBOSA were based on shared conserved regions of *Sesamum indicum* and *Salvia miltiorrhiza* at the 5'- and 3'-ends of the genes. The successful amplification of GLOBOSA and DEFICIENS of all *Salvia* species and additionally species from other genera (*Rosmarinus*, *Perovskia*, *Ocimum*, *Mentha*, *Lavandula* and *Dracocephalum*), that were part of the study, leads to the assumption that those primers are most likely universal for the Nepetoideae and could be used for upcoming projects concerning flowering in this subfamily of Lamiaceae.

The general evaluation of the first GLOBOSA (as well as the DEFICIENS) sequences revealed that those genes are build up by seven exons, enclosing six introns, the latter makeup around two thirds of the complete length of the gene. The varying length of the genes (displayed in figure 34 and elaborated in results 4.5.1) derives from differing intron lengths, which might point to so far unknown regulatory regions within these introns. As example, the second intron of the C-class gene AGAMOUS has been demonstrated to possess regulatory functions [140]. The exon length and thus the amino acid sequence length seems to be strongly conserved and shows an interesting pattern of fixed amino acid numbers in the different subgenera of *Salvia* (see figure 35). The East Asian and New World amino acid length of GLOBOSA is reduced compared to the ones of Eurasian *Salvia* species. In the light of the global dispersal history of the genus *Salvia*, the shorter (213 amino acid) state appears to be the ancestral state, since the early spreads to East Asia and to the New World via the Bering land bridge, both in the late Oligocene (around 24 mya ago), show shorter exon lengths, compared to the more recent dispersals of Eurasian subgenera *Salvia*, *Sclarea* and *Heterosphace* (assumed around mid Miocene, 10 mya ago) that display enhanced exon lengths. The insertion of nucleotides to the GLOBOSA exons and the subsequent addition of amino acids seems to be a restricted event to the Western Old World.
The earlier mentioned proposed ancestral dispersal event to the New World, that putatively gave rise to both subgenera *Calosphace* and *Audibertia* (or genera *Lasemia* and *Ramona*) [112][36] could be viewed with caution as confirmed since the exons length of GLOBOSA is conserved in *Calosphace* (e.g. *Salvia elegans* or *Salvia hispanica*) and *Audibertia* (e.g. *Salvia columbariae*), pointing to a single dispersal event

and a subsequent restriction of amino acid length in the New World. To put it in other words, the exon length of GLOBOSA might be an antiproportional indicator of species age / subgenus age in genus *Salvia*.

The general evaluation of the proteins of the first B-class gene GLOBOSA sequences in *Salvia* species, revealed the typical structure of the MIKC-type MADS proteins, including the four typical MADS (M), intervening (I), K-Box (K) and C-terminal (C) protein domains, confirmed by the BLAST search of the amino acids sequence. The presence of those protein domains supports the conserved nature of the flowering gene proteins, as well as the working hypothesis for the duplication of GLOBOSA transcription factors (see figure 42).
The MADS-domain (the first 60 amino acids in the sequence) in all evaluated species is strongly conserved, suggesting binding to similar DNA regions respectively and thus most likely controlling the same downstream genes during the flowering process. The higher variable protein domains (namely K-Box and C-terminal) with their putative responsibility for protein dimerization and tetramerization and the deduced resulting variation of floral quartet complex formation might be a more likely explanation of *Salvia* floral structural differences, rather than DNA-region binding patterns of the transcription factors.

In contrast to the ARF1 fragment, the B-class genes have been sequenced as a whole, starting with the start codon and ending with the stop codon.
The paraphyletic nature of *Salvia* is once more confirmed using the trait-related floral developmental B-class gene GLOBOSA (see figure 36). This is the first time to specifically construct a gene-tree phylogeny based on an ABCDE gene of flowering that is demonstrably involved in floral development, especially in corolla and stamen development (see qPCR results, e.g. figure 18 or 23) and therefore coupled with the evolutionary background of the genus *Salvia*. The strongly significant posterior probabilities values of the GLOBOSA gene-tree, even for very closely related species, including species that are able to hybridize endorses this approach clearly.
Geographic regions are separated with 100% significance values and underline a presumed origin of *Salvia* in Southwest Asia and a subsequent early dispersal event to East Asia and the New World (24 to 22 million years ago) and a dispersal event to the Mediterranean Region (around 10 million years ago) [36][112].

The *Salvia pratensis*-clade (see figure 36 uppermost clade), that was named like that because all species included (*S. hierosolymitana*, *S. nemorosa* and *S. virgata*) have been considered as subspecies of *Salvia pratensis* in some determination keys [39], shows a clear resolution. Those synonyms are still listed in the plant list, although all species were lifted to the species taxon level, from the subspecies level. Hybrids of *S. pratensis* and *S. nemorosa* were already described in the 19^{th} century [138], with possible hybridization events of small *S. pratensis* flowers and large *S. nemorosa* flowers [34]. The illustration of the very similar shape of those two species, and only a shift of the flower size has been shown in figure 14 or figures 13a and 13b and previously discussed in 5.1. The statistically significant separation of such possibly hybridizing species proves the usability of this trait-related marker for the European region.

Applying this marker to New World *Salvia* species obviously blurs the interpretation of such results and might lead to false conclusions, because of the duplication event of GLOBOSA in the New World. The New World duplication has been deduced by three sequences of *Salvia elegans* and *Salvia farinacea*, respectively, where the two different GLOBOSA copies could be found or, to be more precise, were found by the phylogenetic analysis of the B-class GLOBOSA gene. This finding was surprising since in the gel there was no sign of double bands (other than DEFICIENS). In contrast to this, for the European species *Salvia pratensis* in total 13 sequences from different individuals were evaluated and the missing sign of duplication (because of monophyletic clustering in the GLOBOSA gene tree) strongly supports the assumption that this duplication event is limited to the New World *Salvia*. This has to be verified for other distant related *Salvia* species for Eurasia.
Nevertheless, it is a promising approach to use trait-related markers to get an easy and simultaneously more intrinsic grasp on the evolutionary relation of Old World *Salvia*. This is possible without laborious genomics experiments and therefore there is an encouragement for the usage of this method for upcoming Old World *Salvia* phylogenies. The approach should not be applied to American *Salvia* species, though, because of the proved duplication event. From a phylogenetic point of view this is a drawback, however, this result brings up an explanation for the radiation of the New World *Salvia*, based on a geographically restricted duplication event of a floral organ developmental gene.
The congruence of the radiation and the duplication of one of *the* major control genes for corolla and stamen identity and development may be interconnected, since the duplication of GLOBOSA could be a

facilitator of floral diversity, and hence for plant-pollinator co-evolution and ultimately species diversification. On the functional level, the formation of slightly varying floral quartet complexes by exchange of the single GLOBOSA protein by duplicated and neofunctionalized or subfunctionalized GLOBOSA proteins might be the working hypothesis for origin of the extension of the floral trait dimensions (especially in corolla, as presented in figure 42 and discussed in 5.1) and therefore be one of the driving forces for the manifold diversity of *Salvia*.

5.5 Encrypting B-class DEFICIENS

The very first experiment, the amplification of the B-class DEFICIENS gene by PCR (using primer from table 10), already indicated that the situation in this B-class gene is different from its heterodimer partner GLOBOSA.
In figure 39, the varying fragment sizes as well as the clearly visible amplification of two fragments suggests that in DEFICIENS something is going on.
The huge length variances that makeup up to 500 nucleotides (see results 4.7.2) can mainly ascribed to differences in intronic length. As in GLOBOSA the six introns that are flanked by seven exons account for around two thirds of the complete length of the gene. The differences in exon length ranges from 675 and 723 nucleotides (or 222 to 241 amino acids) and those rather huge variances are translated to the C-terminal domain of the transcription factor (as illustrated in figure 40) that is putatively responsible for transcriptional activation and in forming the tetramers. In the light of the mode of action for this flowering gene the varying length of the C-terminal domain might have an influence on how the distinct DEFICIENS proteins interact with the other members of a particular floral quartet complex.
As for GLOBOSA, the amino acid sequence of the MADS protein domain (the first 60 amino acids) of DEFICIENS displays a high conversation supporting the assumption that the downstream controlled genes by this transcription factor are similar in all investigated species.

As in the results, for the discussion it is presupposed that the topology of the DEFICIENS gene-tree is based on highest (mostly 100%) posterior probability significance values (see figure 41).
The DEFICIENS gene shows a clear duplication, since the phylogenetic gene tree demonstrates two main clusters, that both include sequences from all geographic regions, with an exception and absences of East Asian *Salvia* species in the DEFICIENS variant 1 (upper) cluster, an observation that can be explained with the limited data set for this region.
The DEFICIENS variant 2 (lower) cluster resembles the topology of trait-independent markers from other publications as well as the trait-related GLOBOSA gene tree presented in this thesis (see figure 36) concerning the evident split of geographic regions. The difference is here that the *Rosmarinus* / *Perovskia* cluster is closer related to the New World / East Asian *Salvia* cluster and is delineated from European *Salvia* species (other than in GLOBOSA and plastidic-based trees). As for GLOBOSA, for the B-class DEFICIENS gene, different colonies of the same individual were sequenced and evaluated, which led to more ambiguous results compared to the other B-class gene. As example, the lower DEFICIENS variant 2 cluster contains two sequences of *Salvia officinalis* and *Salvia lavandulifolia*, whereby the two copies of *Salvia officinalis* do not cluster together, since the *Salvia officinalis* K1 shows a closer relation to the *Salvia lavandulifolia* cluster, pointing to a not so clear species delineation of this gene for the mentioned species. This pattern might be due to the general possible hybridization events in those species or to a general less discriminative power of this gene.

