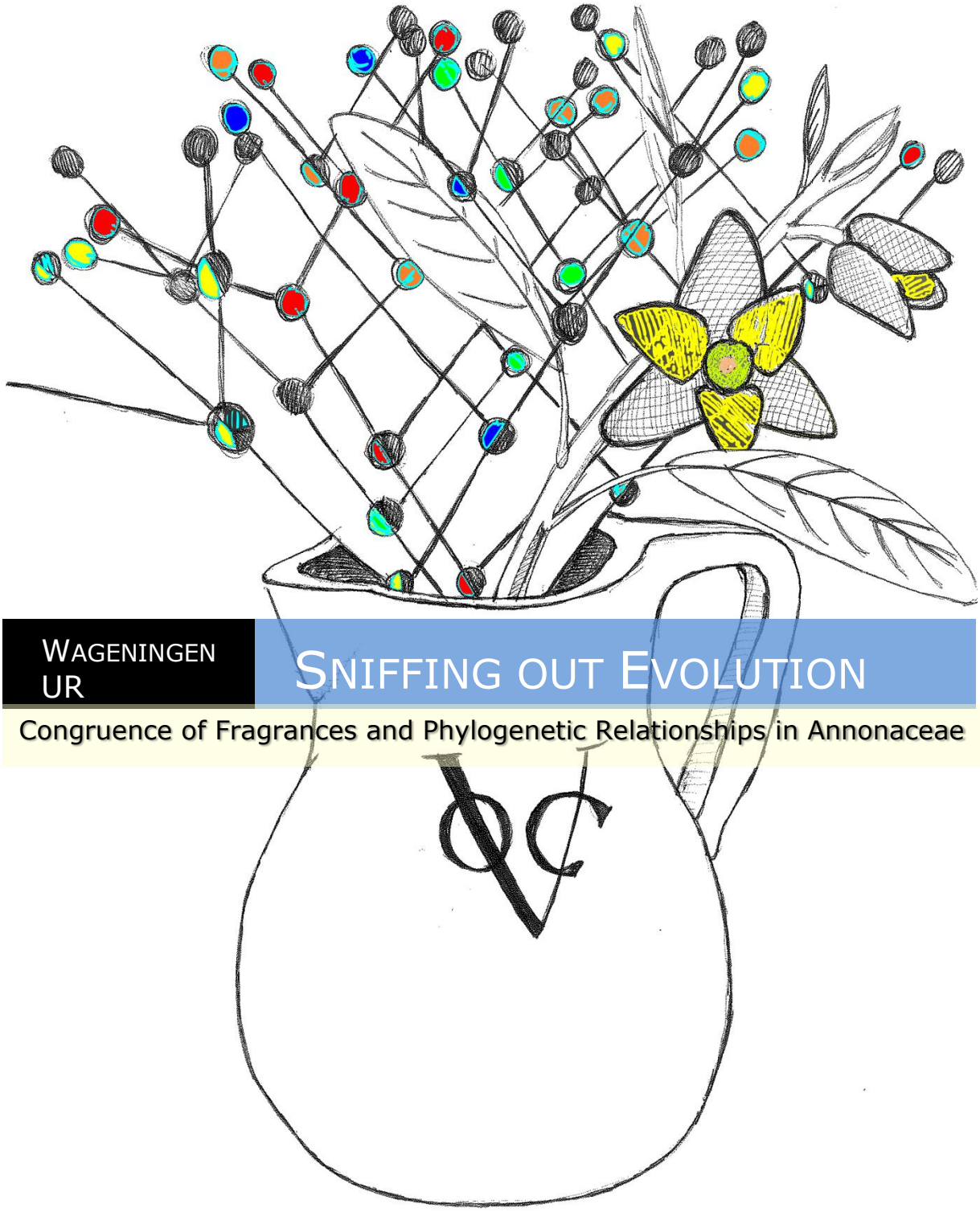


WAGENINGEN UR

SNIFFING OUT EVOLUTION

Congruence of Fragrances and Phylogenetic Relationships in Annonaceae

Jeike L. van de Poel BSc. – MBI 890421659010

Under supervision of Dr. Lars Chatrou and Dr. Kate Goodrich

Biosystematics group, October 2012
Course code: BIS-80439

Abstract

The flowering plants are a relatively young group that has reached high levels of diversity and harbors an enormous assortment of chemical components. Annonaceae encompass an astounding amount of different flower fragrances. Floral odor is important in the attraction of pollinators, especially if the pollinators in question are beetles, which is the case for most Annonaceae. A floral fragrance consists of approximately 100 Volatile Organic Compounds. From two genera within the Annonaceae (*Asimina* Adans. and *Deeringothamnus* Small) the complete chemical composition of the floral scent has been studied. These chemical data were provided in order to perform optimizations over a simplified version of an existing phylogenetic tree based on chloroplast markers (ITS, accD-psaI, matK-trnK, psbA-trnH, psbM-ycf6, rpL16 intron, rpl32-trnl, rpoB-trnC, trnC-ycf6, trnL-trnL-trnF, trnS-psbC, trnS-trnfM, ycf1). Due to a lack of previous studies in which phylogenetic analysis and extensive collection of fragrance data are combined, there is no real agreement as to what is the best method to use fragrance data in phylogenetic analyses. The aim of this study is to find out which method (within Maximum Parsimony and Maximum Likelihood optimization) is best to use for optimization of components from floral fragrances, and whether it is possible to use chemical data as a means to resolve polytomies. Two major biosynthetic pathways of floral volatiles are reconstructed with the help of the KEGG-database and the Plant Metabolic Network. These pathways enable the creation of stepmatrices, which assign costs to the transition of one volatile organic compound into another. Chemical components are treated as character states (polymorphic characters) or all components from one species' fragrance are grouped together into a different character state, which is called a 'scenario'. Maximum Parsimony optimization with an asymmetric stepmatrix as a cost-model was expected to lead to the best results. However, a symmetric stepmatrix performs better during tests. Scenarios are interesting, but optimization-results are hard to interpret, due to the large amount of components that are present in one scenario. Therefore it is suggested to do a Maximum Likelihood optimization (with a symmetric Mk1-model) as a follow up. Optimization of chemical data seems apt to resolve polytomies. However, a lot of improvements have to be implemented in future studies, before these (somewhat preliminary) conclusions can be fully supported.

Keywords

Annonaceae, *Asimina*, chemical pathways, cladogram, *Deeringothamnus*, floral fragrance, optimization, volatile organic compounds.

Table of Contents

List of Tables and Figures

Tables

Appendices

Figures

Appendices

Preface

By now, I have spent so much time at the Biosystematics Group in Wageningen that I have begun to regard most of the staff members as (distant) family. I can honestly say I will miss their personalities, their wit, and not in the least their intelligence. My fellow students, or the other *Island-inhabitants*, will always retain a special place in my heart. I therefore wish to thank all the staff and students from the Biosystematics Group in Wageningen, in particular of course my supervisor, Dr. Lars Chatrou.

My second supervisor, Dr. Kate Goodrich from Widener University (Chester), is someone whom I would like to triple-thank. First because she agreed to be my supervisor, and thus helped me to finish this thesis. Secondly because of her hospitality when I came to the US to learn about Gas Chromatography – Mass Spectrometry (GC-MS). And last but not least because she agreed to me using her data for this study. Another person that was willing to help me out, this time by providing me with a tree, was Dr. Kurt Neubig. So, "Thank You!", Dr. Kurt Neubig.

The original plan for this thesis differed quite a bit from the eventual execution, due to some problems with the Costa Rican government. This was a real disappointment, but some Costa Ricans did not disappoint at all. Special thanks to Nelson Zamora, Reinaldo Aguilar and other people from the InBio-institute who helped us find the Annonaceae during our fieldtrip. To get to Costa Rica and the United States of America, I got financial help from the Christine Buisman Fonds, Fonds Taxon Ondersteuning, Stichting het Kronendak and the Alberta Mennega Stichting, whom I would like to thank cordially for their generosity.

During the preparation of this fieldtrip our team got a lot of help from the Plant Physiology Group in Wageningen, in the person of Francel Verstappen. He gave us a lot of advice regarding the material needed to extract floral fragrances, and would have helped me with the GC-MS, had I ever gotten my data. I would like to thank him very much, as well as Dr. Eric Visser who helped us find a good glass blower. The third person I would like to thank for his help is Dr. Klaus Schliep, the writer of the Phangorn Package in R. He readily helped me with my scripts when I would falter.

Introduction

The high-impact journal Science celebrated its 125th anniversary in 2005. At the occasion, 25 imperative questions facing the life sciences were enumerated (Science vol 309, issue 5731, pp. 1-204). One of these question is 'what determines species diversity?' (Pennisi 2005). The seemingly straightforward question is difficult to address. Baseline data are poor, and the interplay of factors that shape species diversity requires multidisciplinary projects. Yet, it is crucial to reveal the mechanisms shaping diversity to understand the nature of the wave of extinctions the world is experiencing, and to determine strategies to mitigate it.

The flowering plants are a relatively young group that has reached high levels of diversity. Pollen of flowering plants appeared in the Early Cretaceous (ca 140 Ma; Feild and Arens 2005), but only some 30 Ma later they became significant elements of ecosystems. Several hypotheses have been coined to explain the evolutionary success of flowering plants. In general, these hypotheses are non-exclusive, each of them highlighting biological features that have uniquely evolved in flowering plants. Recently, Jiao et al. (2011) corroborated the hypothesis that whole genome duplication has enlarged the genetic repertory, which could have been a major factor in the evolution of life history innovations causing the rapid diversification and dominance of flowering plants. Other hypotheses relate to the origin of deciduousness (Hickey et al. 1983; Hill and Scriven 1995), or the origin of fruit facilitating the dispersal of seeds by animals (e.g. Smith 2001).

This project focuses on the importance of the evolution of flowers for the evolution of flowering plants (Labandeira et al. 1994). Rather than the mere possession of flowers, it is the variation in floral characters that promote diversity. Support for this hypothesis, coming from large-scale studies that associate the relationships between floral traits and diversification rates on a broad scale (Dodd et al. 1999; Davies et al. 2004) is moderate when statistical rigour is applied (Kay and Sargent 2009). The lack of robust support is likely to be related to the large scale of these studies, inevitably involving sampling issues (Ricklefs and Renner 2000), and the inclusive and complex nature of the characters analysed.

On a small scale, there are few studies suggesting a relationship between diversification and diversity in floral traits. Armbruster et al. (1994) demonstrated that divergence in floral morphology occurred as a means to avoid competitive exclusion in a group of sympatric, congeneric species. In an assemblage of eight species of tropical palms of the genus *Geonoma*, floral scents were clearly differentiated (Knudsen 1999). However, the relationship between floral odour diversity and pollination remained unclear, as data on pollinating insects were almost absent. In a study on two subspecies of *Antirrhinum majus*, Suchet et al. (2011) found clear olfactory differences between the subspecies, which affected pollinator attraction. The three studies, however, did not address whether character differences occurred after speciation, i.e. serves a role in species maintenance, or whether character evolution was a prerequisite for speciation. The only way in which this question can be addressed is by studying floral phenotype from a phylogenetic perspective.

In the Annonaceae a number of studies in different relevant fields has been completed to warrant the study of floral scents and their role in the origin and maintenance of species diversity. Although they are very diverse taxonomically, the basic setup of their flowers is remarkable alike (van Heusden 1992; Silberbauer-Gottsberger et al. 2003): Annonaceae flowers typically are trimerous, having a single whorl of sepals and two whorls of petals, which are morphological distinct (Saunders 2010). However, flower morphology is diverse in some clades, notably in clades that are species-rich, such as *Duguetia* (Maas et al. 2003) and *Annona* (Rainer 2007). For several of these groups species-level phylogenetic studies are on-going (e.g. Chatrou et al. 2009). Moreover, various detailed observational studies on pollination ecology have been done, demonstrating diversity in the pollinating insects within clades (e.g. Gottsberger 1989; 1999; Maas et al. 2003). Finally, analyses of volatile organic compounds that characterise the floral fragrance show high levels of variation between congeneric species (Jürgens, Webber et al. 2000; Goodrich and Raguso 2009).

Floral Fragrance

This master's project was performed under the general assumption that the arrival of flowers in evolutionary time has increased the chance of survival in plants significantly. Flowering plants harbour an enormous diversity in chemical components that are thought to protect them against attacks from all sorts of enemies in changing environments (Agrawal and Fishbein, 2006; Gershenzon and Dudareva, 2007). Flower's fragrances possibly were derived from these (Pellmyr and Thien, 1986), and, according

to van der Pijl (1960), floral odours preceded floral colours as a pollinator attractant. Floral odour is important for the attraction of pollinators (Faegri and van der Pijl 1966; Dobson 1994; Miyake and Yafuso 2003; Salzmann et al. 2007), especially if the pollinators in question are beetles (Proctor, Yeo et al. 1996; Gottsberger 1999; Armbruster and Muchhala 2009; Saunders 2010; Teichert, Dötterl et al. 2011), which is the case for most Annonaceae (Gottsberger 1989; 1999; Silberbauer-Gottsberger et al. 2003). Waelti et al. (2008) proposed that floral scent plays an important role in the reproductive isolation between two closely related species. Even if this would not be the case, floral odour would be particularly interesting as it is highly complex and diverse, both within and among taxa (Raguso et al. 2001; Dobson 2006; Chess et al. 2008). The chemical composition of floral fragrances has been studied quite extensively (Kaiser 1993; Dudareva and Pichersky 2000; Dobson 2006; Knudsen et al. 2006; Pichersky et al. 2006). However, the evolutionary factors responsible for this composition (and the variation within) are not well-known (Schiestl 2010).

Annonaceae

Within the Annonaceae, an astounding amount of different flower fragrances is present (Goodrich, 2012). This is possibly due to the large amount of Annonaceae that are pollinated by beetles (e.g. Gottsberger, 1999). Beetles often feed on the flowers they pollinate (Gottsberger, 1989), one of the suggested reasons why Annonaceae flowers have thick petals (Gottsberger, 1999). Since floral fragrance was possibly derived from defensive compounds used to deter herbivores (Pellmyr and Thien, 1986), this could have led to the vast array of fragrances in this plant family. A correlation between floral scent and pollinator attraction is probable, as was shown in many studies on pollination biology in Annonaceae (Gottsberger 1988; Silberbauer-Gottsberger et al. 2003; Teichert 2008). Only two genera of Annonaceae have scent composition analyses for all species (*Asimina* and *Deeringothamnus*; (Goodrich and Raguso 2009)). Besides these two genera, only fourteen other species have had their scent analyzed using Gas Chromatography-Mass Spectrometry (GC-MS) (Ma et al., 1998, Jürgens et al. 2000; Ratnayake et al. 2007; Teichert 2008; Teichert et al. 2009; Pripdeevech 2011; Goodrich, 2012).

Algorithms

If the volatile organic compounds (VOCs) that make up the floral fragrance are quantified, these data can be used in further studies. To gain more insight in the evolution of this floral trait one could build a cladogram with these data, or optimize the chemical data over existing DNA-based phylogenies. Due to a lack of previous studies in which phylogenetic analysis and extensive collection of fragrance data are combined, there is no real agreement as to what is the prefered method to use fragrance data in phylogenetic analyses (Levin et al., 2003).

Maximum Parsimony (MP) makes use of the criterion that the tree with the minimum number of changes is always the best (Farris, 1983). This algorithms has been used in evolutionary research often, and in many different ways. In the chemosystematic studies performed by Bolick (1983); Humphries and Richardson (1980); Richardson (1983); Nandi et al. (1998), the data was coded as either present or absent and used to build a tree using maximum parsimony. Every chemical component was treated as an independent character, and every transition between character states was thought to be of equal weight. Any numerical analysis requires the characters to be independent (Sneath and Sokal, 1973). However, many chemical components are not statistically independent, as they are biogenetically related (part of the same pathway) and coding characters that are likely to be produced via identical pathways could violate the assumption of independence (Barkman, 2001). Also, plant volatiles are usually ill-suited to phylogenetic reconstruction (Azuma et al., 1997; Barkman et al., 1997; Williams and Whitten, 1999) since they show quite some homoplasy (Azuma et al., 1999) or fail to render phylogenetic signal (Jürgens et al., 2006). Therefore Givnish and Sytsma (1997), as well as Levin et al. (2003,) suggest to optimize floral fragrance data over already existing phylogenies.

While optimizing volatile organic compounds over phylogenies, the characters can be treated as being independent or they can be ordered (for instance: character C can only originate from character B, which can only originate from character A). Some argue (e.g. Barkman, 2001) that it would be even better to take the pathway[1] into account in the analyses. This could be achieved by simply using a stepmatrix. When one works with a stepmatrix, the pathway would become the character, and the VOCs the

[1] VOCs are chemical components that are synthesized by enzymes in chemical pathways.

character's states. Every grid value in a stepmatrix provides the cost (the total number of steps[2]) to get from one character state to the other.

The Maximum Likelihood (ML) algorithm calculates the probability of expecting a certain character at the ancestral (internal) nodes of a tree. The likelihood that a tree is the best tree is gathered from these probabilities (Felsenstein, 1981).

- The symmetrical (Mk1, Lewis (2001)) model for categorical data works with an equal probability as to changing state i into j and changing j into i. Characters are treated as independent unities.
- Another ML-approach that considers the characters to be independent of each other is the asymmetrical (2 parameter Markov k-state) model. This is a maximum likelihood-model for categorical data that enables the user to assign different costs to the gain and loss of a character (Maddison and Maddison, 2011).

A third technique that could be used for optimization of chemical data over existing trees is Stochastic Character Mapping. Unfortunately, a Master's Thesis just does not provide enough time to look into all techniques, and it was therefore decided to not consider this algorithm in this particular study.

Questions and Hypotheses

The aim of this study is to find out which of the methods, described above, would be preferable to use for optimization of components from floral fragrances. The preferability of a model will depend on the probability that the model's output is realistic. Realistic in this context would mean that there is a certain congruence between the pathway that is optimized and the tree that it is optimized over. In other words, chemical components, reconstructed to be ancestral (occurring in one or more of the four most ancestral nodes in the trees), should not be mandatorily derived from a component that is not calculated to be ancestral, but does occur in the fragrance of one or more species in the tree.

The most preferable model is to be found by testing the following six hypotheses:

- Maximum Parsimony optimization, using the unordered approach yields realistic results.
- Maximum Parsimony optimization, using the ordered approach yields realistic results.
- Maximum Parsimony optimization, using a symmetric stepmatrix yields realistic results.
- Maximum Parsimony optimization, using an asymmetric stepmatrix yields realistic results.
- Maximum Likelihood optimization, using a symmetrical model yields realistic results.
- Maximum Likelihood optimization, using an asymmetrical model yields realistic results.

Based on a study by Barkman (2001), Maximum Parsimony optimization with an asymmetric stepmatrix as a cost-model is expected to lead to the best results. However, Barkman (2001) used an asymmetric stepmatrix to build a new cladogram, while in this study, the stepmatrix will be used as a cost-model during optimization over a pre-existing dendrogram.

Another objective of this project is to find out more about the applicability of floral fragrances in phylogenetics. As the genetic code directly governs the expression of enzymes, which are needed to produce the different chemical components, it is hypothesized that optimizing volatile organic compounds over trees can help resolve polytomies. In order to test this hypothesis, the pscores (parsimony scores) of all optimizations will be compared.