Generally, the relationships depicted in this gene-tree should be handled with care, since the separation into two clusters leads the interpretation of species taxa relation slightly ad absurdum.
However, the results are interesting in the context of genetic biogeography. The main separation of the two DEFICIENS genes in the genus *Salvia* has occurred at least 25 mya ago, approximated by the dispersal events to the New World and to East Asia and the occurring split of the genus [36][112]. The presence of both gene copies in all geographic regions confirms its presence in the MRCA or the early representatives of *Salvia* in Southwest Asia and also the duplication of this gene before the worldwide dispersal.
This duplication might be more ancient as well (as elaborated below), however, the duplication in DEFICIENS variant 1 of European *Salvia pratensis* took place after geographic separation of the genus and can therefore be dated as relatively recent, after the assumed dispersal to Central Europe (which would be 9.8 mya to 6,8 mya, according to [112]). Comparable recent duplication event cannot be excluded for

the other regions (the New World and East Asia), since the sequence data set was limited, but the double peaks in American DEFICIENS variant 1 transcripts in the qPCR experiment points in this direction.
It has to be evaluated in the future whether the closest related genera and species of the genus *Salvia* sensu lato, that served as outgroups for the provided gene tree (e.g. *Lavandula* or *Ocimum*), display those duplication events as well. To clear the question whether the duplication of DEFICIENS only took place in *Salvia* (which would be very unlikely though, since the early split of DEFICIENS / APETALA3 has been demonstrated several times (reviewed in [90])), will help to explain the floral diversity in Lamiaceae or *Salvia* in particular.

However, to pinpoint the phylogenetic origin of the DEFICIENS duplication event, a more intense evaluation of closely related genera of *Salvia* needs to be examined. With the information gathered in this thesis the duplication rests at the base of the genus *Salvia* sensu lato, since the monophyly of *Salvia* sensu lato is inviolated. A clustering of *Salvia* sensu lato species with other genera from the Nepetoideae, leading to separation from the other DEFICIENS cluster containing *Salvia* sensu lato species would shift and confirm the DEFICIENS duplication event to an earlier branching of the Nepetoideae. Furthermore the presence of East Asian *Salvia* species in both DEFICIENS clusters needs to be proven, because so far the dataset is limited and the East Asian *Salvia* sequences are restricted to the DEFICIENS variant 1 cluster.

5.6 ABCDE, or about plan B of propagation in flowering plants

The relationship of the ABCDE class genes of flowering, which was inferred from protein sequences of *Arabidopsis thaliana* (a rosid outgroup), *Sesamum indicum* (a Lamiales outgroup) and finally the *Salvia* sequences shows a clear and statistically supported pattern.
As illustrated in figure 43, there are two main clusters, namely a B-class gene cluster and an ACDE gene cluster that are separated by a 99,7% posterior probability value.
Within both B-class clusters, the respective *Arabidopsis thaliana* protein sequence serves as outgroup, proving that the observed duplication events (as shown in figure 36 and 41) in *Salvia* are most likely limited to this genus. However, as elaborated before, the time point of the DEFICIENS duplication needs to be pinpointed more exactly.

The strict separation of the B-class genes from the other floral transcription factors, points to an ancestral important functional interrelation of corolla and stamen: (1) attraction of pollinators, that were present long before flowering plants [141] (2) and the spread of the haploid stage of flowering plant development by insects. Even further, certain processes of flowering plant sexuality are partly beyond the strict control of the plants and naturally those processes includes the involvement of unrelated species (like insects or birds).

The phylogenetic split of the ABCDE genes amazingly fits to a putative subdivision of plant sexuality into intrinsic and extrinsic processes.
There are intrinsic processes (represented by ACDE class genes) that include procedures, which take place physically within the plant, like initial growth of the flower and coverage of the more sensitive parts by the calyx or the development of carpel, fruits and seeds in later stages. And naturally the production of pollen and the development of the corolla takes place within the plant as well, however the *meaning* of these processes and structures shifts to outside the seemingly closed system “flower” or even ”plant”. Those organs (corolla and stamen) ought to initiate extrinsic processes (represented by B-class genes) that evoke the cooperation with pollinators reacting to a corolla color or to a pollination mechanism. The mainly B-class gene controlled development of corolla and stamen structures induces processes that evades the control of the plant for some extant, by shifting the spatial mode of action physically outside of the plant. This includes the uptake and processing of a signal (the corolla shape or color) by an outsourced organism as well as the entrusting of the precious haploid cargo to that organism, that could be possibility lost for ever. However, this ”trust” (or mutual interaction) was built up over millions of years in the course of evolution. An early separation of these intrinsic and extrinsic floral developmental processes can be read from this data.

5.7 Gretchenfrage - "how do you feel about species?"

5.7.1 Figuring out a species concept – walking on penrose stairs

Trying to find a species concept or even *the* concept of species is comparable to walking on penrose stairs as illustrated in M.C. Escher's "ascending and descending"; falsely assuming reaching higher ground or an approximation with more and more seemingly sophisticate ideas, while in reality infinite circling around the object of thought eventually results in loosing sight on it.

5.7.2 Application of concepts

"*No one definition has as yet satisfied all naturalists; yet every naturalist knows vaguely what he means when he speaks of a species.*"
- from Charles Darwins "The origin of species" [142]

Consequently, *no new* plant species concept should be given here, instead an advice on what should be considered when describing or working plant "species", which might have an increase in usefulness rather than execute another circling on the penrose stairs. In retrospective on this thesis the genus *Salvia* shall be used for elaboration on this issue.

Charles Darwin, though consequently avoiding a species concept still provided a nomenclatural framework and their levels ascend from (1) small differences of single plant individuals, to (2) lesser varieties, to (3) well-marked varieties, to (4) subspecies and ultimately (5) the rank of species [142]. And almost "naturally" the transitions between these terms / states are fluent. Additionally Darwin stressed that a clear line between the definitions he called sub-species and species has not been drawn or provided or defined by biologists (or philosophers).
Using the example of the "species" *Salvia officinalis* and its immediate "inter- (or intra-(?)) specific environment", this transition can be recognized and depicted on the ecological, morphological, but also on the anthropogenic based semantic level. The ability to hybridize and the formation of intermediate hybrids has been shown in a large scaled study of *S. officinalis* and *S. fruticosa* [139], and at the same time these species could be separated from each other when looking at the trait-related ARF1 marker (see figure 33). Also, the separation of the closely related species *S. officinalis* and *Salvia lavandulifolia* in the ARF1 and GLOBOSA phylogenetic trees (see figure 33 and 36) as well as in the limited shift in the statistical evaluation of flower geometry along the size-based x-axis shows both, the similarity as well as the possibility to successfully separate those similarities.
Additionally, an anthropogenical induced boundary comes into play here, because *S. lavandulifolia* is also ascribed as subspecies of *S. officinalis* (both can be found in the plant list, [37]).

To quote Mr. Charles Darwin the last time:
It should be remembered that systematists are far from pleased at finding variability in important characters, and that there are not many men who will laboriously examine internal and important organs, and compare them in many specimens of the same species. [142]

Though we live in a world of permanent flow, taxonomy relies to a large degree on framing nature to a static state, which obviously is a paradox, but the necessity of static depiction of something unstatic is inevitable. The same thing is true for phylogeny, here molecular snapshots are as well conducted to explain a world which is in endless flow. The clear dichotomous ordering of species results in a taxonomical sigh of relief, however, the fact, that such illustration of nature only exists in our mind should always be kept in mind.

For biodiversity investigations a holistic view on the organisms (or populations) of interest should get the greatest attention; for practical applications plant floral traits should be examined, with focus on the identification of the individuals and their allocation to a term that is generally referred to as "species". A term that might be under biological or philosophical revision (that kind of ironically depicts the mode of evolution), while the anthropogenic driven goal of those revisions is ultimately a temporarily framed picture, which is the inevitable basis of practical and scientific considerations.

6 Summary and outlook

The indented holistic approach covered morphological, developmental and phylogenetical investigations. The morphological examination included the identification of reference plants on the basis of their floral characteristics using identification literature. This was extended by statistic description of floral features of the available *Salvia* individuals. Based on the identification and their statistic evaluation those single *Salvia* individuals became the representatives of "their" species taxa for downstream experiments.
Measurements of *Salvia* floral traits have been performed in the past and though measurements for single species taxa were carried out extensively, limitations of those approaches often have been the restriction to a geographic region or limited downstream evaluations.
Here, the floral measurements of cultivated species of genus *Salvia* from Europe, East Asia and the New World have been conglomerated and a statistic evaluation of those traits were combined into a global context.
The scope of the study addressed the question, whether a morphological classification can be achieved by evaluation of *Salvia* corolla floral traits.
The statistical niches for closely related species as well as the dispersal in the whole PCA plot displays both, the practical usability in a closely related species context as well as for broadened worldwide applications. However, a statistical deduced geographic separation could not be achieved because of homologous floral geometries that are independent of geographic origin. Though already informative, this approach is far from exhausted, since (1) only a small fraction of *Salvia* species has been examined yet and (2) more measure sites or patterns, including stamen structure or corolla color, could be integrated into forthcoming measurements of *Salvia* species taxa.

- Measurements and statistical evaluation of *Salvia* floral traits appears to be a suitable tool to classify the species of this genus

The developmental background of the formation of before measured floral structures has been evaluated in-depth for model plants (like *Arabidopsis thaliana* or *Antirrhinum majus*), however, not transmitted to the species rich Lamiaceae family, that includes the genus *Salvia*.
The question from this thesis' scope includes the transmission of the flowering model to the genus *Salvia* by applying quantitative PCR experiments in order to investigate the expression patterns in the context of flowering.
The idea to separate the single floral organs specifies the general statement of floral gene activity to the distinct building blocks of the flower, allowing more detailed conclusions. This concept and its practical approach generally verified the pattern of the ABCDE model in *Salvia* but also highlighted the alternating expression during development for some genes. For example, C-class AGAMOUS is predominantly expressed in the inner whorls of closed flowers, with exception of the ovaries of *Salvia pratensis*. In comparison, E-class gene SEPALLATA3 showed higher expression for all organs of opened flowers, with exception of calyx and corolla of *Salvia elegans*. The B-class gene GLOBOSA displays a vast elevation of expression in its (coming from the ABCDE model) expected organs, namely corolla and stamen, however surprisingly, the expression even increased after the development of corolla was apparently completed by opening the flowers.
Taken together, the general pattern resembles the ABCDE model, however, the before mentioned alternations imply the importance of separating the floral building blocks to get a deeper insight into the development of the flower. The closed development tissue was derived from a relative late stage of flowering, since for example the calyx was already at almost complete size, hence for further experiments an earlier developmental stage should be included. However, here the clear separation of floral organs might be difficult, making in-situ hybridization experiments for small flower buds of *Salvia* an alternative. Additionally, other candidate developmental genes could be added to further complete the view.

- The ABCDE model of flowering was successfully transferred to species of the genus *Salvia*

The full length amplification of the B-class developmental genes GLOBOSA and DEFICIENS for the genus *Salvia* and other genera of the Lamiaceae has been successful. The phylogenetically evaluated multiple sequence alignments yielded highly significant phylogenetic trees that endorse the topology of non-trait related investigations (e.g. using barcodes) and also separates closest related species, that are naturally able to hybridize (like *Salvia pratensis* and *Salvia nemorosa*).
The question that was asked in this context at the beginning of this thesis addressed the issue whether it is possible to usw such trait-related markers to infer relationships.

However, since both genes display a certain degree of duplication, deduction of relationships should be cherished with care. Especially for DEFICIENS a full gene duplication for *Salvia* could be demonstrated. The limitation of gene duplication in GLOBOSA to the New World allows for some astonishing conclusions to be drawn.
The unraveled different degrees of duplication have been put into a model that might serve as an answer to the species richness of *Salvia* in the New World and is grounded on the assumed multiplication of MADS-MIKC transcription factor tetramerization allowing the possible development of extended floral structures, pushing the speciation process that subsequently gives rise to new species. The additional evaluation of other European and New World *Salvia* species should help to strengthen this model.

- B-class gene phylogeny provides both, a robust trait-related phylogenetic interrelation of the genus *Salvia*, and a model for the radiation of this genus in the New World

The ABCDE phylogeny including *Salvia* and rosid *Arabidopsis thaliana* amino acid sequences strengthens this proposed model, because *Arabidopsis thaliana* B-class genes (PISTILLATA and APETALA3) form outgroups for the *Salvia* GLOBOSA and DEFICIENS clusters respectively, indicating that the shown duplication is at least restricted to Lamiales. The restriction of duplication of GLOBOSA to *Salvia* New World is evident, however, the exact "phylogenetic origin" of the DEFICIENS duplication has yet to be determined, since the split is evident for the whole genus *Salvia*. Closely related genera of *Salvia* needs to be examined concerning DEFICIENS duplication to pinpoint the exact origin of gene duplication.
In this thesis floral morphological, floral developmental and floral genetic approaches have been interwoven and the genus *Salvia* used to apply the attempt of a holistic view. Naturally, additional approaches, like e.g. karyotyping which would be of special interest in plant families with frequent hybridization events, could be incorporated in the future to perfect the view.

7 References

[1] H. Schenk A. F. W. Schimper E. Strasburger, F. Noll. *Strasburger – Lehrbuch der Pflanzenwissenschaften*, volume 37. Springer-Verlag, Berlin Heidelberg, 2014.

[2] Charles Lyell. *Principles of geology*, volume 1. Murray, London, 1830.

[3] G Brent Dalrymple. *The age of the Earth*. Stanford University Press, 1994.

[4] Madeline C Weiss, Filipa L Sousa, Natalia Mrnjavac, Sinje Neukirchen, Mayo Roettger, Shijulal Nelson-Sathi, and William F Martin. The physiology and habitat of the last universal common ancestor. *Nature microbiology*, 1(9):1–8, 2016.

[5] Susana Magallón, Sandra Gómez-Acevedo, Luna L. Sánchez-Reyes, and Tania Hernández-Hernández. A metacalibrated time-tree documents the early rise of flowering plant phylogenetic diversity. *New Phytologist*, 207(2):437–453, 2015.

[6] Evolutionary Biology, New Haven, Evolutionary Biology, New Haven, and Volunteer Boulevard. Heterogeneous Rates of Molecular Evolution and Diversification Could Explain the Triassic Age Estimate for Angiosperms. 64(5):869–878, 2015.

[7] Jared Diamond. *Arm und Reich: Die Schicksale menschlicher Gesellschaften*, volume 10. Fischer Taschenbuch, 2006.

[8] Gregor Mendel. *Versuche über Pflanzenhybriden*, volume IV. Band. Naturforschender Verein in Brünn, 1865.

[9] Joseph Gottlieb Kölreuter. *Vorläufige Nachricht von einigen das Geschlecht der Pflanzen betreffenden Versuchen und Beobachtungen*. Gleditsch, 1761.

[10] Christian Konrad Sprengel. *Das entdeckte Geheimnis der Natur im Bau und in der Befruchtung der Blumen.*, volume 1. Vieweg, Berlin, 1793.

[11] O. Schmeil and J. Fitschen. *Flora Von Deutschland und Angrenzender Länder*, volume 95. Quelle Meyer Verlag, Wiebelsheim, 2011.

[12] Murray W Nabors. *Introduction to botany*, volume 1. Pearson, San Francisco, 2004.

[13] Johann Wolfgang Goethe. *Die Metamorphose der Pflanzen*, volume 5. Verlag Freies Geistesleben, Stuttgart, 1985.

[14] John J. W. Rogers and M. Santosh. *Continents and Supercontinents*, volume 1. Oxford University Press, 2004.

[15] William L Crepet. The fossil record of angiosperms: Requiem or renaissance? *Annals of the Missouri Botanical Garden*, 95(1):3–33, 2008.

[16] Ge Sun, David L Dilcher, Shaoling Zheng, and Zhekun Zhou. In search of the first flower: a Jurassic angiosperm, *Archaefructus*, from northeast China. *Science*, 282(5394):1692–1695, 1998.

[17] Qiang Fu, Jose Bienvenido Diez, Mike Pole, Manuel García Ávila, Zhong Jian Liu, Hang Chu, Yemao Hou, Pengfei Yin, Guo Qiang Zhang, Kaihe Du, and Xin Wang. An unexpected noncarpellate epigynous flower from the Jurassic of China. *eLife*, 7:1–24, 2018.

[18] Mario Coiro, James A. Doyle, and Jason Hilton. How deep is the conflict between molecular and fossil evidence on the age of angiosperms? *New Phytologist*, pages 83–99, 2019.

[19] Dmitry D Sokoloff, Margarita V Remizowa, Elena S El, Paula J Rudall, and Richard M Bateman. Supposed Jurassic angiosperms lack pentamery, an important angiosperm-specific feature. *New Phytologist*, 2019.

[20] Hervé Sauquet, Maria Von Balthazar, Susana Magallón, James A. Doyle, Peter K. Endress, Emily J. Bailes, Erica Barroso De Morais, Kester Bull-Hereñu, Laetitia Carrive, Marion Chartier, Guillaume Chomicki, Mario Coiro, Raphaël Cornette, Juliana H.L. El Ottra, Cyril Epicoco, Charles S.P. Foster, Florian Jabbour, Agathe Haevermans, Thomas Haevermans, Rebeca Hernández, Stefan A. Little, Stefan Löfstrand, Javier A. Luna, Julien Massoni, Sophie Nadot, Susanne Pamperl, Charlotte Prieu, Elisabeth Reyes, Patrícia Dos Santos, Kristel M. Schoonderwoerd, Susanne Sontag, Anaëlle Soulebeau, Yannick Staedler, Georg F. Tschan, Amy Wing-Sze Leung, and Jürg Schönenberger. The ancestral flower of angiosperms and its early diversification. *Nature Communications*, 8(May), 2017.