[2] 'Cost' is the same as 'steps' in this case, as a stepmatrix is read as a cost matrix: the larger the number of steps, the higher the cost of the transition.

Materials and Methods

Scent sampling

The link between floral fragrance and pollinators has been a topic of interest for many years. In 1919 it was shown experimentally by Frisch that bees have a specific sense of smell, and in the thirties it was suggested that the cotton plant used the 'odor of a chemical compound' to attract the boll weevil (Bosart, 1936). Clearly, scientists were interested in the fragrance of flowers and the role they played in our world. However, their lack of descriptive adjectives led to complications. In the early 20th century only a few descriptions were in use (such as 'floral', or 'oriental'), and these were interpreted in various ways (Bosart, 1936). However, in 1934, the gaseous plant hormone ethylene was first described, and during the last decennia, the understanding of the biosynthetic routes (and their molecular background) that lead to the production of plant volatiles has increased enormously (Pichersky and Gershenzon, 2002). This progress is due to advances in three scientific areas: biochemistry, molecular techniques and the collection and analysis of volatile substances (Millar and Sims, 1998). Two main collection techniques are described below.

Dynamic Headspace Sampling

During dynamic headspace sampling (also known as active air sampling), the sample matrix is actively led over a filter. This leads to a relatively high sample rate in comparison to passive air sampling, which is only useful when measurements persist for at least a week (Begerow et al. 1995; Harner et al., 2006). However, to lead the matrix over the filter, an extra device (pump) is needed. Also, there must be an inflow into the sample matrix, to preclude a vacuum. Figure 1 depicts the experimental set-up that was used during the fieldwork that preceded this study. The VOCs that stuck to the filter are trapped in an extraction matrix, to which an internal standard (a highly volatile, known, component) can be added. Since the amount of this internal standard is known, it enables the calculation of the quantities of the other volatiles in the extraction matrix.

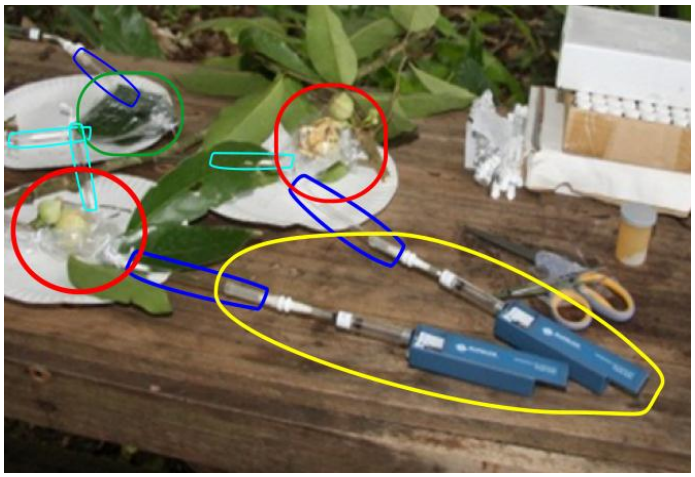


Figure 1 Scent sampling in the field, using dynamic headspace sampling
The yellow circle contains two PAS-500 personal air sampler, the dark-blue rectangles are drawn around the filters (glass tubes filled with 150 mg Tenax®) in which the floral VOCs are caught, the lighter blue variant represents the filters (idem) that clear the incoming air. The red circles comprise the Annonaceae flowers in their Turkey-bags, while the green circle harbours a leaf (to catch some wounding-volatiles).

Static Headspace Sampling (SPME)

Although Solid Phase Microextraction (SPME) is not the only static headspace (also known as passive air) sampling-technique, it is the one that will be described here, as the data that were used for this study were obtained using SPME. As the name suggests, in SPME, the amount of solvent extracted is very small in comparison to the sample volume. The sample matrix will not become depleted, rather an equilibrium between the sample matrix and the extraction phase will be reached (Pawliszyn et al., 1997). Because volatiles are attached to a fibre and not trapped in a matrix, no internal standard can be added. If there is no internal standard no quantities can be measured, and thus one can only work with relative percentages (Goodrich, personal communication, 2012).

GC-MS

Gas Chromatography was first introduced 60 years ago (Snow & Slack 2002). Nowadays, the GC has an extra –MS attached to it: Gas Chromatography-Mass Spectrometry (GC-MS). In the GC-MS, the eluates are heated until the entire content of the extraction matrix is evaporated (dynamic headspace sampling) or all components have let go of the SPME-fibre, upon which the volatiles are led to an oven which holds a glass column. The initial temperature of the oven is typically 40ºC, a temperature at which the volatiles will bind to the glass wire. Slowly, temperature is increased, until eventually a temperature of c. 250ºC is obtained. The slow increase in temperature causes the docked compounds to start 'walking' the wire, triggered by their boiling point. Thus, compounds that are highly volatile (have a low boiling point) become active before others do, and move up the line more quickly. The single compounds enter the

mass-spectrometer one by one. The period of time it takes them to enter this part of the machine gives information as to their boiling point. Once the compounds are out of the oven, they get ionised, a process that splits them into smaller units. The spectrum of these subunits' masses is registered by the mass-spectrometer and is known as the compound's 'fingerprint'. Every chemical compound has its own unique fingerprint. The identification of each individual compound in a fragrance blend is performed by the comparison of the compound's fingerprint to those in existing libraries (e.g. NIST and Wiley). If the fingerprint matches that of a compound in one (or more) of the libraries for over 90 percent, the compounds are considered to be one and the same. The best method would be to identify a chemical component by comparing its mass spectrum to that of a known standard run on the same machine. However, given limitations of time and access to standards, the 90 percent library match is often chosen as an acceptable method (Goodrich, personal communication 2012). The (relative) quantity of each compound can be derived from the area under the compounds peak in the graph, produced by the GC-MS.

Data

Tree

A phylogenetic tree based on chloroplast markers (ITS, *accD-psal*, *matK-trnK*, *psbA-trnH*, *psbM-ycf6*, *rpL16* intron, *rpl32-trnl*, *rpoB-trnC*, *trnC-ycf6*, *trnL-trnF*, *trnS-psbC*, *trnS-trnfM*, *ycf1*) was provided by Dr. Kurt Neubig (Neubig and Abbott, unpublished data). Figure 2 shows a concensus tree without branch lengths. This tree comprises all nine *Asimina* species, plus the two *Deeringothamnus* species, and several outgroups. These outgroups (*Annona, Cymbopetalum, Disepalum, Oxandra,* and *Rollinia*) were selected based on previous work (Chatrou et al., 2012; Doyle et al., 2004; Richardson et al., 2004).

Chemical components

Floral fragrance data were provided by Dr. Kate Goodrich (Goodrich and Raguso, 2009). These chemical data comprise 269 volatile organic compounds (VOCs), of which 114 have been identified. Although this is less than half of the total amount of the volatiles in the dataset, the identified chemical components form the majority of the relative scent composition, and each species' fragrance is composed of these identified VOCs.

Data Editing

Tree

The Perfect Match

With a view to optimizing the floral fragrance data, as provided by Dr. Kate Goodrich (Goodrich and Raguso, 2009) the tree as provided by Dr. Kurt Neubig (Neubig and Abott, unpublished data; see Figure 2) was simplified. It was altered in such a way that only the species for which Dr. Kate Goodrich's chemical dataset provided information were present. Also, it was made sure that no species occurred in a tree more than once. As a result, sampling was reduced to a single accession per species, except for two species: *Asimina pygmaea* and *Asimina obovata*. The first species had chloroplast markers that occurred in two different clades, and it was decided to place *Asimina pygmaea* within in the clade with *Deeringothamnus*, *Asimina*

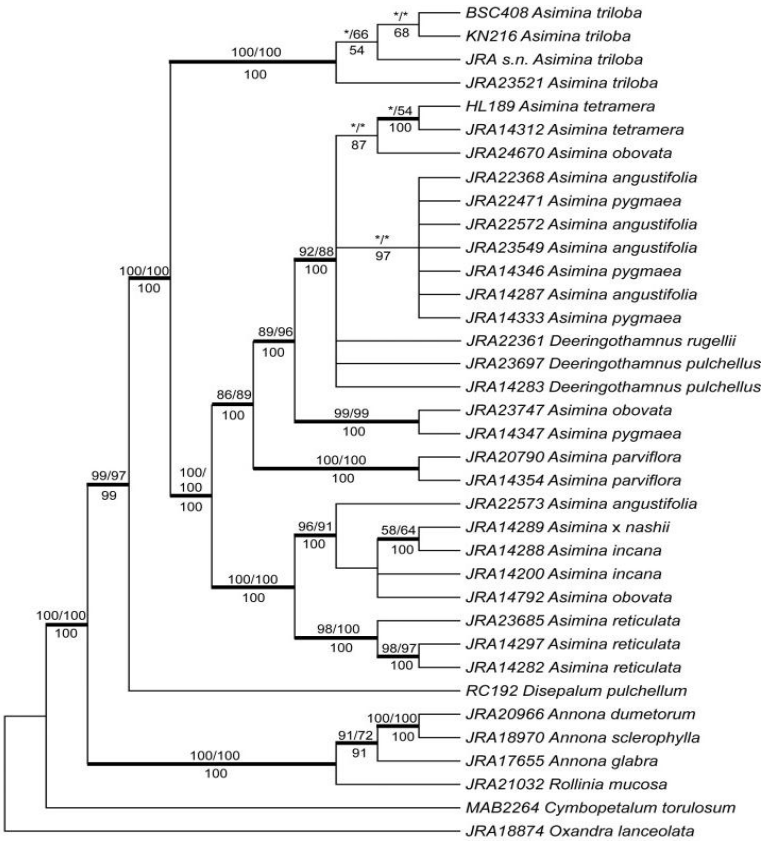


Figure 2:. Tree generated using all 11 plastid regions. Combined bootstrap consensus tree under maximum parsimony (MP), maximum likelihood (ML), and Bayesian analyses. Values above branches indicate MP and ML percentages respectively, values below branches indicate Bayesian posterior probabilities. Asterisks indicate values of less than 50% and thickened branches indicate consistent clades among the strict consensus of the 128 trees found in the MP search. – modified from Neubig and Abbott (unpublished data).

tetramera, and *Asimina longifolia*, in which most of this species' chloroplast markers occurred. The temperate *Asimina obovata* (common name: Bigflower Pawpaw) proved to be a bit harder a nut to crack.

Where To Put Bigflower Pawpaw?

There were no grounds to prefer one placement of *Asimina obovata* over the other and therefore all three occurrences were seen as the, possibly, best event (Figure 3). Three trees with polytomies were created (Figure 4), one for every possible occurrence of *Asimina obovata*. The tree in which *Asimina obovata* was placed in the position as taken by 'Obov 1' in Figure 3 was called "Obov1", etcetera.

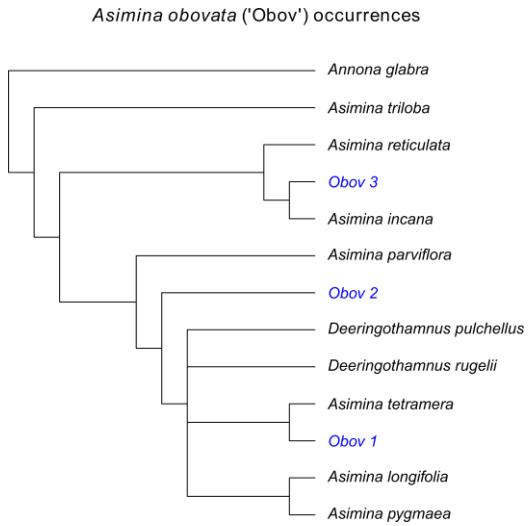


Figure 3 Simplified version of the tree provided by dr. Kurt Neubig (Neubig and Abbott, in press). The three occurrences of *Asimina obovata* are represented by 'Obov1', 'Obov2', and 'Obov3'.

Resolutions

To avoid problems that possibly occur while optimizing (using MP) over polytomies, the polytomies in the tree trees were resolved manually. The unresolved trees were now called Obov1+, Obov2+ and Obov3+. The plus sign indicates that the tree is unresolved, the one, two and three represent the first, second or third placement of *Asimina obovata*. Subsequently, each of these three trees was resolved, which led to nine resolved trees per Obov+-tree. Thus, 27 new trees were created, which were given the names 1a – 1i, 2a – 1i, 3a – 3i, in which the number is an indication of the unresolved tree it is derived from. In total, 30 trees were created, all without branch lengths.

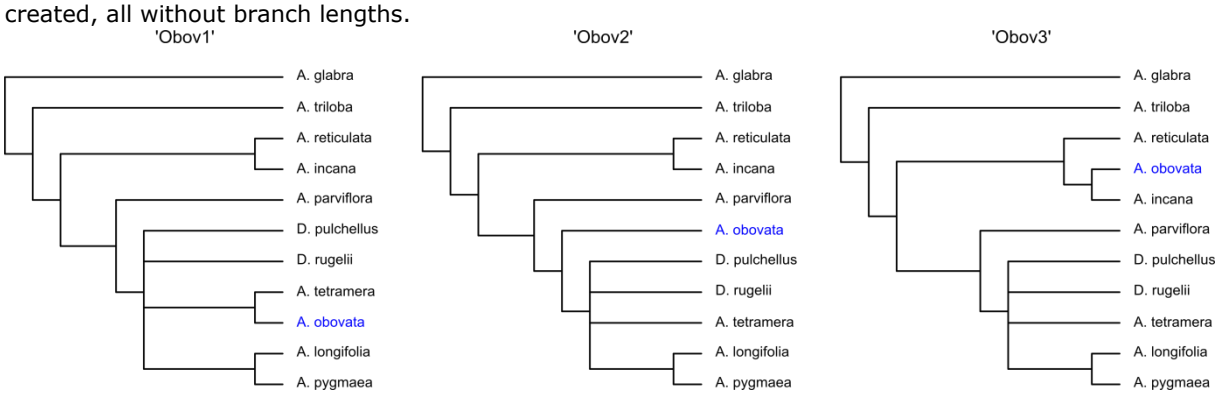


Figure 4 Trees with different placements of *Asimina obovata* (see also Figure 3). Polytomies are preserved.

Chemical components

Scoring

The dataset as provided by Dr. Kate Goodrich (Goodrich and Raguso, 2009) consisted of relative percentages (average and standard deviation) for each component in the female and male flowers of every species. SPME collections never yield absolute quantities, as there is no internal standard (Goodrich, personal communication 2012). The Biosystematics group at Wageningen University and Research Centre has access to active air-sampling equipment, which could provide absolute quantities. Since the goal of this study is to clarify and to provide a better understanding of optimizations of chemical data over existing phylogenetic trees, and to keep things as simple as possible, it was decided to work with qualitative data only. Whenever a component was present in either the female stage or male stage of a flower (or in both), it was scored as present in the species' floral fragrance.

Links

With the help of the free chemical structure database 'ChemSpider' (www.chemspider.com) the names assigned to the 114 VOCs could be complemented with synonyms. Isomers were treated as one component only (for instance four Lilac alcohol isomers were treated as the composite component: "Lilac alcohol isomer") as isomers are bound to originate from the same precursor and in a similar way (Barkman, 2001). The databases provided by the Kyoto Encyclopedia for Genomes and Genes (www.kegg.jp) and the Plant Metabolic Network (www.plantcyc.org) provided information on several

pathways, reactions and components. As the pathways provided by these databases were usually small and/or incomplete, they were linked when components overlapped (e.g. a pathway in which acetone is formed was linked to a pathway in which acetone is degraded). This was done under the assumption that different plant species and/or families use very similar, if not the same, pathways (Croteau et al., 2000; KEGG.jp).

Pathway-Creation

Two major pathways were reconstructed (Figure 5), which in turn were simplified until they showed only the most important compounds, plus the amount of chemical transitions it took to change one component into another. Figure 6 and Figure 7 show the pathways for the Terpenes and components derived from Chorismate, respectively.
Only the 34 VOCs that could be incorporated in one of these two pathways were used in the analyses.

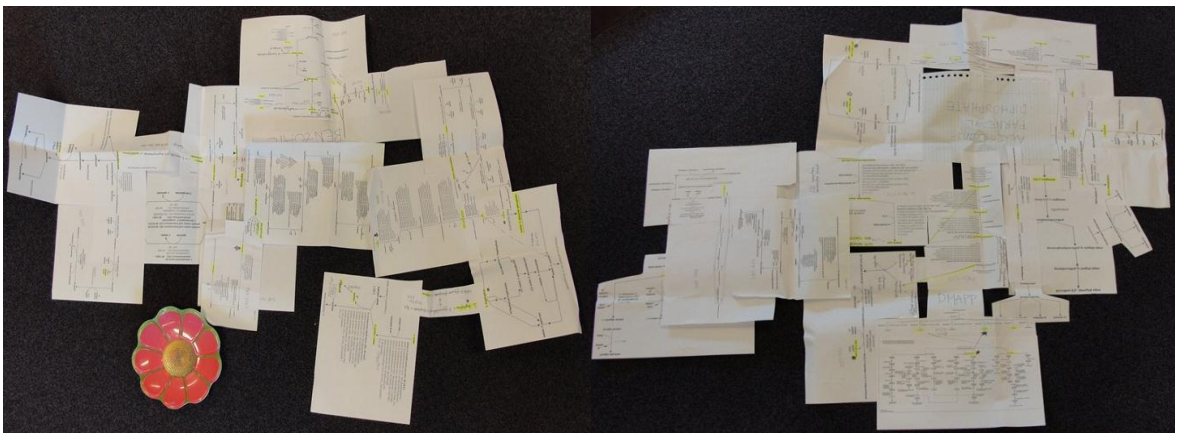


Figure 5 Being creative with pathways and reactions from plantcyc.org and KEGG.jp. On the left the Chorismate-pathway. The Terpene-pathways is on the right.

Scenarios

Every species' fragrance composition consists of a different subset of the 34 identified VOCs described earlier. Let us use a metaphor to explain fragrance composition and the perception of this fragrance.
If a chemical component, such as ethanol, would be a flower, the fragrance would resemble a bouquet. Although most people would not opt for a bouquet with only flowering grasses, they are often used to complement the other, usually more colorful, flowers. Although we could experience the smell of a single component to be repugnant, if it is part of a set of chemicals, we might actually like the fragrance. The same might hold for insects and thus, in the light of pollination as a driver of evolution, it is important to also take a look at the set of chemicals that make up the flower's fragrance.
The identified VOCs from one pathway, as present in one species' floral fragrance, was called a 'scenario'. Scenarios were described for the Monoterpenes, Sesquiterpenes and Terpenes (Figure 6) and for the compounds derived from Chorismate (Figure 7). Scenarios were treated as character states. In this case, the character would be 'Fragrance'.

Simulation

Considering the vast amount of chemical data, the magnitude of the chemical pathways and the number of (resolved) trees that are used in this study, it was decided to first perform a simulation. The data, tree and pathway used in this simulation are a fictional representation of the real data, tree and pathway, but with the amount of data and species seriously reduced. All methodological hypotheses will be tested by optimizing the simulation-data (sData) over a simulation-tree (sTree), while using a simulation-pathway (sPathway) to construct stepmatrices. The objective of this simulation is to gain more insight in the different methods to optimize chemical data over a tree and the simulation's outcomes will be used to decrease the amount of tests that have to be performed with the real data[3], the actual tree[4] and the Terpene- and Chorismate- pathways.