[21] Klaus Kubitzki, JG Rohwer, and V Bittrich. *The families and genera of vascular plants*, volume 7. Springer, 2004.

[22] P.F. Stevens. Angiosperm phylogeny website. version 14, july 2017 [and more or less continuously updated since]. `http://www.mobot.org/MOBOT/research/APweb/`, 2017. Accessed: 2020-03-30.

[23] Bo Li, Philip D Cantino, Richard G Olmstead, Gemma LC Bramley, Chun-Lei Xiang, Zhong-Hui Ma, Yun-Hong Tan, and Dian-Xiang Zhang. A large-scale chloroplast phylogeny of the lamiaceae sheds new light on its subfamilial classification. *Scientific reports*, 6(1):1–18, 2016.

[24] Bo Li and Richard G Olmstead. Two new subfamilies in Lamiaceae. *Phytotaxa*, 313(2):222–226, 2017.

[25] Grant T Godden, Taliesin J Kinser, Pamela S Soltis, and Douglas E Soltis. Phylotranscriptomic analyses reveal asymmetrical gene duplication dynamics and signatures of ancient polyploidy in mints. *Genome biology and evolution*, 11(12):3393–3408, 2019.

[26] Bryan T. Drew and Kenneth J. Sytsma. Phylogenetics, biogeography, and staminal evolution in the tribe Mentheae (Lamiaceae). *American Journal of Botany*, 99(5):933–953, 2012.

[27] EM Reid and MEJ Chandler. The bembridge flora—catalogue of cainozoic plants in the department of geology. *British Museum (Natural History) London*, page 206, 1926.

[28] Marcela Martínez-Millán. Fossil record and age of the asteridae. *The Botanical Review*, 76(1):83–135, 2010.

[29] Hui Li, Jingrui Li, Hongtong Bai, Lei Shi, and Huafang Wang. The complete chloroplast genome sequence of *Lavandula angustifolia* (Lamiaceae), an introduced aromatic plant growing in China. *Mitochondrial DNA Part B: Resources*, 4(1):1352–1353, 2019.

[30] Miaomiao Li, Dongfen Zhang, Qiang Gao, Yingfeng Luo, Hui Zhang, Bin Ma, Chunhai Chen, Annabel Whibley, Yu'e Zhang, and Qun Li Han Guo Junhui Li Yanzhai Song Yue Zhang Lucy Copsey Yan Li Xiuxiu Li Ming Qi Jiawei Wang Yan Chen Dan Wang Jinyang Zhao Guocheng Liu Bin Wu Lili Yu Chunyan Xu Jiang Li Shancen Zhao Yijing Zhang Songnian Hu Chengzhi Liang Ye Yin Enrico Coen Yongbiao Xue Cao, Yinghao. Genome structure and evolution of *Antirrhinum majus L. Nature plants*, 5(2):174–183, 2019.

[31] Benoît Boachon, C. Robin Buell, Emily Crisovan, Natalia Dudareva, Nicolas Garcia, Grant Godden, Laura Henry, Mohamed O. Kamileen, Heather Rose Kates, Matthew B. Kilgore, Benjamin R. Lichman, Evgeny V. Mavrodiev, Linsey Newton, Carlos Rodriguez-Lopez, Sarah E. O'Connor, Douglas Soltis, Pamela Soltis, Brieanne Vaillancourt, Krystle Wiegert-Rininger, and Dongyan Zhao. Phylogenomic Mining of the Mints Reveals Multiple Mechanisms Contributing to the Evolution of Chemical Diversity in Lamiaceae. *Molecular Plant*, 11(8):1084–1096, 2018.

[32] Jay B Walker, Kenneth J Sytsma, Jens Treutlein, and Michael Wink. *Salvia* (Lamiaceae) is not monophyletic: implications for the systematics, radiation, and ecological specializations of *Salvia* and tribe Mentheae. *American Journal of Botany*, 91(7):1115–1125, 2004.

[33] Jay B. Walker and Kenneth J. Sytsma. Staminal evolution in the genus *Salvia* (Lamiaceae): Molecular phylogenetic evidence for multiple origins of the staminal lever. *Annals of Botany*, 100(2):375–391, 2007.

[34] Regine Claßen-Bockhoff, Thomas Speck, Enikö Tweraser, Petra Wester, Sascha Thimm, and Martin Reith. The staminal lever mechanism in *Salvia* L. (Lamiaceae): A key innovation for adaptive radiation? *Organisms Diversity and Evolution*, 4(3):189–205, 2004.

[35] Bryan T. Drew, Jesús Guadalupe González-Gallegos, Chun Lei Xiang, Ricardo Kriebel, Chloe P. Drummond, Jay B. Walker, and Kenneth J. Sytsma. *Salvia* united: The greatest good for the greatest number. *Taxon*, 66(1):133–145, 2017.

[36] Maria Will and Regine Claßen-Bockhoff. Time to split *Salvia* s.l. (Lamiaceae) – New insights from Old World *Salvia* phylogeny. *Molecular Phylogenetics and Evolution*, 109:33–58, 2017.

[37] The Plant List (2013). Version 1.1. `http://www.theplantlist.org/`. Last accessed: 2020-05-30.

[38] Regine Claßen-Bockhoff. Stamen construction , development and evolution in *Salvia* s.l. *Natural Volatiles and Essential Oils*, 4(4):28–48, 2017.

[39] Gustav Hegi. *HEGI - Illustrierte Flora von Mitteleuropa*, volume Band V, 4. Teil. Carl Hanser Verlag, München, 1927.

[40] Maria Will and Regine Claßen-Bockhoff. Why Africa matters: Evolution of Old World *Salvia* (Lamiaceae) in Africa. *Annals of Botany*, 114(1):61–83, 2014.

[41] Lawrence J Hobbie. Auxin: Molecular genetic approaches in *Arabidopsis*. *Plant Physiology and Biochemistry*, 36(1-2):91–102, 1998.

[42] John William Chandler. Auxin Response Factors. *Plant, cell & environment*, 39(5):1014–1028, 2016.

[43] Si-Bei Li, Zong-Zhou Xie, Chun-Gen Hu, and Jin-Zhi Zhang. A review of Auxin Response Factors (ARFs) in plants. *Frontiers in plant science*, 7:47, 2016.

[44] Elmar Weiler and Lutz Nover. *Allgemeine und molekulare Botanik*, volume 11. Georg Thieme Verlag, Stuttgart, 2008.

[45] Eva Zažímalová, Jan Petrášek, and Eva Benková. *Auxin and its role in plant development*. Springer-Verlag, Heidelberg, 2014.

[46] Heidi Szemenyei, Mike Hannon, and Jeff A Long. Topless mediates auxin-dependent transcriptional repression during *Arabidopsis* embryogenesis. *Science*, 319(5868):1384–1386, 2008.

[47] Tom Guilfoyle. Sticking with auxin. *Nature*, 446(7136):621–622, 2007.

[48] Renhou Wang and Mark Estelle. Diversity and specificity: Auxin perception and signaling through the TIR1/AFB pathway. *Current opinion in plant biology*, 21:51–58, 2014.

[49] Emilie Varaud, Florian Brioudes, Judit Szécsi, Julie Leroux, Spencer Brown, Catherine Perrot-Rechenmann, and Mohammed Bendahmane. Auxin Response Factor 8 regulates *Arabidopsis* petal growth by interacting with the bHLH transcription factor BIGPETALp. *The Plant Cell*, 23(3):973–983, 2011.

[50] Marc Goetz, Adam Vivian-Smith, Susan D Johnson, and Anna M Koltunow. Auxin Response Factor 8 is a negative regulator of fruit initiation in *Arabidopsis*. *The Plant Cell*, 18(8):1873–1886, 2006.

[51] Punita Nagpal, Christine M Ellis, Hans Weber, Sara E Ploense, Lana S Barkawi, Thomas J Guilfoyle, Gretchen Hagen, José M Alonso, Jerry D Cohen, Edward E Farmer, Joseph R Ecker, and Jason W Reed. Auxin response factors ARF6 and ARF8 promote jasmonic acid production and flower maturation. *Development (Cambridge, England)*, 132(18):4107–4118, 2005.

[52] Cédric Finet, Chloé Fourquin, Marion Vinauger, Annick Berne-Dedieu, Pierre Chambrier, Sandrine Paindavoine, and Charles P Scutt. Parallel structural evolution of Auxin Response Factors in the angiosperms. *The Plant Journal*, 63(6):952–959, 2010.

[53] M F Wu, Q Tian, and J W Reed. Arabidopsis microRNA167 controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. *Development*, 133(21):4211–4218, 2006.

[54] Ning Liu, Shan Wu, Jason Van Houten, Ying Wang, Biao Ding, Zhangjun Fei, Thomas H Clarke, Jason W Reed, and Esther Van Der Knaap. Down-regulation of Auxin Response Factors 6 and 8 by microRNA167 leads to floral development defects and female sterility in tomato. *Journal of experimental botany*, 65(9):2507–2520, 2014.