[3] The data, as derived from the dataset that was provided by Dr. Kate Goodrich (Goodrich and Raguso, 2009), that is used in the rest of this study.
[4] The tree, as derived from the one provided by Dr. Kurt Neubig (Neubig and Abbott, unpublished data) that is used in the rest of this study.

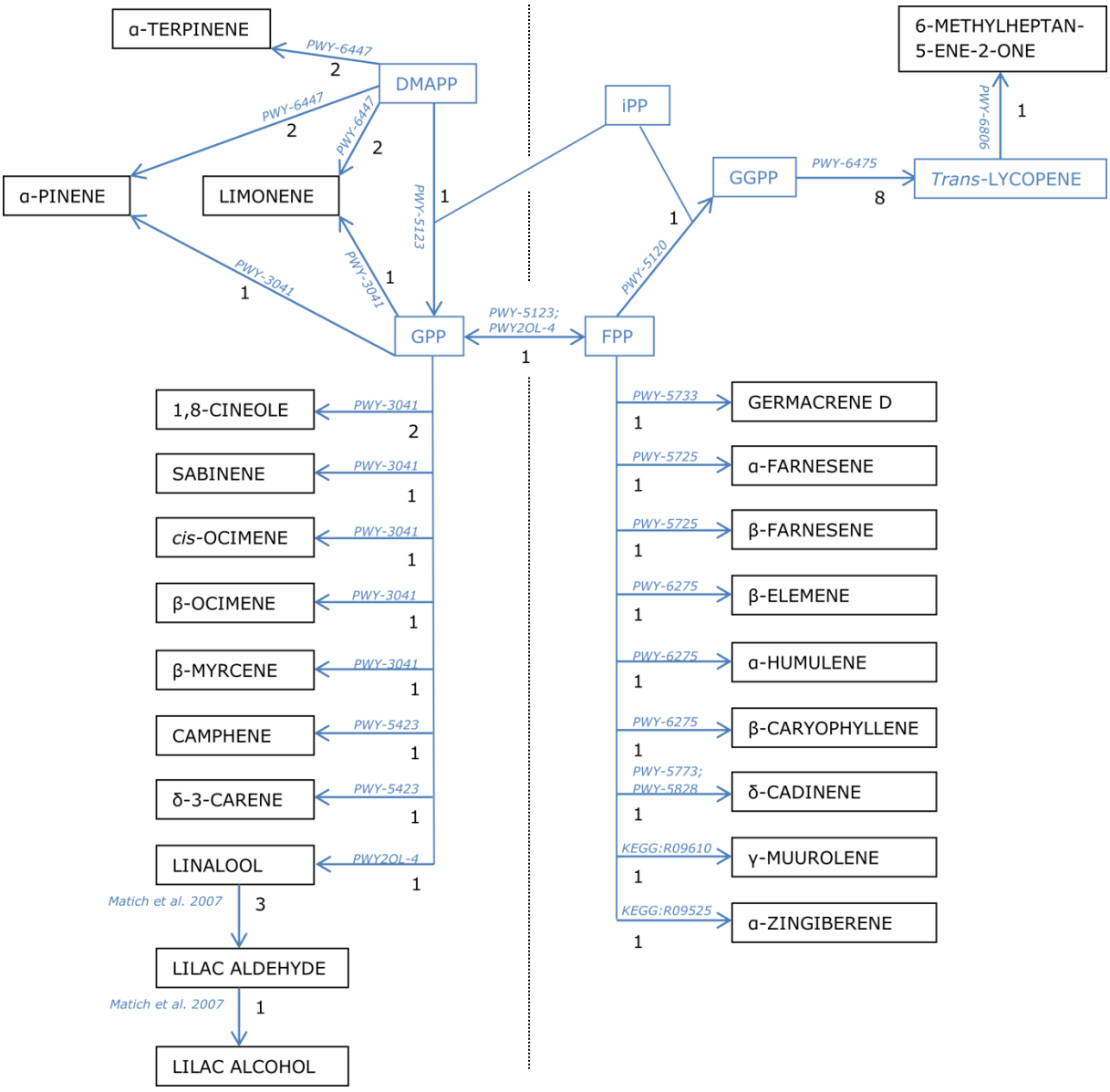


Figure 6: Pathway showing formation of Terpenes (Monoterpenes on the left and Sesquiterpenes on the right side of the dashed line). Volatile Organic Compounds as identified in the floral fragrances (Goodrich and Raguso, 2009) are shown in black. Numbers represent the amount of steps (chemical transitions with the help of a different chemical component, such as an enzyme) needed to change one compound into the other.
Steps in this figure are based on pathways which are given in blue. Most are based on plantcyc.org, when they are based on kegg.jp, the pathway's name is preceded by 'KEGG'.

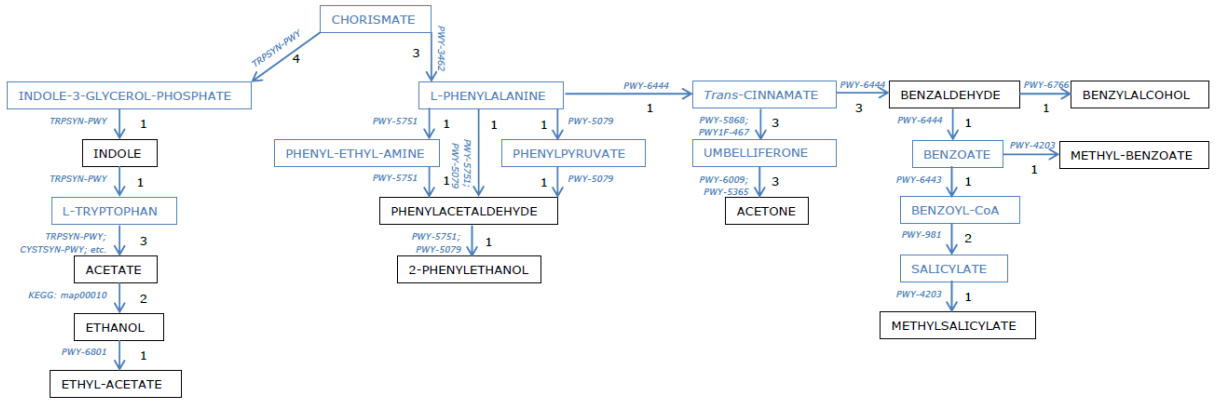


Figure 7: Pathway showing formation of Chorismate-derived compounds. Volatile Organic Compounds as identified in the floral fragrances (Goodrich and Raguso, 2009) are shown in black. Numbers represent the amount of steps (chemical transitions with the help of a different chemical component, such as an enzyme) needed to change one compound into the other.
Steps in this figure are based on pathways which are given in blue. Most are based on plantcyc.org, when they are based on kegg.jp, the pathway's name is preceded by 'KEGG'.

Fiction

sTree

The sTree consisted of 7 taxa: Species 1 till Species 6 and an Outgroup. Because the actual tree did not have branch lengths, the sTree was not given branch lengths either. In most phylogenetic studies the appropriate assumption would be that polytomies are 'soft', which means that the relationships between the species that form a polytomy are unclear. However, it is extremely difficult to perform calculations using soft polytomies. Mesquite (Maddison and Maddison, 2010), for instance, can only handle dichotomous trees, or those with 'hard' polytomies for calculations like the ones performed in this study (Maddison and Maddison, 2007). When polytomies are stated to be hard, the program will interpret the polytomies to be sure facts, in other words, it will assume that all species in a polytomy descended from the same ancestor, at the exact same time. Often this is an unrealistic assumption, like is the case with the actual tree (and the sTree) in this study. Unfortunately though, this debatable assumption is the best option, since Mesquite will not be able to perform the right optimizations over trees that contain soft polytomies. To make sure that the sTree resembles the actual tree as much as possible, a (soft) polytomy was introduced in the sTree, which was then stated to be hard.

sPathway

The sPathway was kept as simple as possible, without completely diminishing resemblance to the Terpene- and Chorismate-pathway. Out turns into A (1 step), A turns into B (2 steps) and B turns into C (1 step), see also Figure 8.

Figure 8 Graphical representation of fictional pathway, used in this Test. Each arrow represents one transition (step).

sData

The left half of Table 1 holds the simulational characters (sCharacters) that were assigned to the different species. These VOCs were given the names 'Out', 'A', 'B', and 'C', and are all products (and intermediates) of the same sPathway.

Table 1 Species and characters. The left part of the table ("Binary, Original") provides the chemical components as they occur in the simulation-floral fragrances. The right side ("Binary Extra") gives the chemical components plus their precursors in the pathway.

	Binary, Original				Binary, Extra			
	Out	*A*	*B*	*C*	*Out*	*A*	*B*	*C*
Species 1	0	1	0	1	1	1	1	1
Species 2	0	1	0	0	1	1	0	0
Species 3	0	0	0	1	1	1	1	1
Species 4	0	0	1	0	1	1	1	0
Species 5	0	1	0	0	1	1	0	0
Species 6	0	1	0	0	1	1	0	0
Outgroup	1	0	0	0	1	0	0	0

Data Editing

Precursor Character States

The right part of Table 1 shows the sCharacters as observed per species, plus their precursors in the sPathway. As character A can only originate from character Out, it would be logical to conclude that Out must have been present in a species that produces A. The same accounts for B (A and Out must also have been present) and C (Out, A and B are all needed to get to C). Extras were not implemented in scenarios, as this would be (almost) impossible to do for real data.

Table 2 Polymorphic characters and Scenarios, without definition of Outgroup's character state. "Polymorph, Original", "Polymorph, Extra" and "Scenario" are characters, and their states are 'A', 'B', 'C', '1', '2', '3', '4' and '?'.

	Polymorph, Original	Polymorph, Extra	Scenarios
Species 1	A, C	A, B, C	4
Species 2	A	A	1
Species 3	C	A, B, C	3
Species 4	B	A, B	2
Species 5	A	A	1
Species 6	A	A	1
Outgroup	?	?	?

Polymorphisms

Polymorphic sCharacters were derived from the binary sCharacters. These polymorphic sCharacters had the binary sCharacters as their states (e.g. instead of having character A and B, the character had states A and B). This was done in order to perform optimizations with a stepmatrix, which could assign costs to transitions from one state to another. In the polymorphic sCharacters, as in the binary sCharacters, the precursors were also taken into account (Polymorph, Extra), as can be seen in Tables 2 and 3.

Table 3 Polymorphic characters and Scenarios, with definition of Outgroup's character state. "Polymorph, Original", "Polymorph, Extra" and "Scenario" are characters, and their states are 'Out', 'A', 'B', 'C', '0', '1', '2', '3', and '4'.

	Polymorph, Original	Polymorph, Extra	Scenarios
Species 1	A, C	Out, A, B, C	4
Species 2	A	Out, A	1
Species 3	C	Out, A, B, C	3
Species 4	B	Out, A, B	2
Species 5	A	Out, A	1
Species 6	A	Out, A	1
Outgroup	Out	Out	0

Scenario

Figure 9 Fictional pathway. Crosses represent places in which a loss could occur that would lead to de disappearance of C.

As polymorphisms can be hard to handle during calculations (Mesquite will treat either one of the states as present at a node, instead of all, Maddison and Maddison, 2010), scenarios were created (see Table 2 and Table 3). Again, "Scenario" was the character, while '0', '1', '2', '3', '4' and '?' were the character states. State '0' was assigned to a species that had only "Out" as a sCharacter. State '1' meant that a species had "A" as the only sCharacter (precursors were not taken into account for Scenario), state '2' was given to a species that only excreted character "B", state '3' represents "C" as the only VOC in the floral fragrance (in this pathway). State '4' reflected the presence of "A" and "C" at the same time.

Outgroup

There is one big difference between Table 2 and Table 3: The Outgroup was assigned no character (?) or its own character state 'Out'. This was done to gain better insight in the effect of not defining a species' character or character state. The reason that Outgroup was chosen for this test is that in the sCharacters with 'Extra' in their names 'Out' was present in all species, and thus would not provide any information regarding evolutionary kinship.

The sCharacter states were distributed over the sTree in such a way that the 'true' sTree should yield an ancestor with character state A, and or B, but no C present.

Stepmatrices

By disregarding the fact that chemical components are synthesized through a certain pathway, one could lose a lot of information. One way to overcome this is by taking the pathway into account in the analyses, for instance by using a stepmatrix. Every grid value in a stepmatrix provides the cost (the total number of steps) to get from one character state to the other, or from one scenario to another. 'Cost' is the same as 'steps' in this case, as a stepmatrix is read as a cost matrix: the larger the number of steps, the higher the cost of the transition.

Six stepmatrices were created, three for polymorphic characters:

1) Symmetric stepmatrix (Table 4)
2) Asymmetric stepmatrix (Table5)
3) Inverted stepmatrix

and three for scenarios:

1) Asymmetric stepmatrix (Table 6)
2) Symmetric stepmatrix (Table 7)
3) Inverted stepmatrix (Table 8).

PolyAsym

The loss of a VOC can have many causes. It might be due to absence of a complete gene (Yuba et al., 1996), because of failure in transcription (Wang et al., 1997), or it could even have to do with the inability to secrete scent compounds (Barkman 2001). Of course, it could also be because of a mutation that affected a precursor. For instance, if C cannot originate from anything but B, and B is lost, C will be lost as well. Figure 9 illustrates this: the blue crosses represent parts in the pathway in which a loss would lead to the loss of C. Four crosses are shown, which means that C could be lost in at least four ways. On the other hand, C can originate via one way only: Out has to turn into A, which needs to turn into B, so that B can change into C. Thus, for the transition from Out to C, four steps need to be taken, and this transition would be given a 'cost' of four. However, C can turn into Out in several ways, of which the loss of the ability to change Out into A would be one. In other words, at least one step is needed for this transition. A popular belief is that costs should always be kept as low as possible and a follow up would be to assign a cost of 1 to the transition of 'C' into 'Out'.

PolySym

Although the transition from C to Out could need only one step, it could also be that all four steps are taken (first the ability to change B into C, then the ability to change A into B and finally the potency to turn Out into A). This conviction would be the foundation of a symmetric stepmatrix, in which gains and losses are treated equally. For example: A to C would cost 3 steps, as would C to A.

PolyInv

For this stepmatrix, the asymmetric stepmatrix was inverted, which means that the transition A -> B was assigned the costs that were originally assigned to the transition B -> A, etcetera. Inverting the

asymmetric stepmatrix (Table 5) would not have any biological meaning because it assigns higher costs to a loss than to a gain of a character state. However, it could increase our understanding in the mathematics behind the MP-optimizations considerably.

ScenAsym

A scenario can consist of several chemical components. This makes the formation of a stepmatrix a lot harder. The stepmatrix was produced following these guidelines:

- Determine amount (and cost) of losses. A loss only has a cost of 1, as it could occur at the start of a pathway. If one loss could lead to another the cost is accounted for only once (e.g. A -> B -> C; B & C lost, cost is 1).
- Determine gains and their costs. If a component gained is a precursor for another gained component, it is assumed that the precursor was formed first, as reflected in the costs (e.g. A -> B -> C; B and C gained, cost = (cost A -> B) + (cost B -> C))
- The amount of steps is always kept as low as possible.

ScenSym

Almost the same guidelines that were applied during the formation of the asymmetric scenario-stepmatrix are applicable to the symmetric scenario-stepmatrix:

- Determine amount (and cost) of losses. A transition through a loss always has the same cost as a transition through a gain (cost(A -> C) = cost(C -> A)).
- Determine gains and their costs. If a component gained is a precursor for another gained component, it is assumed that the precursor was formed first, as reflected in the costs (e.g. A -> B -> C; B and C gained, cost = (cost A -> B) + (cost B -> C))
- The amount of steps is always kept as low as possible.

Table 4 Symmetric stepmatrix with costs to change from state i into j and from j into i. A loss costs the same as a gain.

	A	B	C	Out
A	0	2	3	1
B	2	0	1	3
C	3	1	0	4
Out	1	3	4	0

Table 5 Asymmetric stepmatrix with costs to change from state i into j and from j into i. A loss always has a cost of 1.

	A	B	C	Out
A	0	2	3	1
B	1	0	1	1
C	1	1	0	1
Out	1	3	4	0

ScenInv

The inverted stepmatrix for Scenario is not really a direct inversion of the asymmetric stepmatrix for scenarios. Although the individual transitions' costs are inverted (from A to B becomes from B to A), the scenario-states usually comprise multiple characters, and thus the total costs for a transition from one scenario state to another should be calculated from scratch.

Table 6 Asymmetric stepmatrix with costs to change from state i into j and from j into i. A loss always has a cost of 1.

	A	B	C	A&C	Out
A	0	2	3	3	1
B	1	0	1	1	1
C	1	1	0	0	1
A&C	1	1	0	0	1
Out	1	3	4	4	0

Table 7 Symmetric stepmatrix with costs to change from state i into j and from j into i. A loss costs the same as a gain.

	A	B	C	A&C	Out
A	0	2	3	3	1
B	2	0	1	1	3
C	3	1	0	0	4
A&C	3	1	0	0	4
Out	1	3	4	4	0

Table 8 Asymmetric stepmatrix with costs to change from state i into j and from j into i. A gain always has a cost of 1.

	A	B	C	A&C	Out
A	0	1	1	1	1
B	2	0	1	1	3
C	3	1	0	1	4
A&C	3	1	1	0	4
Out	1	1	1	1	0

Maximum Parsimony (MP)

Results

Figure 10, Figure 11 and Figure 12 show optimizations with and without specification of Outgroup's character 'Out' under four approaches, for the polymorphic, original sCharacter ("Polymorph, Original"), the polymorphic sCharacter with extra states ("Polymorph, Extra"), and the scenarios ("Scenario"), respectively. Results will be discussed per sCharacter.

Polymorphic, Original

Figure 10 shows eight trees, four with and four without a character state Out assigned to Outgroup. If no state was assigned to Outgroup (character state = ?), this taxon is assigned the character state that is present in the most recent common ancestor. Overall, the trees per approach are very much alike, only when using the symmetric matrix some minor differences can be distinguished.