[55] Paul H Reeves, Christine M Ellis, Sara E Ploense, Miin-Feng Wu, Vandana Yadav, Dorothea Tholl, Aurore Chételat, Ina Haupt, Brian J Kennerley, Charles Hodgens, et al. A regulatory network for coordinated flower maturation. *PLoS genetics*, 8(2), 2012.

[56] Xiaodong Liu, Jian Huang, Yao Wang, Kanhav Khanna, Zhixin Xie, Heather A Owen, and Dazhong Zhao. The role of floral organs in carpels, an *Arabidopsis* loss-of-function mutation in microRNA160a, in organogenesis and the mechanism regulating its expression. *The Plant Journal*, 62(3):416–428, 2010.

[57] Zhichao Xu, Aijia Ji, Jingyuan Song, and Shilin Chen. Genome-wide analysis of Auxin Response Factor gene family members in medicinal model plant *Salvia miltiorrhiza*. *Biology open*, 5(6):848–57, 2016.

[58] Enrico S Coen and Elliot M Meyerowitz. The war of the whorls: genetic interactions controlling flower development. *Nature*, 353(6339):31–37, 1991.

[59] John L Bowman, David R Smyth, and Elliot M Meyerowitz. Genetic interactions among floral homeotic genes of *Arabidopsis*. *Development*, 112(1):1–20, 1991.

[60] Elliot M Meyerowitz, David R Smyth, and John L Bowman. Abnormal flowers and pattern formation in floral. *Development*, 106(2):209–217, 1989.

[61] Günter Theißen, Rainer Melzer, and Florian Rümpler. MADS-domain transcription factors and the floral quartet model of flower development: linking plant development and evolution. *Development*, 143(18):3259–3271, 2016.

[62] Beverley Glover. *Understanding flowers and flowering second edition*. Oxford University Press, 2007.

[63] Günter Theissen and Heinz Saedler. Floral quartets. *Nature*, 409(6819):469–471, 2001.

[64] Gerco C Angenent and Lucia Colombo. Molecular control of ovule development. *Trends in plant science*, 1(7):228–232, 1996.

[65] Gary S. Ditta, Elvira Baumann, Martin F. Yanofsky, Soraya Pelaz, and Ellen Wisman. B and C floral organ identity functions require SEPALLATA MADS-box genes. *Nature*, 405(6783):200–203, 2000.

[66] Günter Theißen, Jan T Kim, and Heinz Saedler. Classification and phylogeny of the MADS-box multigene family suggest defined roles of MADS-box gene subfamilies in the morphological evolution of eukaryotes. *Journal of Molecular Evolution*, 43(5):484–516, 1996.

[67] Kerstin Kaufmann, Rainer Melzer, and Günter Theißen. MIKC-type MADS-domain proteins: Structural modularity, protein interactions and network evolution in land plants. *Gene*, 347(2 SPEC. ISS.):183–198, 2005.

[68] John L Bowman. Making cauliflower out of *Arabidopsis*: The specification of floral meristem identity. *Flowering Newsletter*, (14):7–19, 1992.

[69] M Alejandra Mandel, Cindy Gustafson-Brown, Beth Savidge, and Martin F Yanofsky. Molecular characterization of the *Arabidopsis* floral homeotic gene APETALA1. *Nature*, 360(6401):273–277, 1992.

[70] Peter Huijser, Joachim Klein, WE Lönnig, Hans Meijer, Heinz Saedler, and Hans Sommer. Bracteomania, an inflorescence anomaly, is caused by the loss of function of the MADS-box gene SQUAMOSA in *Antirrhinum majus*. *The EMBO journal*, 11(4):1239–1249, 1992.

[71] Cindy Gustafson-Brown, Beth Savidge, and Martin F Yanofsky. Regulation of the *Arabidopsis* floral homeotic gene APETALA1. *Cell*, 76(1):131–143, 1994.

[72] Jens F Sundström, Naomi Nakayama, Kristina Glimelius, and Vivian F Irish. Direct regulation of the floral homeotic APETALA1 gene by APETALA3 and PISTILLATA in *Arabidopsis*. *The Plant Journal*, 46(4):593–600, 2006.

[73] Medard Ng and Martin F Yanofsky. Activation of the Arabidopsis B-class homeotic genes by APETALA1. *The Plant Cell*, 13(4):739–753, 2001.

[74] Veronica Gregis, Alice Sessa, Carmen Dorca-Fornell, and Martin M Kater. The *Arabidopsis* floral meristem identity genes AP1, AGL24 and SVP directly repress class B and C floral homeotic genes. *The Plant Journal*, 60(4):626–637, 2009.

[75] Detlef Weigel and Elliot M Meyerowitz. Activation of floral homeotic genes in *Arabidopsis*. *Science*, 261(5129):1723–1726, 1993.

[76] K Diane Jofuku, BG Den Boer, Marc Van Montagu, and Jack K Okamuro. Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *The Plant Cell*, 6(9):1211–1225, 1994.

[77] Xuemei Chen. A microRNA as a translational repressor of APETALA2 in *Arabidopsis* flower development. *Science*, 303(5666):2022–2025, 2004.

[78] Koji Goto and Elliot M Meyerowitz. Function and regulation of the *Arabidopsis* floral homeotic gene PISTILLATA. *Genes & Development*, 8(13):1548–1560, 1994.

[79] Emma Keck, Paula McSteen, Rosemary Carpenter, and Enrico Coen. Separation of genetic functions controlling organ identity in flowers. *The EMBO Journal*, 22(5):1058–1066, 2003.

[80] Hans Sommer, Jose-Pio Beltran, Peter Huijser, Heike Pape, Wolf-Ekkehard Lönnig, H Saedler, and Zs Schwarz-Sommer. Deficiens, a homeotic gene involved in the control of flower morphogenesis in *Antirrhinum majus*: the protein shows homology to transcription factors. *The EMBO Journal*, 9(3):605–613, 1990.

[81] Thomas Jack, Laura L Brockman, and Elliot M Meyerowitz. The homeotic gene APETALA3 of *Arabidopsis thaliana* encodes a MADS box and is expressed in petals and stamens. *Cell*, 68(4):683–697, 1992.

[82] W Tröbner, Lucia Ramirez, Patrick Motte, Isabelle Hue, Peter Huijser, WE Lönnig, Heinz Saedler, Hans Sommer, and Zsuzsanna Schwarz-Sommer. GLOBOSA: A homeotic gene which interacts with DEFICIENS in the control of *Antirrhinum* floral organogenesis. *The EMBO Journal*, 11(13):4693–4704, 1992.

[83] Zsuzsanna Schwarz-Sommer, Isabelle Hue, P Huijser, PJ Flor, R Hansen, F Tetens, WE Lönnig, H Saedler, and H Sommer. Characterization of the *Antirrhinum* floral homeotic MADS-box gene DEFICIENS: evidence for DNA binding and autoregulation of its persistent expression throughout flower development. *The EMBO Journal*, 11(1):251–263, 1992.

[84] Hao Yu, Toshiro Ito, Yuanxiang Zhao, Jinrong Peng, Prakash Kumar, and Elliot M Meyerowitz. Floral homeotic genes are targets of gibberellin signaling in flower development. *Proceedings of the National Academy of Sciences*, 101(20):7827–7832, 2004.

[85] Brian McGonigle, Karim Bouhidel, and Vivian F Irish. Nuclear localization of the *Arabidopsis* APETALA3 and PISTILLATA homeotic gene products depends on their simultaneous expression. *Genes & Development*, 10(14):1812–1821, 1996.

[86] Rainer Melzer, Andrea Härter, Florian Rümpler, Sangtae Kim, Pamela S. Soltis, Douglas E. Soltis, and Günter Theißen. DEF- and GLO-like proteins may have lost most of their interaction partners during angiosperm evolution. *Annals of Botany*, 114(7):1431–1443, 2014.

[87] Sangtae Kim, Jin Koh, Mi-Jeong Yoo, Hongzhi Kong, Yi Hu, Hong Ma, Pamela S Soltis, and Douglas E Soltis. Expression of floral MADS-box genes in basal angiosperms: implications for the evolution of floral regulators. *The Plant Journal*, 43(5):724–744, 2005.

[88] Sangtae Kim, Mi Jeong Yoo, Victor A. Albert, James S. Farris, Pamela S. Soltis, and Douglas E. Soltis. Phylogeny and diversification of B-function MADS-box genes in angiosperms: Evolutionary and functional implications of a 260-million-year-old duplication. *American Journal of Botany*, 91(12):2102–2118, 2004.

[89] Yuannian Jiao, Norman J Wickett, Saravanaraj Ayyampalayam, André S Chanderbali, Lena Landherr, Paula E Ralph, Lynn P Tomsho, Yi Hu, Haiying Liang, and Douglas E. Soltis Sandra W. Clifton Scott E. Schlarbaum Stephan C. Schuster Hong Ma Jim Leebens-Mack Claude W. de Pamphilis Soltis, Pamela S. Ancestral polyploidy in seed plants and angiosperms. *Nature*, 473(7345):97–100, 2011.

[90] Helena Augusto Gioppato and Marcelo Carnier Dornelas. When Bs are better than As: the relationship between B-class MADS-box gene duplications and the diversification of perianth morphology. *Tropical Plant Biology*, 12(1):1–11, 2019.

[91] Laura M. Zahn, J. Leebens-Mack, C. W. DePamphilis, H. Ma, and G. Theissen. To B or not to B a flower: The role of DEFICIENS and GLOBOSA orthologs in the evolution of the angiosperms. *Journal of Heredity*, 96(3):225–240, 2005.

[92] Kelsey D Galimba, Jesús Martínez-Gómez, and Verónica S Di Stilio. Gene duplication and transference of function in the paleoap3 lineage of floral organ identity genes. *Frontiers in plant science*, 9:334, 2018.

[93] Michael D Purugganan. The MADS-box floral homeotic gene lineages predate the origin of seed plants: phylogenetic and molecular clock estimates. *Journal of molecular evolution*, 45(4):392–396, 1997.

[94] Martin F Yanofsky, Hong Ma, John L Bowman, Gary N Drews, Kenneth A Feldmann, and Elliot M Meyerowitz. The protein encoded by the *Arabidopsis* homeotic gene AGAMOUS resembles transcription factors. *Nature*, 346(6279):35–39, 1990.

[95] Toshiro Ito, Kian-Hong Ng, Tze-Soo Lim, Hao Yu, and Elliot M Meyerowitz. The homeotic protein AGAMOUS controls late stamen development by regulating a jasmonate biosynthetic gene in *Arabidopsis*. *The Plant Cell*, 19(11):3516–3529, 2007.

[96] Lei Li, Chuanyou Li, and Gregg A Howe. Genetic analysis of wound signaling in tomato. evidence for a dual role of jasmonic acid in defense and female fertility. *Plant Physiology*, 127(4):1414–1417, 2001.

[97] Desmond Bradley, Rosemary Carpenter, Hans Sommer, Nigel Hartley, and Enrico Coen. Complementary floral homeotic phenotypes result from opposite orientations of a transposon at the PLENA locus of *Antirrhinum*. *Cell*, 72(1):85–95, 1993.

[98] Brendan Davies, Patrick Motte, Emma Keck, Heinz Saedler, Hans Sommer, and Zsuzsanna Schwarz-Sommer. PLENA and FARINELLI: redundancy and regulatory interactions between two *Antirrhinum* MADS-box factors controlling flower development. *The EMBO Journal*, 18(14):4023–4034, 1999.

[99] Anusak Pinyopich, Gary S Ditta, Beth Savidge, Sarah J Liljegren, Elvira Baumann, Ellen Wisman, and Martin F Yanofsky. Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature*, 424(6944):85–88, 2003.

[100] Rebecca Favaro, Anusak Pinyopich, Raffaella Battaglia, Maarten Kooiker, Lorenzo Borghi, Gary Ditta, Martin F Yanofsky, Martin M Kater, and Lucia Colombo. MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *The Plant Cell*, 15(11):2603–2611, 2003.

[101] Richard GH Immink, Isabella AN Tonaco, Stefan De Folter, Anna Shchennikova, Aalt DJ van Dijk, Jacqueline Busscher-Lange, Jan W Borst, and Gerco C Angenent. SEPALLATA3: the 'glue' for MADS box transcription factor complex formation. *Genome biology*, 10(2):R24, 2009.

[102] Hua-Ying Fan, Yi Hu, Matthew Tudor, and Hong Ma. Specific interactions between the K domains of AG and AGLs, members of the MADS domain family of DNA binding proteins. *The Plant Journal*, 12(5):999–1010, 1997.

[103] Takashi Honma and Koji Goto. Complexes of MADS-box proteins are sufficient to convert leaves into floral organs. *Nature*, 409(6819):525–529, 2001.

[104] Susan L Urbanus, Stefan De Folter, Anna V Shchennikova, Kerstin Kaufmann, Richard GH Immink, and Gerco C Angenent. *In planta* localisation patterns of MADS domain proteins during floral development in *Arabidopsis thaliana*. *BMC Plant Biology*, 9(1):5, 2009.

[105] Ernst Mayr. What is a species, and what is not? *Philosophy of science*, 63(2):262–277, 1996.

[106] Brent D Mishler and John S Wilkins. The hunting of the SNaRC: a snarky solution to the species problem. 2018.

[107] David L Hull. Are species really individuals? *Systematic zoology*, 25(2):174–191, 1976.

[108] Charles Linnaeus. *Species plantarum*, volume 3. Impensis GC Nauk, 1799.

[109] Carolus Linnaeus. *Systema naturae*, volume 1. Stockholm Laurentii Salvii, 1758.

[110] Ernst Mayr. Animal species and evolution. *Animal species and evolution.*, 1963.

[111] M.F. Claridge, H.A. Dawah, and M.R. Wilson. *Species - The units of biodiversity*. Chapman Hall, 1997.

[112] Ricardo Kriebel, Bryan T. Drew, Chloe P. Drummond, Jesús G. González-Gallegos, Ferhat Celep, Mohamed M. Mahdjoub, Jeffrey P. Rose, Chun Lei Xiang, Guo Xiong Hu, Jay B. Walker, Emily M. Lemmon, Alan R. Lemmon, and Kenneth J. Sytsma. Tracking temporal shifts in area, biomes, and pollinators in the radiation of *Salvia* (sages) across continents: leveraging anchored hybrid enrichment and targeted sequence data. *American Journal of Botany*, 106(4):573–597, 2019.

[113] Joseph Sambrook, Edward F Fritsch, and Tom Maniatis. *Molecular cloning: a laboratory manual*. Number Ed. 2. Cold spring harbor laboratory press, 1989.

[114] J Patterson, B Chamberlain, and D Thayer. Finch tv version 1.4. 0. *Publ. by Authors*, 2004.

[115] Sudhir Kumar, Glen Stecher, and Koichiro Tamura. MEGA7: molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Molecular biology and evolution*, 33(7):1870–1874, 2016.

[116] J Allaire. RStudio: integrated development environment for R. *Boston, MA*, 537:538, 2012.

[117] Alexei J Drummond, Marc A Suchard, Dong Xie, and Andrew Rambaut. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and evolution*, 29(8):1969–1973, 2012.

[118] A Rambaut. Figtree 1.4. 2 software. *Institute of Evolutionary Biology, Univ. Edinburgh*, 2014.

[119] Mark Johnson, Irena Zaretskaya, Yan Raytselis, Yuri Merezhuk, Scott McGinnis, and Thomas L Madden. Ncbi blast: a better web interface. *Nucleic acids research*, 36(suppl_2):W5–W9, 2008.

[120] Herbal plant genomes. `http://herbalplant.ynau.edu.cn/html/Genomes/`. Last accessed: 2017-09-30.

[121] Promega M13 primers. `https://www.promega.de/products/cloning-and-dna-markers/molecular-biology-enzymes-and-reagents/puc_m13-sequencing-primers/?catNum=Q5401`. Last accessed: 2020-05-12.

[122] Yanfang Yang, Shuang Hou, Guanghong Cui, Shilin Chen, Jianhe Wei, and Luqi Huang. Characterization of reference genes for quantitative real-time PCR analysis in various tissues of *Salvia miltiorrhiza*. *Molecular Biology Reports*, 37(1):507–513, 2010.

[123] Anthony R Brach and Hong Song. efloras: New directions for online floras exemplified by the Flora of China project. *Taxon*, 55(1):188–192, 2006.

[124] Botanisches Institut KIT - nicklab.de. `https://www.nicklab.de`. Accessed: 2020-05-17.

[125] Santiago Benitez-Vieyra, Jessica Pérez-Alquicira, Federico D. Sazatornil, César A. Domínguez, Karina Boege, Rubén Pérez-Ishiwara, and Juan Fornoni. Evolutionary transition between bee pollination and hummingbird pollination in *Salvia*: Comparing means, variances and covariances of corolla traits. *Journal of Evolutionary Biology*, (April):1–11, 2019.

[126] Robert C Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–1797, 2004.

[127] Wenliang Wei, Xiaoqiong Qi, Linhai Wang, Yanxin Zhang, Wei Hua, Donghua Li, Haixia Lv, and Xiurong Zhang. Characterization of the sesame (*Sesamum indicum* L.) global transcriptome using illumina paired-end sequencing and development of EST-SSR markers. *BMC genomics*, 12(1):451, 2011.

[128] Raw sequence reads of *Salvia* species. `https://www.ncbi.nlm.nih.gov/bioproject/?term=Salvia`. Accessed: 2020-05-26.

[129] Karol Krak, Petr Vít, Jan Douda, and Bohumil Mandák. Development of 18 microsatellite markers for *Salvia pratensis*. *Applications in Plant Sciences*, 8(1):e11316, 2020.

[130] Ai-Xiang Dong, Hai-Bo Xin, Zi-Jing Li, Hui Liu, Yan-Qiang Sun, Shuai Nie, Zheng-Nan Zhao, Rong-Feng Cui, Ren-Gang Zhang, Quan-Zheng Yun, Xin-Ning Wang, Fatemeh Maghuly, Ilga Porth, Ri-Chen Cong, and Jian-Feng Mao. High-quality assembly of the reference genome for scarlet sage, *Salvia splendens*, an economically important ornamental plant. *GigaScience*, 7(7):1–10, 2018.