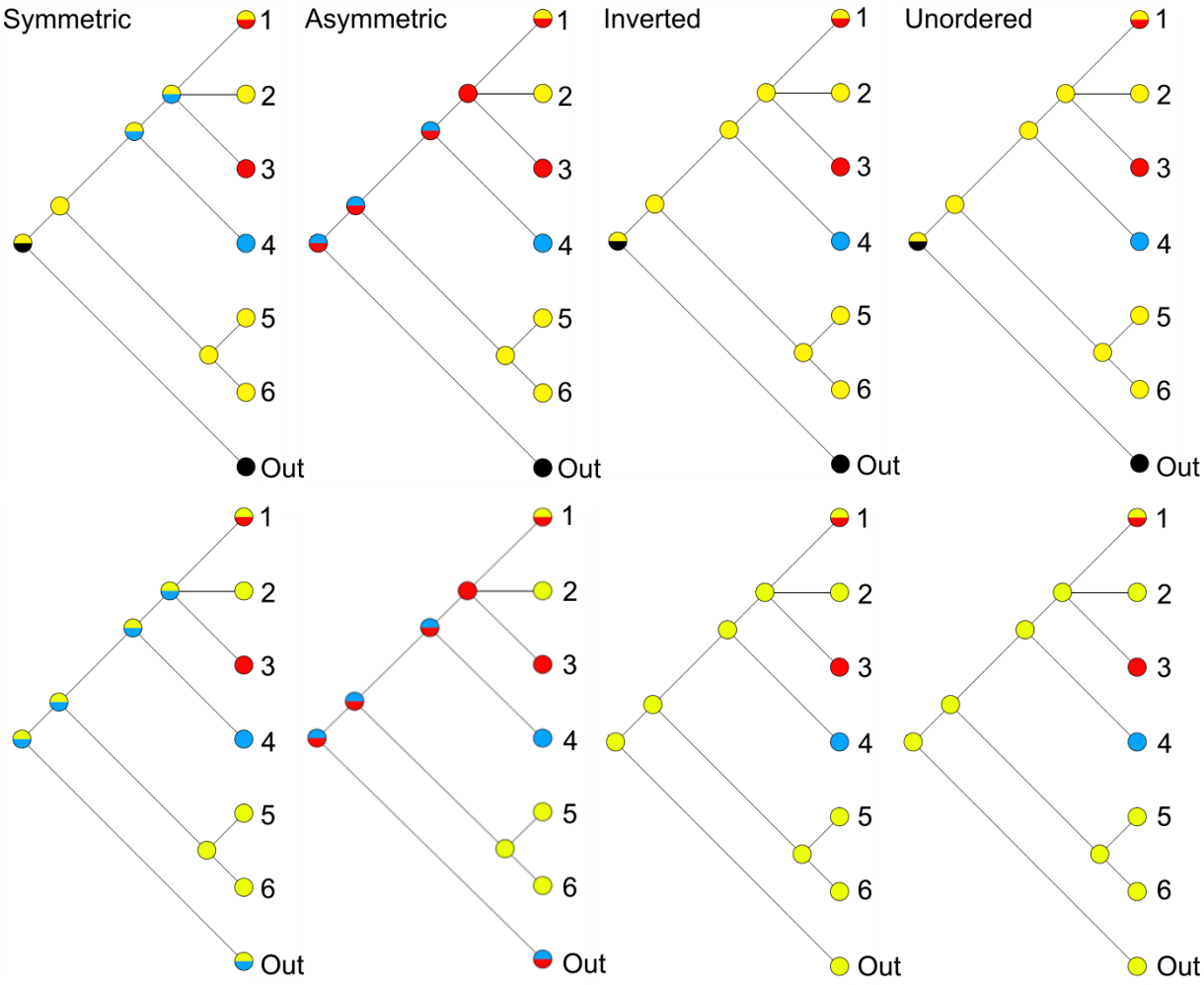


Figure 10 Maximum Parsimony optimization of a polymorphic character ("Polymorph"), with (upper half) and without (lower half) definition of Outgroup. States are: A (yellow), B (blue), C (red) and Out (black). Four approaches are shown: 'Unordered', a current MP model in Mesquite; 'Asymmetric', with use of the asymmetric stepmatrix (Table 5); 'Symmetric', using the symmetric stepmatrix (Table 4); and 'Inverted', with the asymmetric matrix inverted.

The exact similarity between the unordered and inverted approach is striking. All changes occur on the terminal branches only, which would mean that all ancestors had the same character state. This is not a very realistic outcome. Also the inverted stepmatrix on itself has no biological meaning whatsoever.

The asymmetric approach yields the most derived character state as the (possible) ancestral state, a result that makes no sense in view of the biosynthetic pathway. However, it does make sense taking the criterion of parsimony into consideration:

A -> C = 3 steps C -> A = 1 step

Two possibilities:

Gaining C two times cost = 2 x 3 = 6

Losing C four times cost = 4 x 1 = 4

The second option involves fewer steps, and thus this is the option preferred by parsimony optimization. Apparently, the higher cost of the manufacture of a chemical compound pushes this compound back in the tree. The symmetric approach makes less sense biologically, as a loss costs as much as a gain. However, it still is more justifiable than the unordered approach (and definitely more meaningful than

the inversed approach), as it takes into account the chemical pathway through which the compounds originate. This optimization procedure yields some interesting results too, as character state Out, A or B are reconstructed to be present in the trees' most ancestral nodes.

Polymorphic, Extra

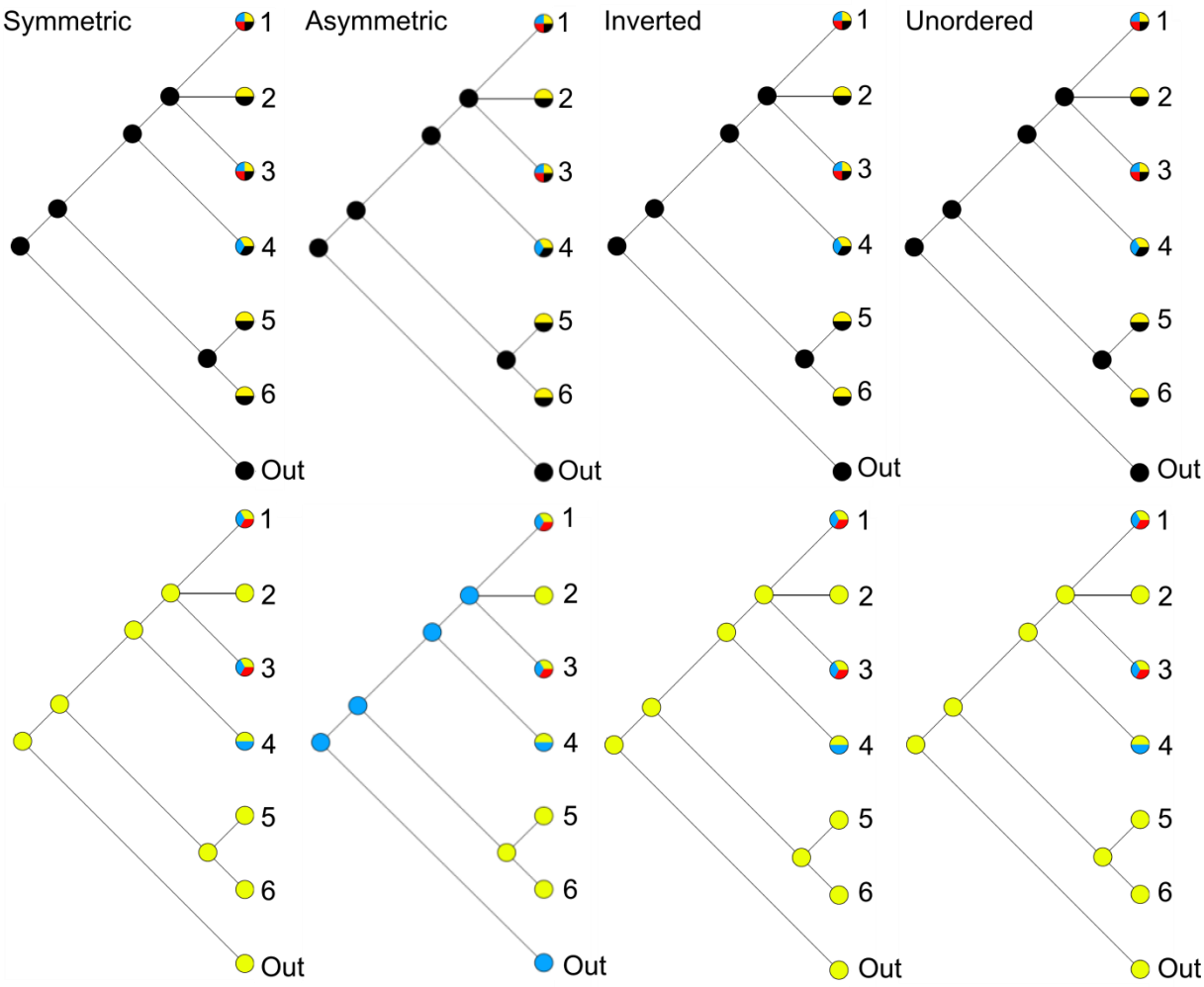


Figure 11 Maximum Parsimony optimization of a polymorphic character ("Polymorph, Extra"), with (upper half) and without (lower half) definition of Outgroup. States are: A (yellow), B (blue), C (red) and Out (black). Four approaches are shown: 'Unordered', a current MP model in Mesquite; 'Asymmetric', with use of the asymmetric stepmatrix (Table 5); 'Symmetric', using the symmetric stepmatrix (Table 4); and 'Inverted', with the asymmetric matrix inverted.

Figure 11 shows eight trees, just as Figure 10, but with extra states added (see Table 1). As character state A was set to be derived from Out, the former character state is present in all taxa in the four trees in which this character state is taken into account. The preponderance of taxa by state Out (in the top four trees) and state A (in the bottom four trees) leads to a preponderance of these states in all nodes, except for the tree (optimized via the asymmetric approach) with no character state specified in Outgroup: This tree yields character state B as the ancestral state in all nodes but one. This phenomenon is a result of the higher costs that were assigned to gaining a more derived character state, compared to the costs of a loss.

Scenarios

Again, eight trees can be found in Figure 12, but this time the character is not polymorphic. Thus, the terminal character states can be interpreted in one way only. This should facilitate things for Mesquite (Maddison and Maddison, 2010). Nonetheless, the exact same pattern that occurred in Figure 10 and 11 presents itself in these trees. Inverted and Unordered yield the same result, the asymmetric approach indicates that the scenarios in which the more or most derived character states are present must have been ancestral. The symmetric approach points to the less-derived scenarios (only character state A and/or B) as ancestral scenarios, which is in agreement with the set-up of this experiment.

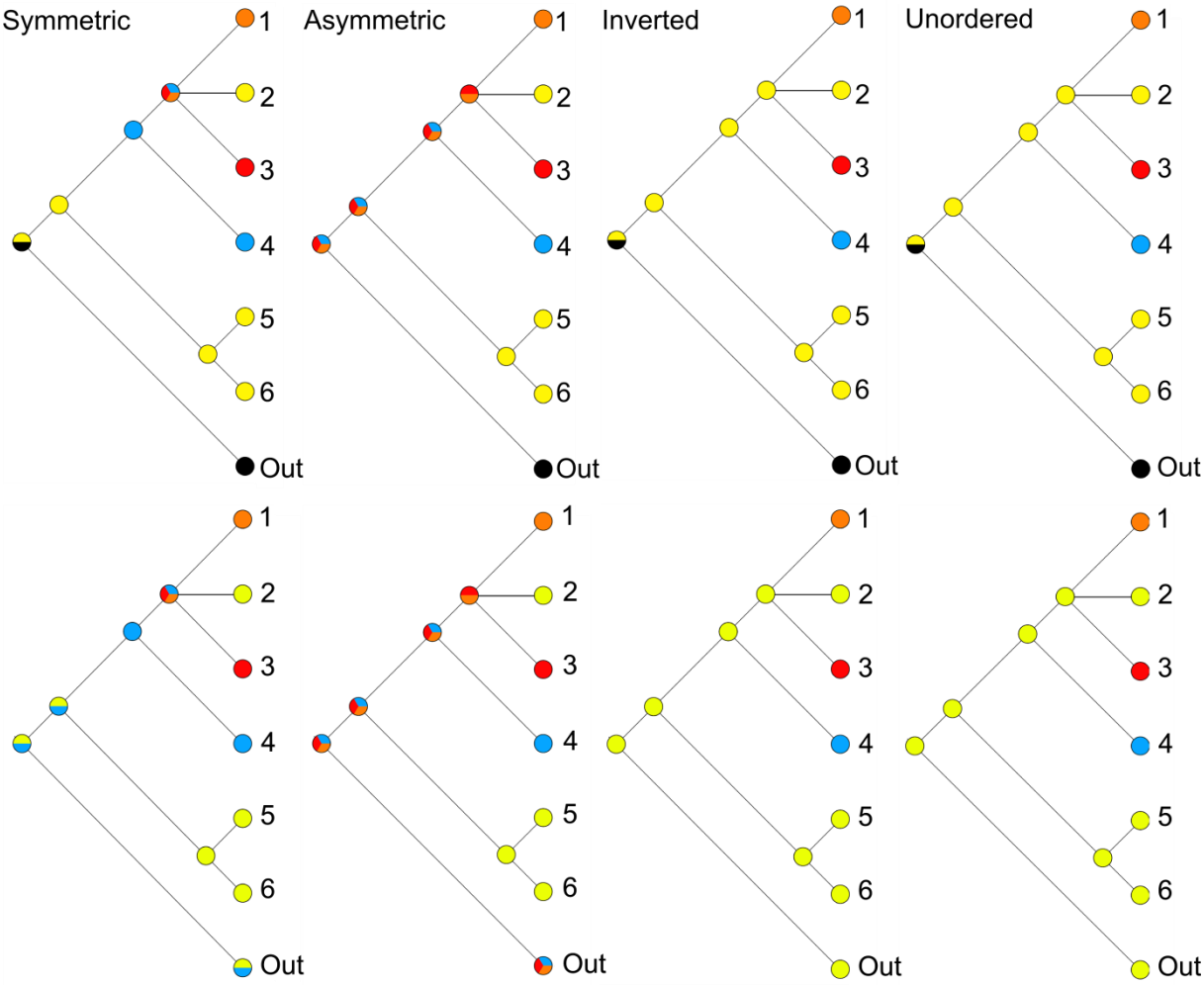


Figure 12 Maximum Parsimony optimization of a character ("Scenario"), with (upper half) and without (lower half) definition of Outgroup. States are: A (yellow), B (blue), C (red), A&C (yellow) and Out (black). Four approaches are shown: 'Unordered', a current MP model in Mesquite; 'Asymmetric', with use of the asymmetric stepmatrix (Table 6); 'Symmetric', using the symmetric stepmatrix (Table 7); and 'Inverted', with the asymmetric matrix inverted.

Conclusions (MP)

In all cases, the unordered approach and the inverted approach yielded the same results. As the inverted approach is wrong, biologically speaking, the unordered approach seems not very apt to use in a study of this kind. In both approaches all changes in character states only happen at the terminal branches of the tree. The asymmetric approach resulted in the more derived character states to be ancestral, which although correct in view of the optimization procedure, does not really make sense from a chemo-biological point of view. For example: if C can only exist if it is preceded by B and A, it just cannot be that an ancestor only had character (state) C as its ancestral state. The approach that seems to be the most apt for optimizations using maximum parsimony is the approach in which a symmetric stepmatrix is used. This approach takes into account the chemical pathway via which the chemical components originate and yields results that are correct, considering the design of this dataset.

Maximum Likelihood (ML)

As stated before the asymmetrical (2 parameter Markov k-state) model for categorical data is a maximum likelihood-model that enables the user to assign different rates to the gain and loss of a character. This can be done by either changing the forward and backward rates, or by changing the bias of one rate over the other. A bias of over 1 means that forward changes are more likely to occur, a bias smaller than one means the opposite (Maddison and Maddison, 2011). If both parameters are unspecified and need to be estimated, one can let Mesquite (Maddison and Maddison, 2010) optimize the parameter values. Three possibilities are given:

- "Use estimated rate from Mk1 model as basis"
- "Try even and two asymmetries"

- "Try both of the above (Mk1 plus even and two asymmetries)",

of which the last one (try both) is the standard setting.

Discussion of Results

This being a test, it was thought best to try with equal rates (backward = forward), or the symmetrical (Mk1, Lewis (2001)) model first. Probabilities of the character being present in the nodes are shown in Table 9. Some of the Mk1-rates were pretty high, and it was striking that the probabilities for these higher rates all showed a pattern, independent of the number of endnodes that actually comprised the character: the probability of its presence was 0.5. As the probabilities for character B showed no such pattern and the rates were relatively low as well, this character was used to further investigate this apparent trend.

Nodes	A	Aextra	B	Bextra	C	Cextra
all	0.5	0.4689	0.0319	0.3644	0.5	0.5
(-Out)	0.5	0.9362	0.0340	0.4444	0.5	0.5
((1,2,3),4)	0.5	0.9924	0.1393	0.6385	0.5	0.5
(1,2,3)	0.5	0.9998	0.0024	0.6558	0.5	0.5
(5,6)	0.5	0.9925	0.0057	0.2576	0.5	0.5
Species 1	**1**	**1**	0	**1**	**1**	**1**
Species 2	**1**	**1**	0	0	0	0
Species 3	0	**1**	0	**1**	**1**	**1**
Species 4	0	**1**	**1**	**1**	0	0
Species 5	**1**	**1**	0	0	0	0
Species 6	**1**	**1**	0	0	0	0
Outgroup	0	0	0	0	0	0
Rate	25	0.1100	0.1260	0.6724	10.945	10.945

Table 9 Probabilities for each character in all nodes. Nodes are represented as the most recent common ancestor of a subset of the tree (e.g. node (1,2,3) is the most recent common ancestor of Species 1, 2 and 3).Endnodes are represented by their taxon's name.

Table 10 shows the probabilities of the occurrence of character B in each node for the Mk1-rates, these rates times ten, and times hundred. When the rates are very high (original times 100) [5], the same pattern emerges. Again, the probabilities become 0.5 if the rates are nearing or transcending this threshold and have a 1:1 ratio. As one plus one is two, and the forward rate (and backward rate) is 1, it was suspected that the probabilities would simply adjust to the distribution of rates. Thus, if the backward rate would be twice or four times the forward rate (asymmetrical, 2 parameter Markov k-state), the probabilities were expected to become 0.3333 or 0.2, respectively (if the rates were very high). This indeed proved to be the case (Table 10).

Conclusions (ML)

Changing the representation from forward and backward rate to overall (Mk1) rate and bias is of no use as these are not independent (e.g. bias = forward rate/backward rate). The Mk1 or overall rate is scaled against the tree, and if a model's parameters are unspecified Mesquite estimates these parameters based on the data (Maddison and Maddison, 2010). Thus the differences in the Mk1 rates as estimated by Mesquite (Maddison and Maddison, 2010) could be less extreme and much smaller in the actual dataset, as provided by Dr. Kate Goodrich

Nodes	x1	x10	x100	1:01	1:02	1:04
(all)	0.0319	0.4595	0.5	0.5	0.3333	0.2
(-out)	0.034	0.4928	0.5	0.5	0.3333	0.2
(4(1,2,3))	0.1393	0.5300	0.5	0.5	0.3333	0.2
(1,2,3)	0.0024	0.3844	0.5	0.5	0.3333	0.2
(5,6)	0.0057	0.4200	0.5	0.5	0.3333	0.2
Species 1	0	0	0	0	0	0
Species2	0	0	0	0	0	0
Species3	0	0	0	0	0	0
Species4	**1**	**1**	**1**	**1**	**1**	**1**
Species5	0	0	0	0	0	0
Species6	0	0	0	0	0	0
Outgroup	0	0	0	0	0	0
Rates						
Forward	0.126	1.26	12.6	12.6	12.6	12.6
Backward	0.126	1.26	12.6	12.6	25.2	50.4

Table 10 Probabilities for character B in all nodes. Nodes are represented as the most recent common ancestor of a subset of the tree (e.g. node (1,2,3) is the most recent common ancestor of Species 1, 2 and 3).Terminals are represented by taxon names.

[5] Based on rates for other characters, the threshold was estimated to lie somewhere between 2.5 and 3.0.

(Goodrich and Raguso, 2009) for the actual tree (Neubig and Abbott, unpublished data).

Unfortunately, prior knowledge on rates is lacking and thus this optimization technique can only be of use in this particular study if the rates as estimated by Mesquite (Maddison and Maddison, 2010) are small. The asymmetric two-parameter model should only be applied if the bias is known.

Recommendations

Based on the outcomes of the tests with simulation-data, optimized over a simulation-tree, it was decided to use a symmetric stepmatrix for maximum parsimonious optimizations performed with real data and over the actual tree in this study. Only the original polymorphous data will be used, no extra (intermediate) character states will be added. Optimizations with maximum likelihood as a criterion will only be performed if the Mk1 rates (as estimated by Mesquite) are low. Although Mesquite allows users to set a bias, this is only employable when one knows the exact bias, as this can influence the outcome considerably. Applying this to the actual study this means that only the Mk1 model should be used, as no prior knowledge on the bias is available.

Computational Methods

Stepmatrices

Polymorph

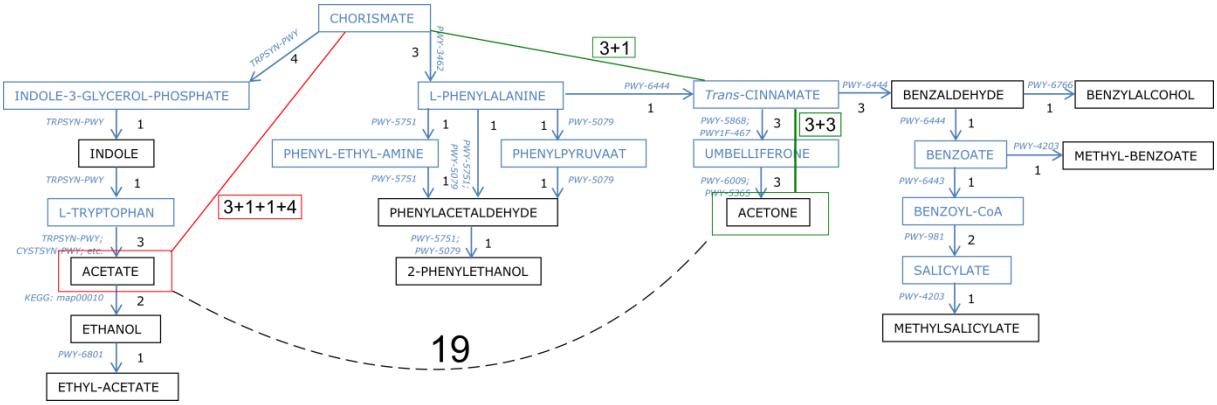


Figure 13 Chorismate pathway. The transition from Acetate to Acetone (or the other way around) would take 19 steps: 3+1+1+4 to get to Chorismate, plus 3+1+3+3 to go from Chorismate to Acetone.

For the polymorph characters (with every VOC, present in a species, as a character state), the pathway was the character. As the Terpenes-pathway was very large it was decided to split this pathway into two parts, at least for the polymorphic analyses. Thus, a total of three polymorphic characters was to be optimized over the trees: the Chorismate-pathway, the Monoterpenes' and the Sesquiterpenes' pathway. Each transition was assigned a certain cost, or number of steps. Figure 13 illustrates how the number of steps was determined for one of many transitions, namely Acetate changing into Acetone. To get from Acetate to Acetone, Acetate should be lost, as should be L-Tryptophan, Indole, Indole-3-glycerol-phosphate, all the way back to Chorismate. As it was decided to use only symmetric stepmatrices (see 'Simulation'), losing these components should be assigned the same number of steps as are needed to gain them. Thus, losing these four components means that 3 + 1 + 1 + 4 = 9 steps were taken. Now that Acetate is completely lost, it is time to calculate the number of steps, needed to get from Chorismate to Acetone. First of all, Chorismate should turn into L-Phenylalanine (3 steps), after which L-Phenylalanine should change into *trans*-Cinnamate (1 step), which should convert into Umbelliferone (3 steps). Last but not least, Umbelliferone should transform into Acetone (3 steps). To change Chorismate into Acetone 10 steps should be taken. As losing Acetate would take 9 steps, and gaining Acetone takes 10, this particular transition would be assigned a cost (number of steps) of 19 in the stepmatrix.
The same method was used for the Sesquiterpenes and the Monoterpenes.

As in both the Chorismate- and the Terpenes-pathway there was at least one transition that could occur via two different paths (with different numbers of steps, see Figure 6 and 7), two stepmatrices were created per character. Since the Monoterpenes and Sesquiterpens were linked to eachother, these two characters could be optimized using one big Terpene-stepmatrix. In total four stepmatrices were created to optimize the polymorphic characters over the trees, which were called PolyChor1, PolyChor2,

PolyTerp1 and PolyTerp2 (the first variants are calculated using the shortest pathway), and can be found in the Appendices.

Scenarios

As discussed on page 13, a scenario is a character state that comprises all components (that are present in a particular species) from one pathway. As the pathway is still the character, each species will have only one character state, and the problem with polymorphism is overcome.
Although this makes the analysis itself a lot simpler, the calculation of the number of steps needed to get from one scenario to another becomes a lot more complex.

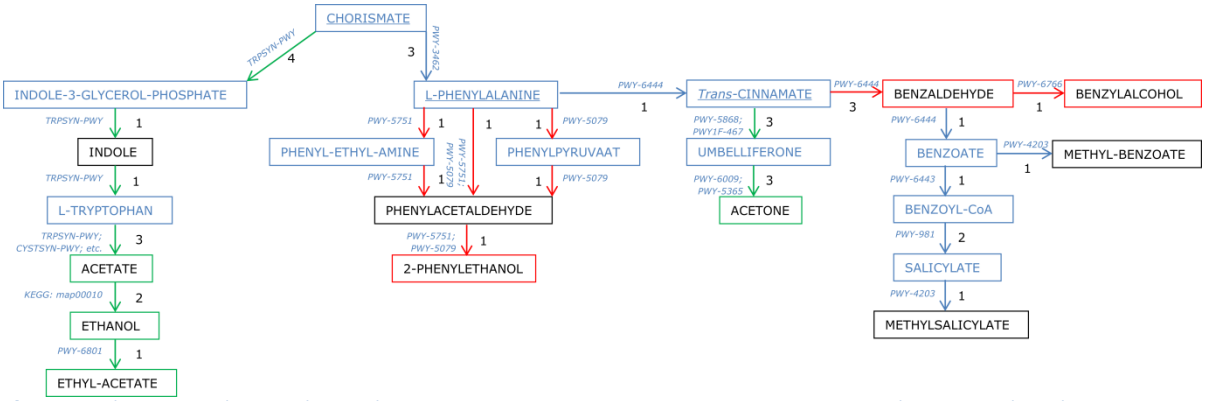


Figure 14: Chorismate pathway. Red rectangles comprises VOCs present in *Asimina tetramera*, green rectangles comprise those that occur in *Asimina parviflora*'s floral fragrance. Components whose names are underscored mark the end of a loss or start of a gain (e.g. Benzaldehyde is lost, but Acetone is gained: Losing stops at *trans*-Cinnamate, as this component is needed to make Acetone). Steps in this figure are based on pathways which are given in blue. Most are based on plantcyc.org, when they are based on kegg.jp, the pathway's name is preceded by 'KEGG'.

Figure 14 illustrates the cost-calculation for a real, but relatively simple, case. To change *Asimina reticulata* into *Asimina parviflora* in the Chorismate pathway, 2-Phenylethanol, Benzaldehyde and Benzylalcohol should change into Acetone, Acetate, Ethanol and Ethyl-acetate.

Losses			
	2-Phenylethanol	->	2 or 3 (depending on pathway)
	Benzaldehyde	->	3
	Benzylalcohol	->	1
Gains			
	Acetate	->	9
	Ethanol	->	2
	Ethyl-acetate	->	1
	Acetone	->	6
Cost			**24 or 25**

In all cases (Monoterpenes, Sesquiterpenes, Terpenes and Chorismate-derived compounds), pathways had multiple versions (Figure 6 and 7). In the case of the Monoterpenes and all Terpenes together, Limonene and α-Pinene can be formed via DMAPP or GPP, yielding different costs. In the Chorismate pathway, Phenylacetaldehyde can be derived from L-Phenylalanine directly, or via an intermediate (Phenyl-ethyl-amine or Phenylpyruvate). All possibilities are taken into account during the formation of different stepmatrices: one for each possible pathway. This led to a total of 7 stepmatrices: ScenChor1, ScenChor2, ScenMono1, ScenMono2, ScenSesqui1, ScenTerp1 and ScenTerp2. The first variant is always the one that takes into account the shortest possible pathway. These stepmatrices can be found in Table III in the Appendices.

Optimizations

Maximum Parsimony optimization was performed in Mesquite, version 2.7.4 (Maddison and Maddison, 2010), Mesquite, version 2.7.5 (Maddison and Maddison, 2011) and in the Phangorn-package in R (Schliep, 2012):

- Parsimony scores per tree and per (non-polymorphic) character were calculated in R (Phangorn package, Schliep, 2012).
- Graphics were created using Mesquite 2.7.4 (Maddison and Maddison, 2010).
- Parsimony scores per tree and polymorphic characters were calculated in Mesquite 2.7.4 (Maddison and Maddison, 2010).

Mesquite

Although Mesquite uses Fitch' Parsimony by default ('unordered') to reconstruct ancestral states, it allows users to create stepmatrices in which one can explicitly assign costs to change a character from one state into another. The maximum number of states is 56 (Maddison and Maddison, 2011). Mesquite supposedly allows hard polytomies when stepmatrices are used (Maddison and Maddison, 2002).

Trees used in the optimizations were the three *Asimina obovata*-placement trees ('Obov1+', 'Obov2+', and 'Obov3+') and their resolved versions (Obov1a – Obov1i, Obov2a – Obov2i, Obov3a – Obov3i). Characters that were defined are:

- Polymorphic Characters:
    - Monoterpenes
    - Sesquiterpenes
    - Chorismate
- Scenarios (non-polymorphic):
    - Monoterpenes (Monoterpene-Scenarios)
    - Sesquiterpenes (Sesquiterpene-Scenarios)
    - Terpenes (Terpene-Scenarios)
    - Chorismate (Chorismate-Scenarios)

And the stepmatrices to assign their costs:

- PolyTerp1 (Monoterpenes, Sesquiterpenes)
- PolyTerp2 (Monoterpenes, Sesquiterpenes)
- PolyChor1 (Chorismate)
- PolyChor2 (Chorismate)
- ScenMono1 (Monoterpene-Scenarios)
- ScenMono2 (Monoterpene-Scenarios)
- ScenSesqui1 (Sesquiterpene-Scenarios)
- ScenTerp1 (Terpene-Scenarios)
- ScenTerp2 (Terpene-Scenarios)
- ScenChor1 (Chorismate-Scenarios)
- ScenChor2 (Chorismate-Scenarios)

Ancestral state reconstruction in Mesquite was performed via the following steps:

- Analyses:
    - Tree Window
    - Trace Character History
    - Reconstruction Method: Parsimony Ancestral States
    - Source of Characters (for Reconstruct Ancestral States)
        - Stored Matrix: Character analysed
    - Source of parsimony models: Stored Parsimony Model
        - Stored Parsimony Model: Stepmatrix used in this analysis
- Save Tree as PDF

Phangorn (R)

In R-package Phangorn (Schliep, 2012), the parsimony scores[6] ('pScores') per tree can be calculated using Sankoff's Parsimony. Sankoff's Parsimony can deal with polytomies and assumes that there is a table that holds the costs for character state changes, which can be assigned by the user (Schliep, personal communication, March 21st, 2012). However, it cannot work with polymorphic characters, and a method to plot the optimizations over existing trees (in R) was not found. The script to calculate pScores, using Sankoff's Parsimony in Phangorn (R) can be found in the appendices ("R-script Phangorn").

[6] The pscore or parsimony score for each tree is the sum of the smallest number of substitutions needed. The tree with the lowest parsimony score is the most parsimonious tree.

Results

Lead-up

Resolved trees

As stated in Material & Methods, three unresolved trees were created, one for each possible placement of *Asimina obovata*. These trees were called Obov1+, Obov2+ and Obov3+. Each of these three trees was resolved, but only three trees will be shown in the result section, while the 27 others can be found in the appendices. The trees in the Appendices should be read from left to right, and down (thus, tree 1a is in the upper left corner, tree 3i is the one in the lower right). All figures containing unresolved trees should be read from left to right, since the Obov1+-tree is the tree on the left, and the Obov3+-tree is the tree on the right.

Stepmatrices

Within the results for all optimizations, stepmatrices for the same pathways were checked for similar pScores. Confidence intervals were calculated for pScores per character in the three tree subgroups (Obov1, Obov2, Obov3) for all characters in all resolved trees together and for all trees.
There were no differences at all in the parsimony scores (pScores) obtained using stepmatrices for the polymorphic characters (PolyChor1 and -2, PolyMono1 and-2), and thus only one stepmatrix per polymorphic character was used for further analyses (PolyChor1 and PolyMono1). Table 11 and 12 show the confidence intervals for the pScores for all trees and resolved trees only, respectively. Most of the confidence intervals overlap and the differences between Table 11 and Table 12 seem to be negligible. Still, both the unresolved and resolved trees will be discussed.

Table 11 Parsimony scores (confidence intervals) for all optimizations in Phangorn, R. Stepmatrices used for optimization are given. The lowest pScores are colored green, the highest pScores are colored red and the intermediates orange.

	Obov1		Obov2		Obov3	
PolyChor1	84.2	85.2	84.2	85.2	84.2	85.2
PolySesqui1	22.0	22.0	22.0	22.0	22.0	22.0
PolyMono1	24.0	24.0	24.0	24.0	24.0	24.0
ScenChor1	92.7	95.5	95.0	95.0	100.7	103.5
ScenChor2	94.7	97.5	97.0	97.0	103.7	106.5
ScenSesqui	44.2	45.2	43.9	45.3	43.9	45.3
ScenMono1	45.9	48.1	48.4	50.4	42.9	44.3
ScenMono2	52.3	55.1	55.1	57.5	50.5	52.5
ScenTerp1	105.3	108.1	106.8	108.2	99.8	101.2
ScenTerp2	112.9	116.1	114.8	116.6	108.8	110.6

Table 12 Parsimony scores (confidence intervals) for resolved optimizations in Phangorn, R. Stepmatrices used for optimization are given. The lowest pScores are colored green, the highest pScores are colored red and the intermediates orange.

	Obov1		Obov2		Obov3	
PolyChor1	84.2	85.2	84.2	85.2	84.2	85.2
PolySesqui1	22.0	22.0	22.0	22.0	22.0	22.0
PolyMono1	24.0	24.0	24.0	24.0	24.0	24.0
ScenChor1	92.5	95.5	95.0	95.0	100.5	103.5
ScenChor2	94.5	97.5	97.0	97.0	103.5	106.5
ScenSesqui	44.2	45.2	43.8	45.3	43.8	45.3
ScenMono1	45.8	47.9	48.3	50.3	42.8	44.3
ScenMono2	52.1	55.0	55.0	57.4	50.4	52.5
ScenTerp1	105.1	108.0	106.7	108.2	99.7	101.2
ScenTerp2	112.8	115.9	114.7	116.7	108.7	110.7

MP, polymorphic characters

Chorismate-pathway

The unresolved (Figure 16) and resolved (Figure II, Appendices) trees do not differ much in the reconstruction of ancestral states. In all cases, the most recent common ancestor possessed either:

- None of the components in the Chorismate-pathway,
- Acetate, or
- Indole

Ancestors higher up in the tree (within the clade that contains *Asimina parviflora*) almost always were reconstructed to have produced only ethanol.

In some of the resolved trees, however, the production of Ethanol was never given as ancestral (1c, 2c, 3c). In one case (1f) Ethyl-acetate was calculated to be a possible ancestral state (as was Ethanol) to the resolved clade containing *Asimina obovata*, *Asimina tetramera*, and *Duguetia rugellii*. The clade consisting of *Asimina incana, Asimina reticulata* and sometimes *Asimina obovata* yielded Benzaldehyde as its ancestral state, always (3a – 3i). In four cases (2h, 2i, 3d, 3e) the three components that were possibly ancestral in the deeper nodes return as a possibility in the higher nodes, after a short period of absence.

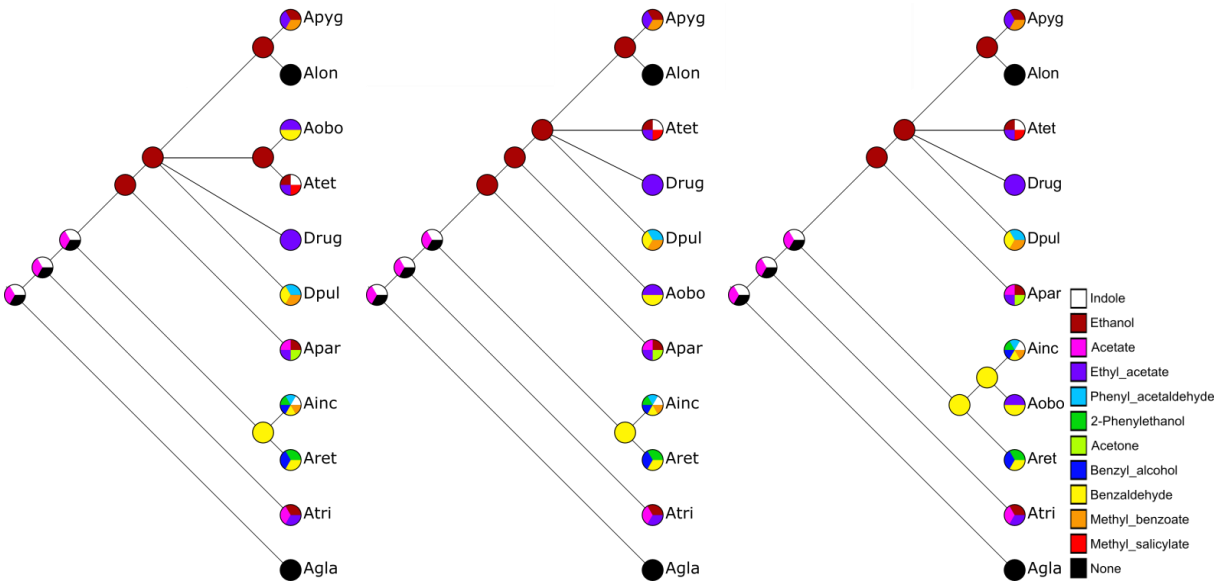


Figure 16 Unresolved trees representing the three different placements of *Asimina obovata*, showing the reconstruction of ancestral states using polymorphic characters (compounds from the Chorismate-pathway) and MP as a reconstruction method. Reconstruction model was a symmetrical stepmatrix based on the Chorismate-pathway (Figure 7).

Terpenes-pathway

Monoterpenes

The unresolved trees (Figure 18) show only one possible state for all ancestors (nodes of the tree):

- β-Myrcene.

Resolved trees (Figure V, Appendices) mostly show the same thing, but a little more than a quarter of all resolved trees (1a, 1f, 1g, 1h, 1i, 2a, 3a) is reconstructed with another, possibly, ancestral state. These trees yield either

- β-Myrcene, or
- α-Pinene

as the compound that is most likely to be produced by the most recent common ancestor of all *Asiminas*.