[131] Elizabeth A Schultz and George W Haughn. LEAFY, a homeotic gene that regulates inflorescence development in arabidopsis. *The Plant Cell*, 3(8):771–781, 1991.

[132] Detlef Weigel, John Alvarez, David R Smyth, Martin F Yanofsky, and Elliot M Meyerowitz. LEAFY controls floral meristem identity in *Arabidopsis*. *Cell*, 69(5):843–859, 1992.

[133] WA Emboden. Pollen morphology of the genus *Salvia* section Audibertia. *Pollen & Spores*, 6(2):527–536, 1964.

[134] Petra Wester and Regine Claßen-Bockhoff. Bird pollination in South African *Salvia* species. *Flora: Morphology, Distribution, Functional Ecology of Plants*, 201(5):396–406, 2006.

[135] N Kharazian. Morphometric study of some *Salvia* L. (Lamiaceae) species in Iran. *Scientific Journal of Biological Sciences*, 1(6):126–137, 2012.

[136] Laura P. Lagomarsino, Elisabeth J. Forrestel, Nathan Muchhala, and Charles C. Davis. Repeated evolution of vertebrate pollination syndromes in a recently diverged Andean plant clade. *Evolution*, 71(8):1970–1985, 2017.

[137] Jeff Ollerton, Ruben Alarcón, Nickolas M Waser, Mary V Price, Stella Watts, Louise Cranmer, Andrew Hingston, Craig I Peter, and John Rotenberry. A global test of the pollination syndrome hypothesis. *Annals of botany*, 103(9):1471–1480, 2009.

[138] Anton Kerner von Marilaun. *Pflanzenleben: bd. Geschichte der Pflanzen*, volume 2. Verlag des Bibliographischen Instituts, 1891.

[139] Ivan Radosavljević, Sandro Bogdanović, Ferhat Celep, Maja Filipović, Zlatko Satovic, Boštjan Surina, and Zlatko Liber. Morphological, genetic and epigenetic aspects of homoploid hybridization between *Salvia officinalis* L. and *Salvia fruticosa* Mill. *Scientific reports*, 9(1):3276, 2019.

[140] M. K. Deyholos. Separable Whorl-Specific Expression and Negative Regulation by Enhancer Elements within the AGAMOUS Second Intron. *The Plant Cell*, 12(10):1799–1810, 2000.

[141] Conrad C Labandeira and J John Sepkoski. Insect diversity in the fossil record. *Science*, 261(5119):310–315, 1993.

[142] Charles Darwin. *The origin of species by means of natural selection: or, the preservation of favored races in the struggle for life*. Penguin Harmondsworth, 1859.

8 Appendix

8.1 Additional samples for the ARF1 phylogeny

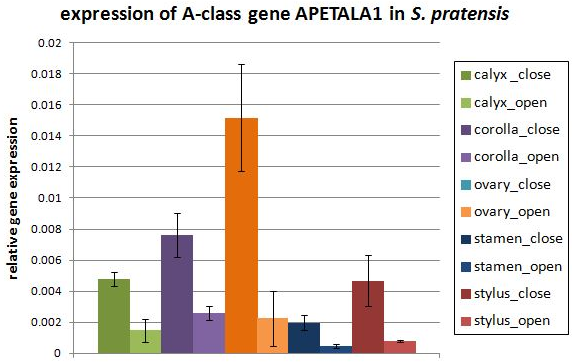
Table 13: Additional *Salvia* species and Lamiaceae outgroups mainly used for the Auxin Response Factor 1 phylogeny experiment (see figure 33) and also the outgroups of B-class gene phylogenies, see figures 36 and 41.

accession ID	**Species taxon**	**Origin**	**Pollinator**	**Subgenus According to** [112]	**Clade / genus According to** [36]
4627	*Lavandula angustifolia* Mill.	Europe	Bee	-	*Lavandula**
5156	*Dracocephalum ruyschiana* L.	Eurasia	Bee	-	*Dracocephalum**
5751	*Ocimum tenuiflorum* L.	Asia	Bee	-	*Ocimum**
5391	*Mentha spicata* cv. *Crispa* L.	Europe	Insects	-	*Mentha**
6574	*Salvia hierosolymitana* Boiss.	Europe	Bee	*Sclarea*	IV-C *Salvia* s.s
7089	*Dracocephalum rupestre* Hance	Asia	Bee	-	*Dracocephalum**
7579	*Mentha spicata* L.	Europe	Insects	-	*Mentha**
8020	*Dracocephalum ruyschiana* L.	Eurasia	Bee	-	*Dracocephalum**
8167	*Dracocephalum rupestre* Hance	Asia	Bee	-	*Dracocephalum**
8174	*Dracocephalum ruyschiana* L.	Eurasia	Bee	-	*Dracocephalum**
8349	*Dracocephalum rupestre* Hance	Asia	Bee	-	*Dracocephalum**
8680	*Mentha aquatica* L.	Eurasia	Insects	-	*Mentha**
8681	*Mentha arvensis* L.	Eurasia	Insects	-	*Mentha**
8753	*Salvia hispanica* L.	America	Bee	*Calosphace*	II-A *Lasemia*
8792	*Salvia officinalis* ssp. *lavandulifolia* (Vahl) Gams	Europe	Bee	*Salvia*	I-D *Salvia* s.s
8831	*Salvia lavandulifolia* Vahl	Europe	Bee	*Salvia*	I-D *Salvia* s.s
8973	*Salvia tiliifolia* Vahl	America	Bee	*Calosphace*	II-A *Lasemia*
8981	*Salvia coccinea* Buc'hoz	America	Bird	*Calosphace*	II-A *Lasemia*
8985	*Salvia dominica* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
8987	*Salvia officinalis* L.	Europe	Bee	*Salvia*	I-D *Salvia* s.s
8994	*Salvia nemorosa* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9097	*Salvia argentea* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9098	*Salvia officinalis* cv. *albiflorum* L.	Europe	Bee	*Salvia*	I-D *Salvia* s.s
9150	*Salvia sclarea* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9265	*Salvia canariensis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9268	*Salvia brousonetii* Benth.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9269	*Salvia somalensis* Vatke	Europe	Bee	*Heterosphace*	I-A *Salvia* s.s
9301	*Salvia viridis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9302	*Salvia viridis* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s
9304	*Salvia fruticosa* Mill.	Europe	Bee	*Salvia*	I-D *Salvia* s.s
9305	*Salvia fruticosa* Mill.	Europe	Bee	*Salvia*	I-D *Salvia* s.s
9341	*Salvia nutans* L.	Europe	Bee	*Sclarea*	I-C *Salvia* s.s

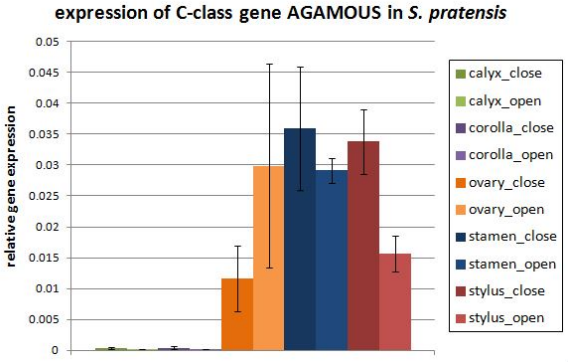
Asterisks indicate exceptions in genus ascription. Those nomenclatures are not from Will and Bockhoff (2017) [36], but from the plant list [37].