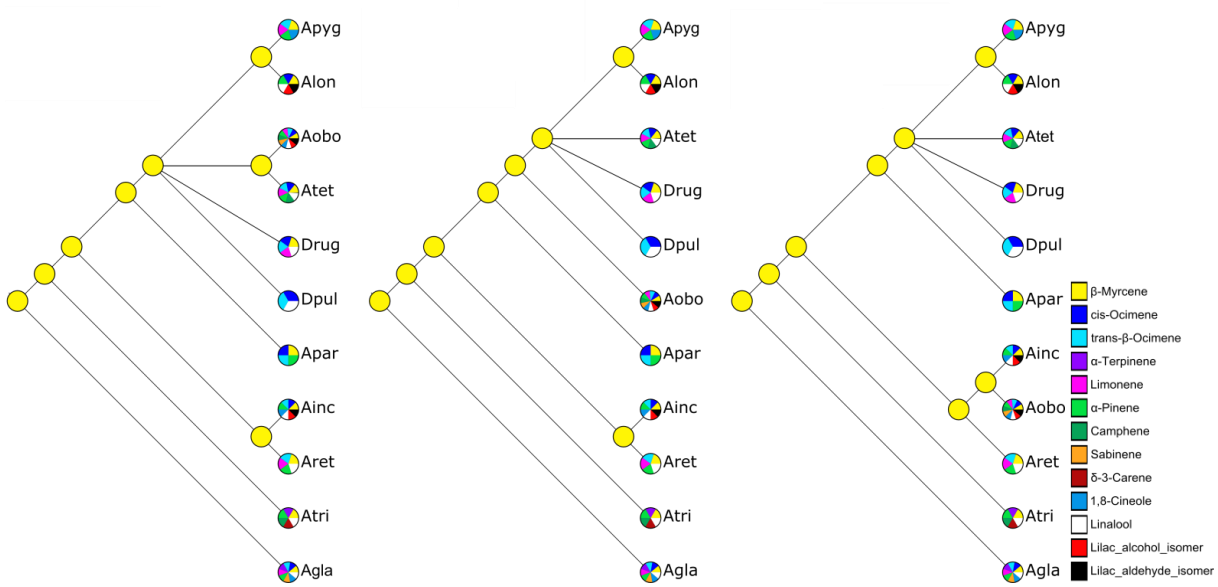


Figure 18 Unresolved trees representing the three different placements of *Asimina obovata*, showing the reconstruction of ancestral states using polymorphic characters (Monoterpenes) and MP as a reconstruction method. Reconstruction model was a symmetrical stepmatrix based on the Terpenes-pathway (Figure 6).

Notable is the fact that the reconstructed states in all nodes of one tree are identical to eachother, and thus whatever was the ancestral state, it was consistent throughout the whole tree.

Sesquiterpenes

In all cases, regardless of the placement of *Asimina obovata* in the tree, or whether the tree is resolved or not, the two deepest nodes are reconstructed with three possible ancestral states, while all ancestors higher up in the trees are calculated to have possibly produced only two of these compounds (see Figure 17). These chemical components are:

- Germacrene-D
- Trans-β-Caryophyllene, and
- δ-Cadinene (for the deeper nodes).

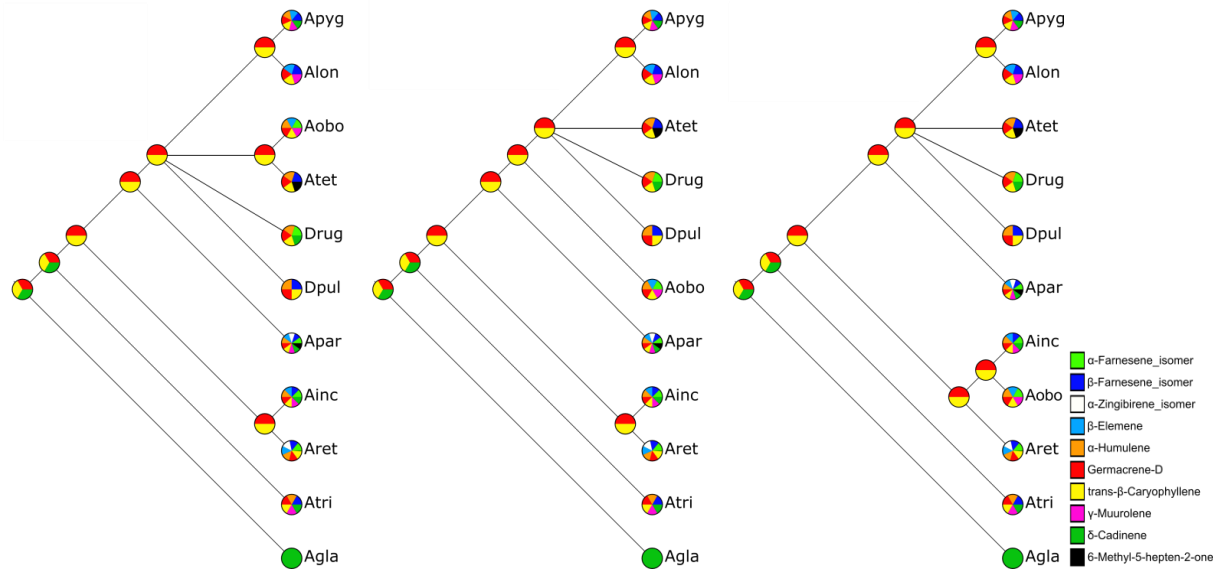


Figure 17 Unresolved trees representing the three different placements of *Asimina obovata*, showing the reconstruction of ancestral states using polymorphic characters (Sesquiterpenes) and MP as a reconstruction method. Reconstruction model was a symmetrical stepmatrix based on the Terpenes-pathway (Figure 6).

MP, scenarios

Chorismate-pathway

There are no visual differences between the optimizations that were created using two different stepmatrices (Chor1 and Chor2, see appendices). However, there are quite a few differences between trees.

Within the unresolved trees (Figure 19) only one thing changes: if *Asimina obovata* is in a clade with *Asimina incana* (trees 3a – 3i), all nodes (except for the two deepest) have *Asimina obovata*'s scenario as their state. If *Asimina obovata* is placed anywhere else (1a – 1i and 2a – 2i), the most recent common ancestor (MRCA) of *Asimina incana* and *Asimina reticulata* is reconstructed to be most likely to produce the same components as *Asimina incana*.
The two deepest nodes of the tree always yield the same, possibly, ancestral states: the *Deeringothamnus rugellii*-scenario and the *Asimina triloba*-scenario. The deepest node also yields the possibility that no components of the Chorismate-pathway were present in the floral fragrance of this most recent common ancestor of all the species in this tree.
Within the resolved trees, a lot is going on in the higher nodes, although the majority (trees 1a, 1c, 1d, 1f, 1h, 1i, 2a, 2c, 2d, 2e, 2h, 2i, 3a, 3c, 3e, 3i) shows the same pattern as the unresolved trees do. If *Asimina pygmaea*, *Asimina longifolia* and *Deeringothamnus rugellii* form a monophyletic clade (1b, 2b, 3b), at least the MRCA's for and within this clade had the *Deeringothamnus rugellii*- or *Asimina triloba*-scenario as their state. In tree 2b, the *Asimina obovata*-scenario was another possible state for these ancestors. Most ancestors between this particular clade en the two deepest nodes were still

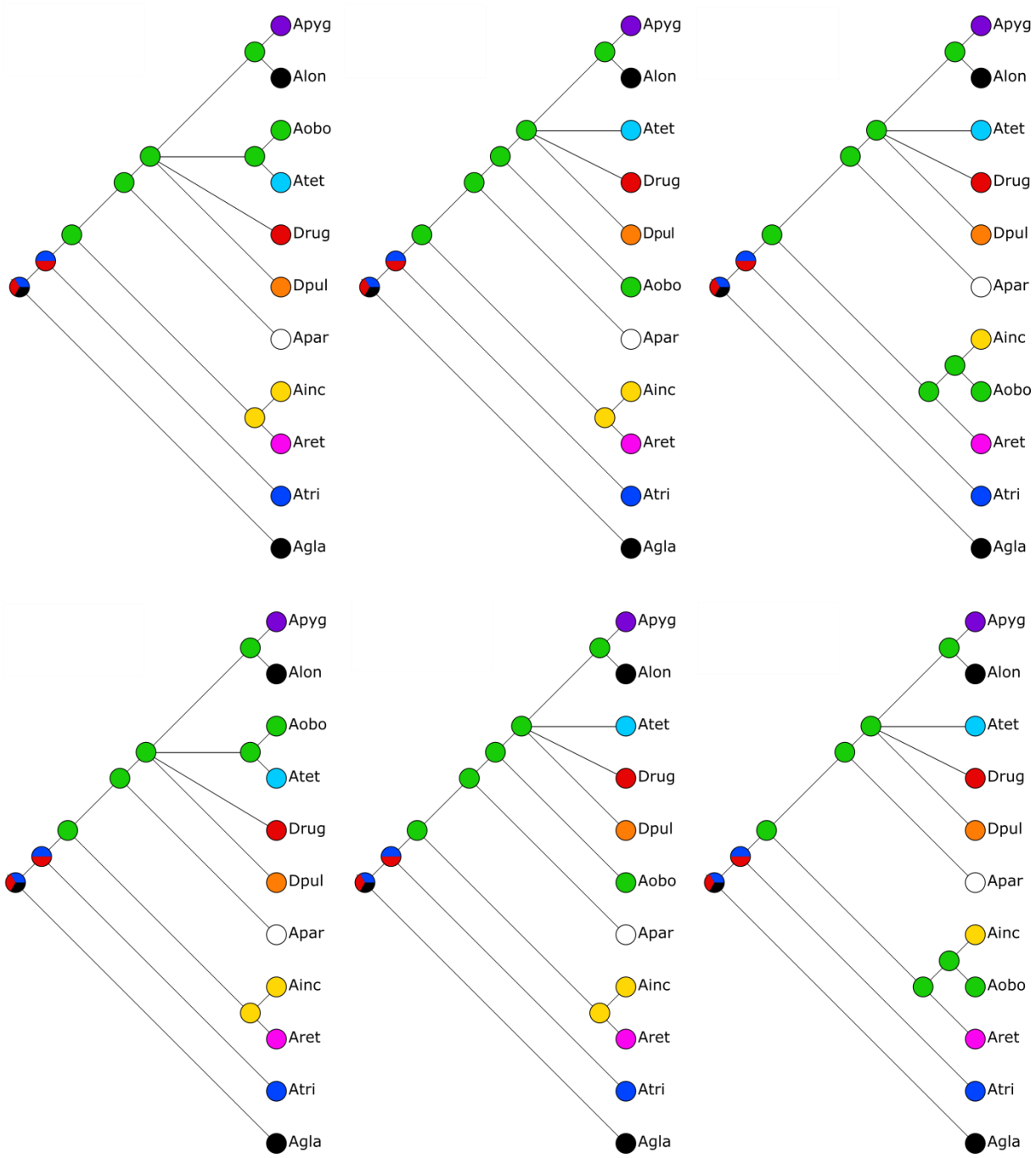


Figure 19 Unresolved trees representing the three different placements of *Asimina obovata*, showing the reconstruction of ancestral states using scenarios as characters (Chorismate-pathway) and MP as a reconstruction method. Every species has its own scenario, only *Asimina longifolia* and *Annona glabra* have the same scenario, as they both produce none of the components in the Chorismate-pathway. The upper three trees were reconstructed using a symmetrical stepmatrix that assigned a cost of 1 to the change of L-phenylalanine into Phenylacetaldehyde, the lower three with a symmetrical stepmatrix that assigned a cost of 2 to this same step. This approach is based on the Chorismate-pathway (Figure 7).

reconstructed to have produced the compounds as present in the *Asimina obovata*-scenario. When *Deerinthamnus pulchellus* and *Asimina tetramera* are grouped together in a monophyletic clade (2b, 3b) something similar happens. Besides the scenario as provided by *Asimina obovata*, their MRCA apparently can have another state: the *Asimina pygmaea*-scenario. Something else that is notable in the MRCA of these two species is the fact that the *Asimina obovata*-scenario is always an ancestral state, no matter what happens in the rest of the tree.

If the MRCA of *Deeringothamnus rugellii*, *Deeringothamnus pulchellus* and *Asimina tetramera* is positioned further back in the tree than the MRCA of only *Deeringothamnus pulchellus* and *Asimina tetramera* (1e, 1g, 3d, 3h), this MRCA has the *Asimina triloba*- or *Deeringothamnus rugellii*-scenario as a state, and so do most of the ancestors that are further back in the tree. However, the MRCA of *Asimina*

parviflora and others is calculated to be most likely to have produced chemical components like those in the scenario of *Deeringothamnus rugellii*.

Terpenes-pathway

Monoterpenes

As is obvious from the unresolved trees alone (Figure 20), there are quite a few differences between the ancestral states as reconstructed with the help of the two different stepmatrices. First, the unresolved trees and patterns that occur in both versions of the resolved trees will be discussed, after which the optimizations over the resolved trees will be discussed per stepmatrix.

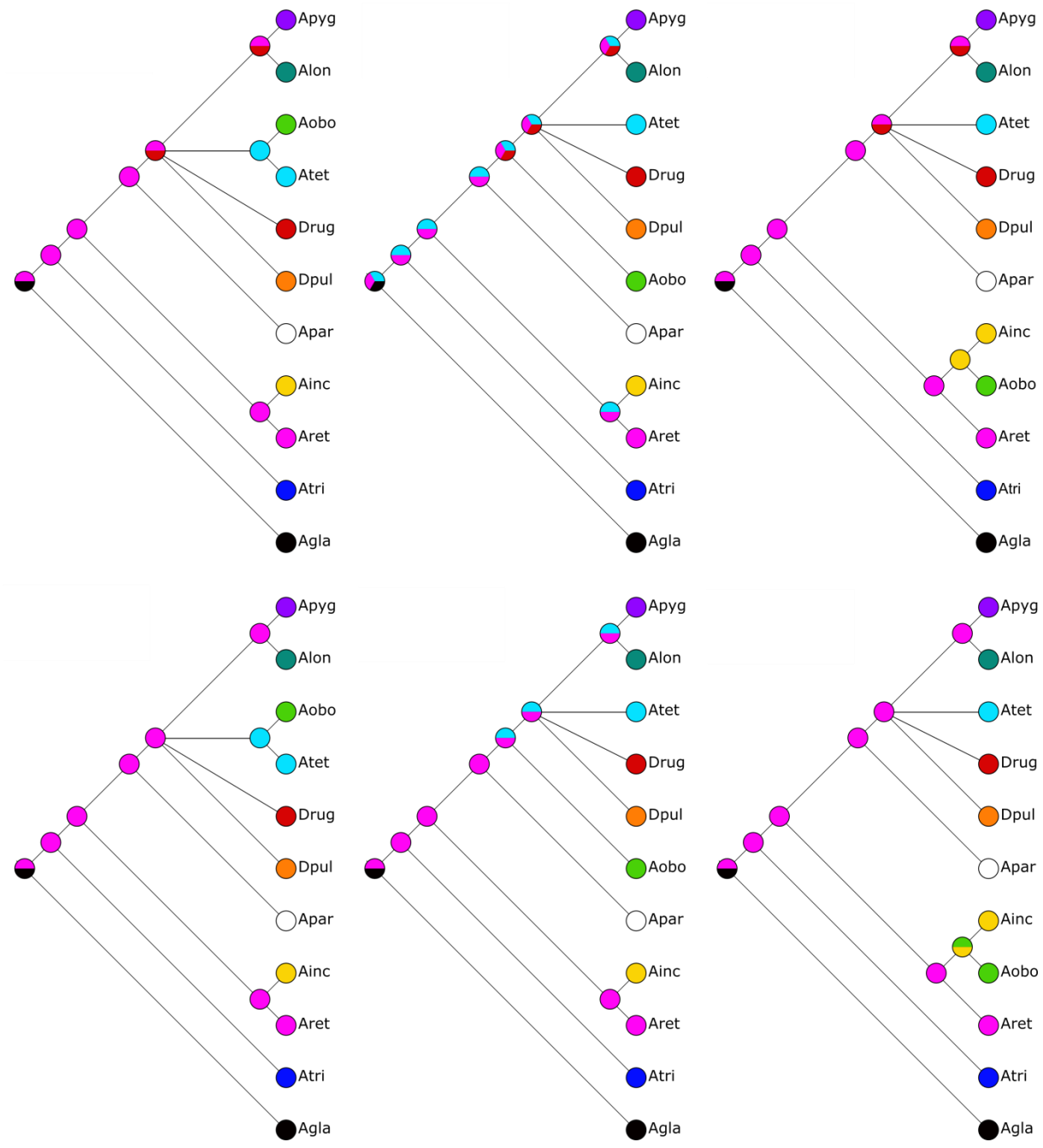


Figure 20 Unresolved trees representing the three different placements of *Asimina obovata*, showing the reconstruction of ancestral states using scenarios as characters (Terpenes-pathway) and MP as a reconstruction method. Every species has its own scenario. The upper three trees were reconstructed using a symmetrical stepmatrix (MonoScen1) that assigned a cost of 1 to the production of Limonene and α-Pinene, the lower three with a symmetrical stepmatrix (MonoScen2) that assigned a cost of 2 to the production of these compounds. This approach is based on the Terpenes-pathway (Figure 6).

Unresolved trees and general patterns

In all unresolved trees, the *Asimina reticulata*-scenario is (one of) the ancestral state(s) in all deeper nodes. Another scenario that is possibly ancestral in some of the deeper nodes is the scenario of *Asimina tetramera*. Within the resolved trees it happens twice (3e and 3g) that the *Asimina reticulata*-scenario is not ancestral to the majority of the tree, while *Asimina tetramera*'s scenario is. Both trees that show this pattern have *Asimina obovata* placed in a monophyletic taxon with *Asimina incana*. In the Obov2-trees something similar happens. If *Asimina tetramera* is placed further back in the tree than *Deeringothamnus rugellii* and *Deeringothamnus pulchellus* (2f, 2i), are the MRCA of these two plus *Asimina tetramera* was calculated to have produced the same components as did *Asimina tetramera*. The same accounts for the MRCA of *Asimina tetramera* and *Asimina obovata*.

In all trees, resolved or unresolved, the most recent common ancestor of all species in these trees could also have produced a fragrance that was similar to that of *Annona glabra*.

Whenever *Asimina tetramera* and *Asimina obovata* were grouped together (1a – 1i), *Asimina tetramera*'s scenario was ancestral to these two species. If *Deeringothamnus rugellii* and *Deeringothamnus pulchellus* form a monophyletic clade (1a, 1h, 1i, 2a, 3a), their ancestor is bound to have produced the same scenario as did *Deeringothamnus rugellii*. Grouping *Asimina incana* and *Asimina obovata* (3a – 3i) led to a most recent common ancestor that had either one of these scenarios as its character state.

If *Deeringothamnus pulchellus* and *Deeringothamnus rugellii* are grouped together in a monophyletic clade, ór if their occurrence in the tree would be (((others), Dpul), Drug) or (((others), Drug), Dpul), like in trees 1b, 1d, 1f, 1g, 1h, 1i, 2a, 2d, 2e, 2f, 2i, 3a, 3e, 3g, 3h, 3i, their MRCA and at least the next ancestor higher up in the tree could have possessed the *Deeringothamnus rugellii*-scenario.