8.2 Floral gene expression in *Salvia pratensis* and *Salvia elegans* by gene

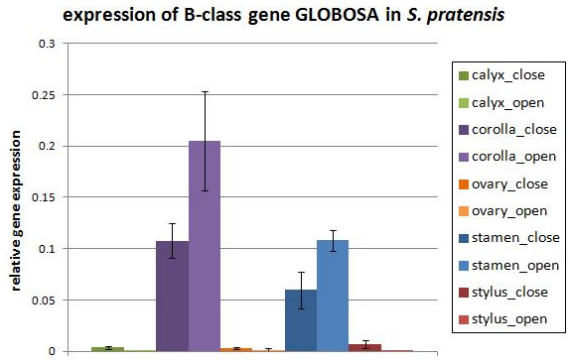
8.2.1 *Salvia pratensis*



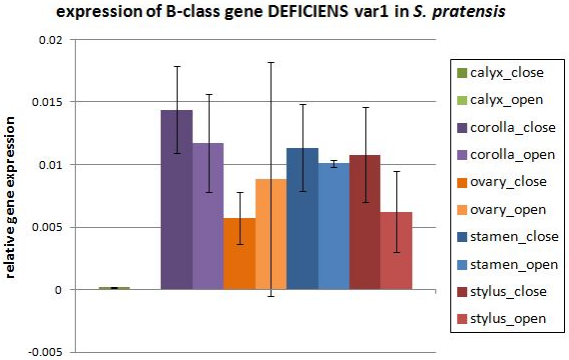
(a) APETALA1 - *Salvia pratensis*



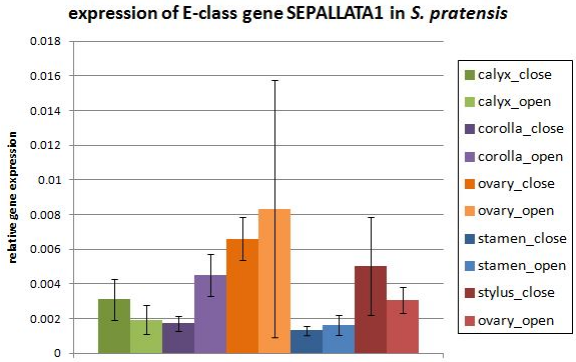
(b) AGAMOUS - *Salvia pratensis*



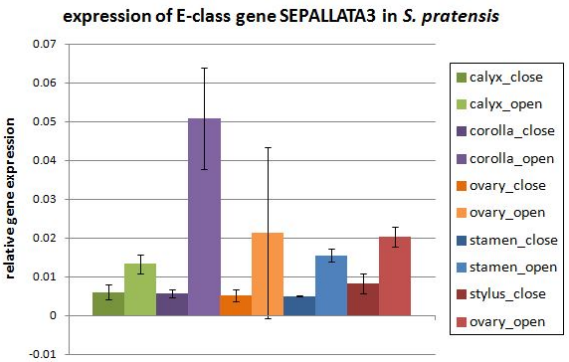
(c) GLOBOSA - *Salvia pratensis*



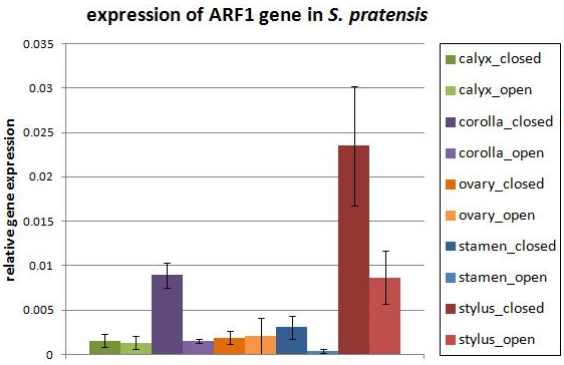
(d) DEF variant1 - *Salvia pratensis*



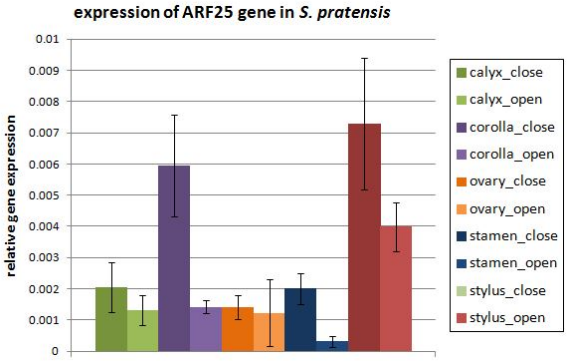
(e) SEPALLATA1 - *Salvia pratensis*



(f) SEPALLATA3 - *Salvia pratensis*



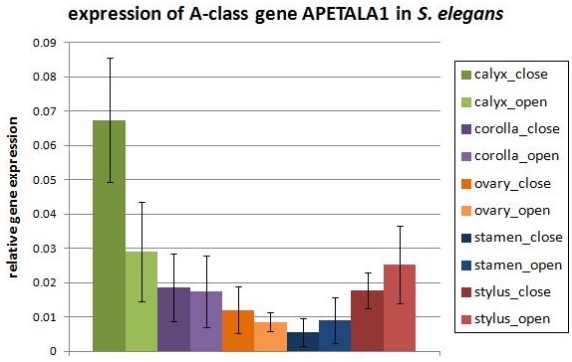
(g) ARF1 - *Salvia pratensis*



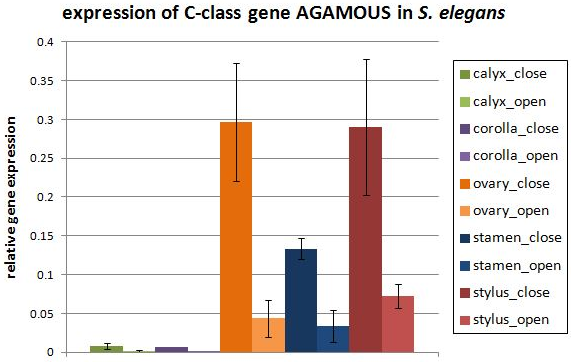
(h) ARF25 - *Salvia pratensis*

Figure 43: Illustration of expression results of *Salvia pratensis* by gene.

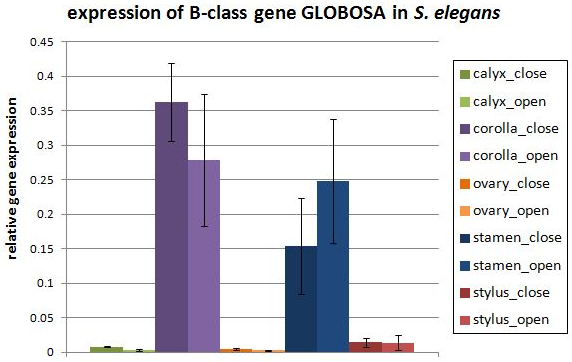
8.2.2 *S. elegans*



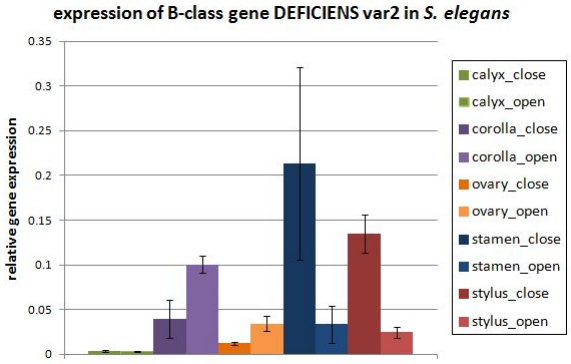
(a) APETALA1 - *Salvia elegans*



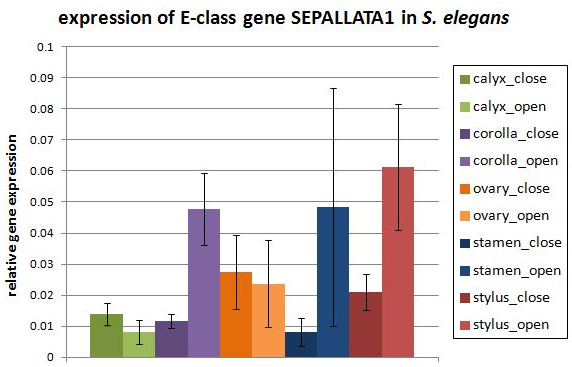
(b) AGAMOUS - *Salvia elegans*



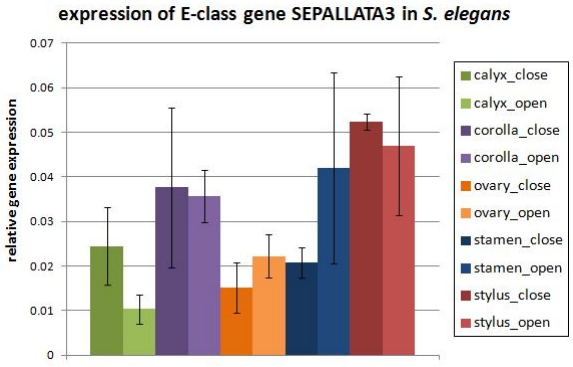
(c) GLOBOSA - *Salvia elegans*



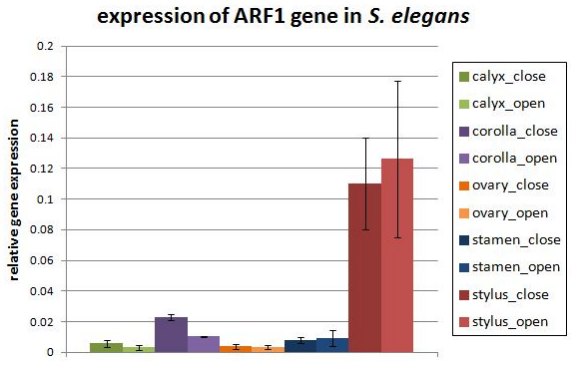
(d) DEFICIENS variant2 - *Salvia elegans*



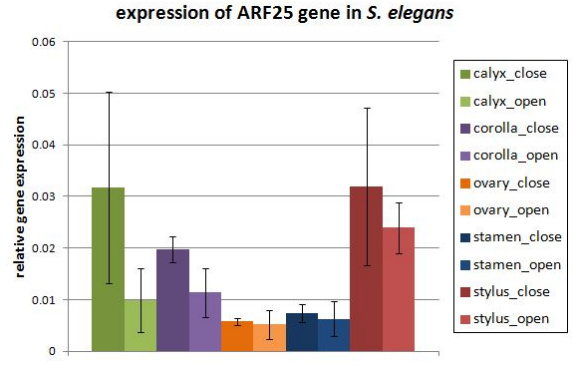
(e) SEPALLATA1 - *Salvia elegans*



(f) SEPALLATA3 - *Salvia elegans*



(g) ARF1 - *Salvia elegans*



(h) ARF25 - *Salvia elegans*

Figure 43: Illustration of expression results of *Salvia elegans* by gene