MonoScen1

All resolved Obov2-trees were reconstructed to have both the *Asimina reticulata*- and the *Asimina tetramera*-scenario as the possible states of the deeper nodes. In tree 1a and 3a, the scenario as produced by *Asimina parviflora* was ancestral in most of the nodes of these two trees. Twice (2c and 2h), this particular scenario also was a possible character state for some ancestors higher up in the tree.

MonoScen2

In trees 1c, 1d, 1f, 3c and 3i, the *Asimina parviflora*-scenario is a possible character state in the five deepest nodes of the trees, at the least. In tree 1a and 3a almost all ancestral nodes have this scenario as their possible state.
All Obov2-trees have the *Asimina reticulata*-scenario as the only state in most of the deeper nodes. Only the MRCA of all species in the tree could have had the same scenario as *Annona glabra*. Higher up in these trees, the scenario, as present in *Asimina tetramera*, becomes a possible ancestral state.
In three trees (1d, 1f and 3i), the MRCA of *Deeringothamnus pulchellus* and any other species possibly had the same scenario as *Deeringothamnus pulchellus* as its state. However, there are always other possibilities.

Sesquiterpenes

All optimizations show that the MRCA of all species in these trees had either the scenario as present in *Asimina triloba* or in *Annona glabra* as their character state. Going higher up in the tree, the *Asimina triloba*-scenario remains ancestral in most cases (20 out of 27 trees, not 1f, 1g, 1i, 2e, 2g, 3d, 3g). In the unresolved trees (Figure 21) this is always the case. In all trees, resolved or unresolved, the MRCA of all species in the tree, minus *Asimina triloba* and *Annona glabra*, had *Asimina incana*'s scenario as (possible) character state.

In almost all trees (26 out of 30, not 2e, 2g, 3d, 3g) the *Asimina pygmaea*-scenario is possibly ancestral to a great deal of the species. Higher up in the trees, the scenario as present in *Deeringothamnus pulchellus* becomes a possible ancestral state. In the case of the Sesquiterpenes, all trees seem to differ just a little bit in the possible ancestral states of the higher nodes, and it was impossible to discern a general pattern. Four trees that are very similar to eachother, but pretty different from the rest are tree 2e, 2g, 3d and 3g. In these trees there is just one possibility as to the character state in each ancestor in the majority of the nodes, and the ancestral states succeed eachother in the following way: *Asimina triloba*-scenario, *Asimina incana*-scenario, *Deeringothamnus rugellii*-scenario, *Deeringothamnus pulchellus*-scenario.

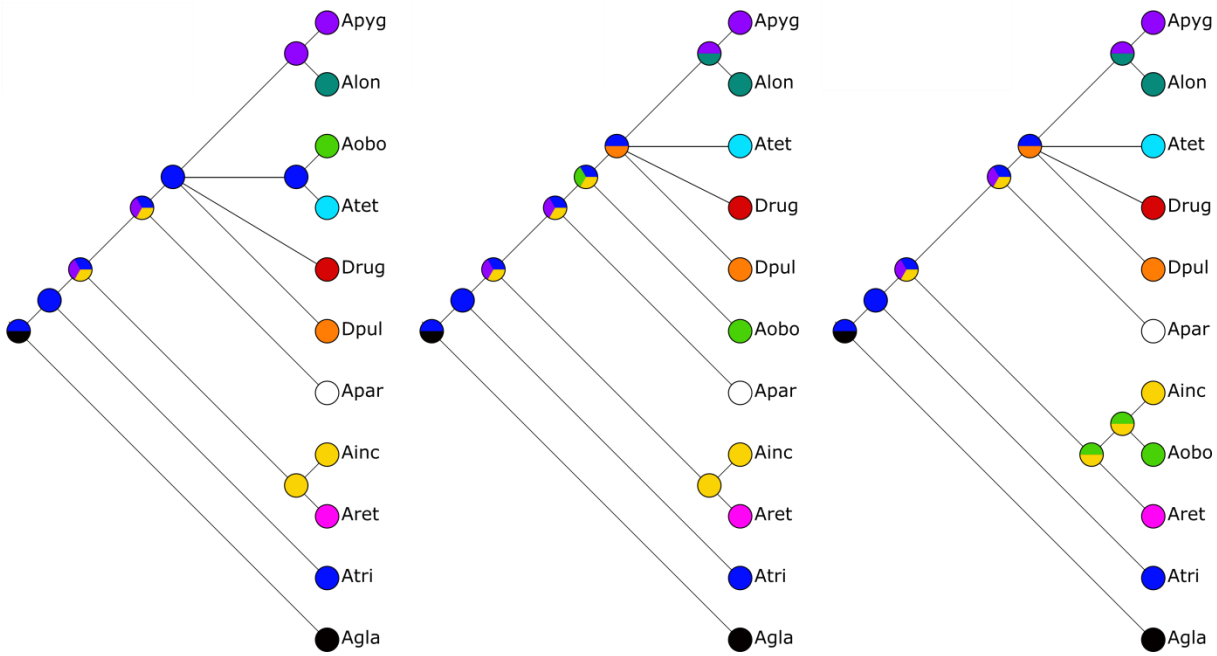


Figure 21 Unresolved trees representing the three different placements of *Asimina obovata*, showing the reconstruction of ancestral states for sesquiterpenes, using scenarios as characters (Terpenes-pathway (Figure 6)) and MP as a reconstruction method. Every species has its own scenario. As there was only one stepmatrix, just three trees are shown.

When *Asimina incana* and *Asimina reticulata* are directly grouped together (1a – 1i, 2a – 2i), their most recent common ancestor always produced the scenario as present in *Asimina incana*. Whenever *Asimina obovata* joins this clade (Obov3-trees), another possible ancestral state would be the scenario from *Asimina obovata,* except for 3d and 3h.

Terpenes (Mono- and Sesqui- together)

As the unresolved trees (Figure 22) already imply, the ancestral states that were reconstructed using the two stepmatrices, TerpScen1 and TerpScen2, differed a lot. Thus, the results will be discussed per stepmatrix.

TerpScen1

The majority of the optimizations (1a, 1b, 1c, 1d, 1h, 2a, 2b, 2d, 2f, 2g, 3a, 3b, 3f, 3g, 3i) has the *Asimina triloba*-scenario as the only ancestral state in the MRCA of all *Asimina*s. Trees 1a, 1b, 1d, 1e, 2a, 2b, 3a and 3b were reconstructed to have only *Asimina reticulata*'s scenario as the ancestral state, higher up in the tree (1h was almost exactly like these trees, were it not that one node had *Deeringothamnus rugellii*'s scenario as its character state). This was partly true in the case of 2d, 2f, 2g, 3f, 3g and 3i, but in this case there was another possibility too. In these trees, the MRCA (and all subsequent ancestors) of *Asimina pygmaea*, *Asimina longifolia*, *Asimina tetramera*, *Deeringothamnus pulchellus* and *Deeringothamnus rugellii* had *Deeringothamnus pulchellus*' scenario as a possible ancestral state as well.

Asimina triloba's scenario being one of two possibilities (the other being the scenario as produced by *Deeringothamnus rugellii*) also occurs in many trees (1c, 2c, 2e, 2h, 2i, 3c, 3d, 3h). However, in three of the trees (1f, 1g, 1i) *Asimina triloba*'s scenario is not presented as possibly ancestral once, while the scenario as present in *Asimina pygmaea* dominates most of the ancestral nodes of these trees. In these cases, the ancestral nodes that are higher up in the tree have the *Deeringothamnus rugellii*-scenario as their one and only character state.

Tree 1c, 2c, 2i, 3c and 3e were some kind of a mix of the trees already described. In the lowest part of the tree it was either *Deeringothamnus rugellii's* or *Asimina triloba*'s scenario that was the ancestral state. A little higher up in the tree *Asimina reticulata* or *Deeringothamnus rugellii* had the possibly most ancestral state, while in the top part it was *Asimina reticulata's* or *Deeringothamnus pulchellus*' scenario that must have been ancestral.
Trees 2e, 2h, 3d and 3h are very similar, were it not that the scenario as produced by *Asimina reticulata* is no longer present in the ancestral nodes that are higher up in the tree.

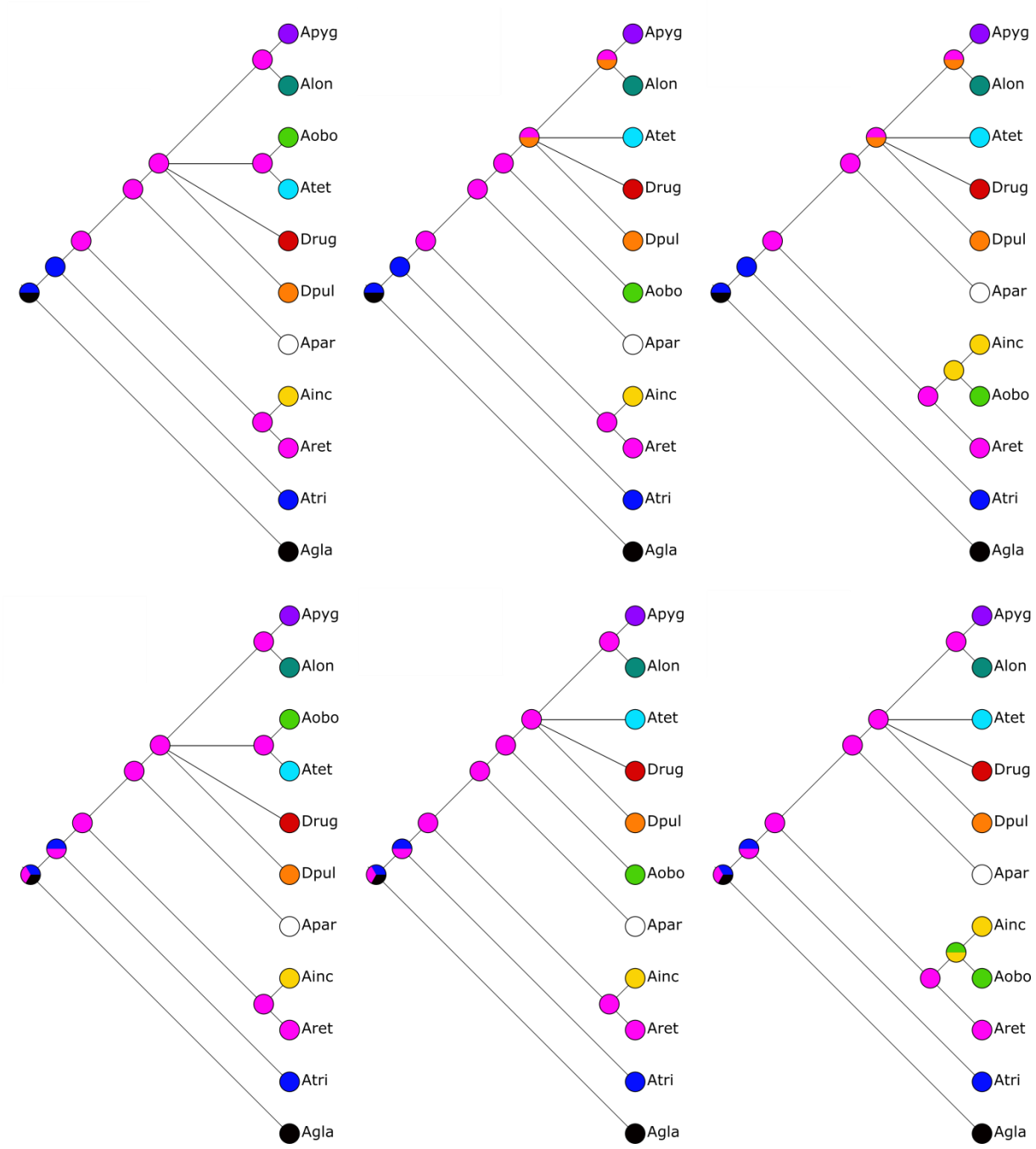


Figure 22 unresolved trees representing the three different placements of *Asimina obovata*, showing the reconstruction of ancestral states using scenarios as characters (Terpenes-pathway) and MP as a reconstruction method. Every species has its own scenario. The upper three trees were reconstructed using a symmetrical stepmatrix (TerpScen1) that assigned a cost of 1 to the production of Limonene and α-Pinene, the lower three with a symmetrical stepmatrix (TerpScen2) that assigned a cost of 2 to the production of these compounds. This approach is based on the Terpenes-pathway (Figure 6).

As the unresolved trees (Figure 22) already imply, the ancestral states that were reconstructed using the two stepmatrices, TerpScen1 and TerpScen2, differed a lot. Thus, the results will be discussed per stepmatrix.

TerpScen1

The majority of the optimizations (1a, 1b, 1c, 1d, 1h, 2a, 2b, 2d, 2f, 2g, 3a, 3b, 3f, 3g, 3i) has the *Asimina triloba*-scenario as the only ancestral state in the MRCA of all *Asimina*s. Trees 1a, 1b, 1d, 1e, 2a, 2b, 3a and 3b were reconstructed to have only *Asimina reticulata*'s scenario as the ancestral state, higher up in the tree (1h was almost exactly like these trees, were it not that one node had *Deeringothamnus rugellii*'s scenario as its character state). This was partly true in the case of 2d, 2f, 2g, 3f, 3g and 3i, but in this case there was another possibility too. In these trees, the MRCA (and all subsequent ancestors) of *Asimina pygmaea*, *Asimina longifolia*, *Asimina tetramera*, *Deeringothamnus*

pulchellus and *Deeringothamnus rugellii* had *Deeringothamnus pulchellus*' scenario as a possible ancestral state as well.

Asimina triloba's scenario being one of two possibilities (the other being the scenario as produced by *Deeringothamnus rugellii*) also occurs in many trees (1c, 2c, 2e, 2h, 2i, 3c, 3d, 3h). However, in three of the trees (1f, 1g, 1i) *Asimina triloba*'s scenario is not presented as possibly ancestral once, while the scenario as present in *Asimina pygmaea* dominates most of the ancestral nodes of these trees. In these cases, the ancestral nodes that are higher up in the tree have the *Deeringothamnus rugellii*-scenario as their one and only character state.

Tree 1c, 2c, 2i, 3c and 3e were some kind of a mix of the trees already described. In the lowest part of the tree it was either *Deeringothamnus rugellii'*s or *Asimina triloba*'s scenario that was the ancestral state. A little higher up in the tree *Asimina reticulata* or *Deeringothamnus rugellii* had the possibly most ancestral state, while in the top part it was *Asimina reticulata'*s or *Deeringothamnus pulchellus*' scenario that must have been ancestral.
Trees 2e, 2h, 3d and 3h are very similar, were it not that the scenario as produced by *Asimina reticulata* is no longer present in the ancestral nodes that are higher up in the tree.

TerpScen2

Tree 1b, 1c, 1d, 1e, 2d, 2e, 2f, 2g, 2h, 2i and all Obov3-trees have either the *Asimina triloba*- or *Asimina tetramera*-scenario as their overall, possibly ancestral state (and ofcourse the *Annona glabra*-scenario). The trees that continue to keep only the scenario as produced by *Asimina tetramera* as their ancestral state are 1b, 1c, 1d, 1e, 2d, 2e, 2f, 2g, 2h and all Obov3-trees, except for 3e. Two of these trees have two other possible states within the ancestors that are higher up in the tree: *Deeringothamnus rugellii*'s or *Deeringothamnus pulchellus*' scenario.

Tree 1f and 1g have only the scenario as present in *Asimina pygmaea* as their utmost ancestral state (except for *Annona glabra*'s scenario), and continue with only *Deeringothamnus rugellii*'s scenario, higher up in the tree.

Trees that have either (*Annona glabra*'s) *Asimina triloba*'s, *Asimina tetramera*'s or *Asimina pygmaea*'s scenario as the character state in their deeper nodes are 1a, 1h, 1i, 2a, 2b and 2c. Tree 1a, 2a, 2b and 2c continue with only the scenarios as produced in *Asimina tetramera* and *Asimina pygmaea* as possible ancestral states, higher up in the tree. In tree 1h, *Deeringothamnus rugellii*-scenario, *Deeringothamnus pulchellus*-scenario, and eventually only *Deeringothamnus rugellii*-scenario are the most possible ancestral states within the higher nodes. Within tree 1i, it is either the scenario from *Deeringothamnus rugellii*, *Asimina tetramera* or *Asimina pygmaea.*

MP, pScores

Relative pScores (pScore divided by highest pScore) were summed for all scenarios, the Chor, Mono and Sesqui scenarios, for the Chor and Terp scenarios together, for Chor, Sesqui, and Mono scenarios plus polymorphic and for all together (see Figure 13). The Obov2-trees usually had the highest pScores, which means they were the least parsimonious for most optimizations. The 'best' trees (the ones for which the summed pscores were lowest) were Obov1g, Obov3a and Obov3b, of which Obov1g was the best most often. If any Obov3 trees had low pScores, a lot of them had low pScores. Obov1-trees on the other hand were 'good' trees often, but there were less of them.

Table 13 Relative pScores per tree, summed over all scenarios, chorismate and monoterpenes and sesquiterpenes, chorismate and terpenes, chorismate and sesquiterpenes and polymorphic, and All. The highest pscore-sums are colored red(dish), the lowest (more parsimoniuous) scores are colored green(ish).

	SUM				
Tree	Scen only	Chor+Mono+ Sesqui	Chor+Terp	Chor+Sesqui +Mono+Poly	All
Obov1+	7.76	5.76	3.84	11.76	11.76
Obov1a	7.63	5.67	3.79	11.67	11.63
Obov1b	7.70	5.70	3.78	11.70	11.70
Obov1c	7.76	5.76	3.84	11.74	11.74
Obov1d	7.76	5.76	3.84	11.74	11.74
Obov1e	7.70	5.70	3.78	11.70	11.70
Obov1f	7.59	5.64	3.78	11.62	11.59
Obov1g	7.53	5.59	3.73	11.59	11.55
Obov1h	7.63	5.67	3.79	11.67	11.63
Obov1i	7.57	5.63	3.78	11.63	11.59
Obov2+	7.84	5.84	3.84	11.84	11.84
Obov2a	7.70	5.74	3.79	11.74	11.70
Obov2b	7.74	5.75	3.83	11.75	11.78
Obov2c	7.84	5.84	3.84	11.81	11.81
Obov2d	7.84	5.84	3.84	11.81	11.81
Obov2e	7.78	5.79	3.83	11.79	11.81
Obov2f	7.76	5.76	3.84	11.76	11.76
Obov2g	7.84	5.84	3.84	11.81	11.81
Obov2h	7.78	5.79	3.83	11.79	11.81
Obov2i	7.76	5.76	3.84	11.76	11.76
Obov3+	7.68	5.79	3.88	11.79	11.68
Obov3a	7.54	5.70	3.84	11.70	11.54
Obov3b	7.52	5.65	3.82	11.65	11.56
Obov3c	7.68	5.79	3.88	11.77	11.65
Obov3d	7.56	5.69	3.82	11.69	11.59
Obov3e	7.64	5.75	3.88	11.75	11.64
Obov3f	7.68	5.79	3.88	11.77	11.65
Obov3g	7.64	5.75	3.88	11.75	11.64
Obov3h	7.56	5.69	3.82	11.69	11.59
Obov3i	7.68	5.79	3.88	11.77	11.65

Table 14 Maximum likelihood rates (Mk1 model) per compound and per unresolved tree. High rates were marked orange.

	Obov1+	Obov2+	Obov3+
INDOLE	0.1593	0.1480	0.1468
ETHANOL	3.3604	2.8804	3.3604
ACETATE	0.1825	0.1911	0.1845
ETHYL_ACETATE	3.3604	2.8804	3.3604
PHENYL_ACETALDEHYDE	0.1481	0.1480	0.1468
2-PHENYLETHANOL	0.0690	0.0691	0.1517
ACETONE	0.0641	0.0647	0.0641
BENZYL_ALCOHOL	0.0690	0.0691	0.1517
BENZALDEHYDE	0.7320	0.4688	0.1649
METHYL_BENZOATE	3.3604	5.7607	3.3604
METHYL_SALICYLATE	0.0634	0.0597	0.0597
β-MYRCENE	0.0597	0.0597	0.0597
cis-OCIMENE	0.3814	0.3911	0.3988
trans-β-OCIMENE	0.1706	0.1713	0.1714
α-TERPINENE	0.0720	0.0721	0.0721
LIMONENE	3.3604	5.7607	3.3604
α-PINENE	0.1560	0.1504	0.1507
CAMPHENE	0.1726	0.5836	3.3604
SABINENE	0.1450	0.1503	0.1501
δ-3-CARENE	0.0684	0.0684	0.0684
1,8-CINEOLE	3.3604	1.0190	0.3195
LINALOOL	0.1608	0.1653	0.1611
LILAC_ALCOHOL_ISOMER	3.3604	0.7432	0.1689
LILAC_ALDEHYDE_ISOMER	3.3604	0.7432	0.1689
α-FARNESENE_ISOMER	0.9956	0.3959	0.3205
β-FARNESENE_ISOMER	0.3367	0.3581	0.3206
α-ZINGIBIRENE_ISOMER	0.1738	0.1826	0.1834
β-ELEMENE	0.6796	0.3940	0.3279
α-HUMULENE	0.1450	0.1452	0.1452
GERMACRENE_D	0.0613	0.0613	0.0613
trans-β-CARYOPHYLLENE	0.0613	0.0613	0.0613
γ-MUUROLENE	3.3604	0.7422	0.5876
δ-CADINENE	3.3604	5.7607	3.3604
6-METHYL-5-HEPTEN-2-ONE	0.1608	0.1533	0.1534

ML

Rates

As described in 'Simulation', sometimes the rates as calculated by Mesquite were too high to yield any interesting probabilities. Therefore, before any of the optimizations' likelihoods were analyzed, all rates were checked (Table 14).

Scenario-components

Special interest was taken in the scenarios that were ancestral to the majority of the tree. In other words, the character states that were reconstructed to be possibly ancestral in the trees' four deepest nodes were deemed most interesting. Since *Annona glabra* is the outgroup, and does not actually belong to the same clade as the other species, the MRCA of this species and all others was not taken in to account. Table 15 shows the ancestral scenarios that were calculated to be the possible character states of the ancestors of *Asimina triloba* and all others (except *Annona glabra*) and of *Asimina tetramera* and all others (except for *Annona glabra* and *Asimina triloba*), and the pathways for which they were found to be ancestral. In Table 16, the chemical components that are present in each scenario are listed. The components that were already found to be ancestral during the polymorphic optimizations were made black. Components that had rates that were too high (the orange compounds in Table 14) were stricken through, as they could not be used for further research on likelihoods. The chemical components that were selected for optimizations using Maximum Likelihood as criterion are:

- Acetate
- Germacrene-D
- trans-β-Caryophyllene
- β-Myrcene
- α-Pinene
- β-Farnesene isomer
- α-Humulene
- trans-β-Ocimene

Table 15 Ancestral scenarios and the pathways in which the components, present in each scenario, occur.

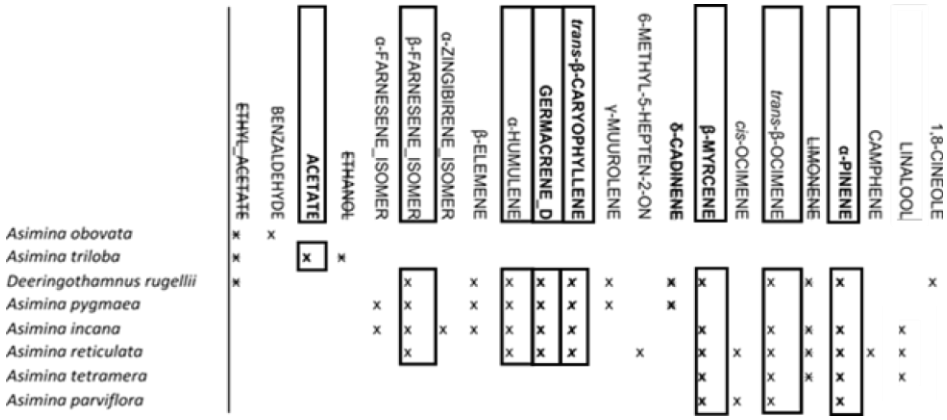
	Chor	Sesqui	Mono	Terp
Asimina obovata	x			
Asimina triloba	x	x		x
Deeringothamnus rugellii	x			x
Asimina pygmaea		x		x
Asimina incana		x		
Asimina reticulata			x	x
Asimina tetramera			x	
Asimina parviflora			x	

These components were selected because they either were found to be ancestral during the MP-optimization of the polymorphic characters, because they occurred in the majority of scenarios that were possibly ancestral (ancestral scenarios as found for the different pathways can be found in Table 15), or both.

Optimizations

To keep things clear, ML probabilities were only registered for the four most ancestral nodes (Figure 23). These nodes are very similar in all trees, resolved or unresolved and independent of the placement of *Asimina obovata*. Optimizations were performed over unresolved trees only. Table 17 shows these probabilities, per component, per tree, per node.

Table 16 Ancestral scenarios and their components. Components that had ML rates that were high (see Table 14) were stricken through, components that were already found to be ancestral during the MP optimizations of polymorphic characters were made black. Components that were selected for further research using Maximum Likelihood optimizations were put in black boxes.



Two components (β-Myrcene and α-Pinene) are very likely to have been present in the most recent common ancestor of all species in the tree, since all nodes (or all except for node 4) show a probability that is higher than 0.95. Germacrene-D and trans-β-Caryophyllene are calculated to have been present in the MRCA of all *Asimina*s, but not in the MRCA of these plus *Annona glabra*.

Trans-β-Humulene, β-Farnesene isomer and α-Humulene have a more complex story to tell. In nodes 3 and 4, α-Humulene is bound to have been present. In the second node the probability is decreased slightly, to 0.92, and in the first node it drops below 0.5. Thus, this component was not present in the MRCA of all species in the tree, but could have been present in the MRCA of all *Asimina*s and is very probably was in the the 3rd and 4th node.

Trans-β-Ocimene was present in the fourth node at least, but is also likely to have been one of the components produced by the ancestors in node 1 till 3. The probabilities for these nodes range between 0.71 and 0.94.

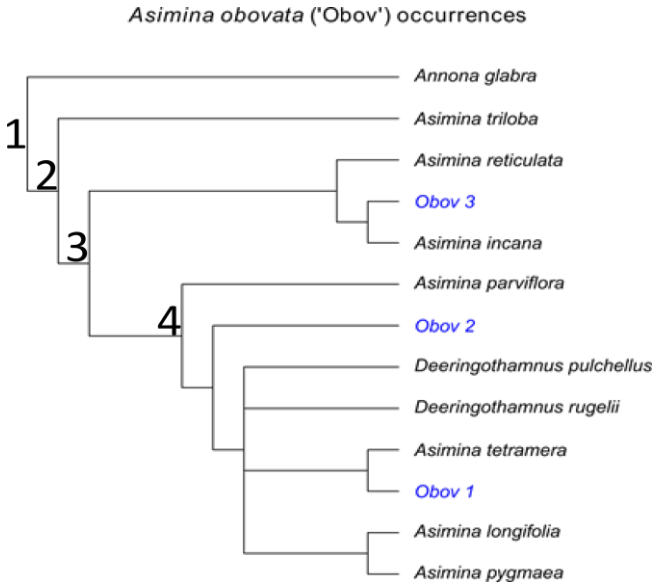


Figure 23 Cladogram displaying all occurrences of *Asimina obovata* at the same time. The four nodes for which the probabilities are given in Table 17 are given the numbers 1 – 4.

Table 17 ML probabilities for the four nodes as specified in Figure 23. Probabilities over 0.95 are colored dark green, between 0.8 and 0.95 green, between 0.6 and 0.8 light green, between 0.4 and 0.6 yellow and between 0.2 and 0.4 orange. The components that were most likely to have been ancestral to (at least) all species in the genus *Asimina* Adans. were made black.

	Obov 1+				**Obov 2+**				**Obov 3+**			
	1	2	3	4	1	2	3	4	1	2	3	4
Acetate	0.230	0.423	0.250	0.328	0.243	0.447	0.280	0.373	0.233	0.429	0.257	0.333
Germacrene-D	0.484	0.968	0.998	1.000	0.484	0.968	0.998	1.000	0.484	0.968	0.998	1.000
trans-β-Caryophyllene	0.484	0.968	0.998	1.000	0.484	0.968	0.998	1.000	0.484	0.968	0.998	1.000
β-Myrcene	0.996	1.000	1.000	1.000	0.996	1.000	1.000	1.000	0.996	1.000	1.000	1.000
α-Pinene	0.972	0.990	0.988	0.965	0.974	0.992	0.994	0.992	0.974	0.991	0.990	0.971
β-Farnesene isomer	0.412	0.784	0.866	0.891	0.404	0.756	0.820	0.804	0.410	0.781	0.851	0.923
α-Humulene	0.462	0.920	0.984	0.995	0.461	0.920	0.983	0.994	0.461	0.920	0.983	0.995
trans-β-Ocimene	0.837	0.714	0.942	0.987	0.836	0.712	0.940	0.985	0.836	0.712	0.940	0.986

The β-Farnesene isomer definitely was not present in the most recent common ancestor of all species' fragrance, but there is a slight chance that it was one of the components in the fragrance of those ancestors that are represented by node 3 and 4. The probability of it being present in these nodes ranges from 0.82 to 0.923.

Acetate on the other hand provides a clearer message: it was not present in any of the more ancestral nodes, according to this optimization using Maximum Likelihood as the algorithm.

Discussion

Many assumptions were made in order to perform this study and every one of these assumptions could be criticized. Since this study was intended as a pilot study, this discussion primarily focuses on future improvements in similar research. The result is a long list of tasks that should be completed before a study such as this could be performed again. However, if all the proposed improvements would be implemented, the outcomes should support all that was assumed in this study. Also, another algorithm to perform the same kind of optimizations is proposed, as only two-third of the possibilities (ML and MP) was reviewed during this project.

Components

Volatile organic compounds that were selected as being (possibly) ancestral to all *Asimina*'s as a result of the MP- and ML-optimizations are Germacrene-D, trans-β-Caryophyllene, β-Myrcene, α-Pinene, α-Humulene, trans-β-Ocimene and Linalool. Notable is the fact that α-Pinene is known to be a toxic to fungi, bacteria and insects (Miresmailli et al., 2006; Bakkali et al. 2008) and β-Caryophyllene has a deterrent and cytotoxic effect on, for instance, beetles (Heil, 2004; Gols et al., 2008; Köllner et al., 2008). Trans-β-Ocimene, β-Myrcene and Linalool are compounds that are present in vegetative plant tissue as well (Courtois et al., 2009), and thus could be wounding volatiles. This all said, it seems that Germacrene-D is the only component that is solely interesting in the light of sexual selection. However, pollinators are very likely to sense fragrances that are emitted by the entire plant, especially if they have to be attracted over long distances. Including both floral and vegetative volatiles could be an important step (Levin et al., 2003) on the way to expanding our knowledge on the role of fragrance in the co-evolution of plants and their pollinators.

Pollination

A correlation between floral scent and pollinator attraction is probable, as was shown in many studies on pollination biology in Annonaceae (Gottsberger 1988; Silberbauer-Gottsberger, Gottsberger et al. 2003; Teichert 2008). However, as stated above, vegetative fragrances could be of great influence on pollinator attraction as well (Levin et al, 2003) and it is very probable that not all components that are emitted by plants play a role in the seduction of pollinators (Strauss and Whitthall, 2007). To investigate this co-dependency further, it is suggested to look into the importance of each component and the (relative) quantities of these components.

GC-AED

GC-AED (Gas Chromatography-Electro Antennographic Detection) can be used to distinguish which compounds are most important in attracting insects (Raguso, 2004). This analytical technique uses an insect antenna as a parallel detector for chemical components that are seperated over a gas chromatography-column. Subsets of components in a fragrance that show biological activity can thus be identified (Raguso, 2004).

By taking only the biologically active components into account during optimizations the dataset should reduce significantly, which will simplify analyses drastically. Especially if vegetative volatiles are included this could be of crucial importance. A smaller dataset means less polytomies, and less assumptions (every component is assumed to have evolved via a certain pathway). Most importantly, this approach would enable researchers to filter most of the noise out of their dataset.

Quantities

In this study data was only analysed quantitatively, as the focus lay on the presence/absence of certain VOCs, instead of their (relative) quantities. Although the qualities are indeed most interesting if one is studying the evolution of the compounds and their pathways, and selective forces are expected to act more on the composition of floral volatiles rather than overall emission rate, quantities might be of great importance in pollinator attraction (Ashman, 2009). Pollinators that have to smell the flowers from a greater distance need the flower to exude the fragrance (or at least part of the fragrance) in higher quantities. The higher a VOC's quantity, the further it can travel.

Environmental Issues

Composition and quantity of plant volatile emissions can be changed by abiotic factors (Takabayashi et al., 1994; Majetic et al., 2009) such as air temperature (i.e. Jakobsen and Olsen, 1994; Nielsen et al., 1995). Benini et al. (2012) even stated that abiotic environmental factors have a greater impact on the

composition of floral fragrances than do the genes.
Biotic environmental factors can have a huge impact too (Takabayashi et al., 1994). For instance, floral volatiles could be influenced by the developmental stage of the flower (Muhlemann et al., 2006) and of course pollinators-mediated selection is thought to play a major role in floral fragrance evolution (Suchet et al., 2011).

Strauss and Armbruster (1997) contradict this last statement. They found that natural enemies have a much larger effect on selection than pollinator-mediated selection. Annonaceae are concentrated in the tropics, which are known to have a higher insect density and diversity, compared to temperate zones (Novotny et al., 2006; Lewinsohn and Roslin, 2008). The predation pressure thus is much higher in the tropics (Janzen, 1970; Coley and Aide, 1991) and it does not seem very farfetched to suggest that especially in the tropics predation pressure could be of greater influence on the composition and quantity of floral (and other) fragrances than pollination alone.

There are so many environmental factors possibly influencing floral fragrances within the Annonaceae that it might be a good idea to gain more knowledge on the effects of these biotic and abiotic factors, before continuing with research on co-evolution between pollinators and floral fragrances. Also, since defensive volatiles are often vegetative, it is once again suggested to take the complete plant's volatile emissions into account.

Chemical pathways

Biosynthetic pathways represent extremely complex processes in nature, and floral phenotypes could change considerably as a result of one mutation within the chemical pathways that are flown through.

Mutations within a biosynthetic pathway can be the result of several events, for instance:

- A gene becomes enzymatically deactivated due to loss of gene expression (Gang, 2005).
- An existing gene evolves in such a way that it loses its original enzymatic function and gains a new one (Gang, 2005).
- Gene duplication followed by divergence (**1** -> **1**+1);
    - The original gene (the one that duplicated) retains its enzymatic function while the new gene's (the copy) enzyme has a new function (Qi et al., 2004).
    - New enzymatic functions can arise independently multiple times: convergent evolution (Gang, 2005).

The fact that there are so many ways to create a new metabolite raises questions on the veracity of the assumption that plants from different species groups make use of similar pathways. To be absolutely sure that the pathway used for the optimizations is correct, one should first gain more knowledge on the chemical transitions (and their order) within each single species.

Stepmatrices

In a perfect world the complete chemical pathway would be known for each and every chemical metabolite, plus the probability of every transition taking place. If this would be the case a stepmatrix assigning costs to every transition should not work with the number of steps alone, but should also take into account the probability (or actually, 1-probability) of each step. For instance, if the pathway would be the same as the fictional pathway that was used for the Test in this study and the probability of A changing into B would be 0.7 and the probability of B changing into C would be 0.4, the costs to change A into B would be 0.3 and the cost to change B into C 0.6. The transition of A into C then would cost 0.9 (0.3 + 0.6). This approach would be much more likely to reflect the real costs of transitions within a pathway.

Unfortunately, however, this world is no Utopia and it will take a lot of years, studies and money before the information needed to create such a stepmatrix is available. Until then, stepmatrices such as those used in this study will just have to make do.

Mesquite

Maximum Parsimony

Although Mesquite gives an error if one tries to calculate some values (such as CI) for a tree using a stepmatrix, it was assumed that the graphical representation is veracious. The parsimony scores for non-

polymorphic characters could be calculated in R (Phangorn). Unfortunately, however, this could not be achieved for characters that had more than one character state in a species. The parsimony scores for these characters were therefore taken from Mesquite. Considering that Mesquite's manual states that Mesquite 2.74 and 2.75 are able to work with polymorphies and hard polytomies, these parsimony scores should be okay, but the error makes one wonder.

Maximum Likelihood

Maximum likelihood optimizations in Mesquite 2.74 are highly influenced by the tree's topology. The expected number of changes per character across a branch is directly depending on the branch length (Lewis, 2001). In this study, the branches were not assigned lengths, but to gain more insight on this matter an extra analysis was performed, with random branch lengths assigned to the trees. There were quite some differences between the rates obtained over trees with branch lengths and over cladograms. It would be interesting to perform a ML-optimization over a tree with actual branch lengths and see what happens to the probabilities of each component being present in the tree's ancestral nodes.

Stochastic Character Mapping

An algorithm that did not pass the review during this project, but probably is very well worth studying is Stochastic character mapping (SCM), also known as Bayesian mutational mapping. SCM (Nielsen, 2002; Huelsenbeck et al., 2003) uses the likelihoods of the ancestral states and the model to simulate character histories (Bollback, 2006). It does not only reconstruct the states at the nodes of the tree, but also along the branches. The rates of character change determine the mapping of the changes on the tree (Huelsenbeck et al., 2003). Although this Bayesian approach is likely to perform better on reconstructing a character's history, it should be recognized that if the model's assumptions and priors are wrong the conclusions will be faulty (Gaut and Lewis, 1995; Lewis et al., 2005).

Conclusions

Best Model

Maximum Parsimony optimization, using the unordered approach yields realistic results.

Rejected. MP optimizations via the unordered approach yield the same results as do MP optimizations performed with an inverted stepmatrix. The inverted stepmatrix assigned higher costs (thus assuming that the probability of the occurrence of an event was lower) to the loss of a metabolite than to the creation of a new one. Biologically or chemically speaking this is not plausible at all, as a chemical component could be lost via many ways (Yuba et al., 1996; Wang et al., 1997; Barkman 2001), but can only arise via a specific series of transitions (see also Figure 9). Since the unordered approach yielded the same results as this ludicrous approach, it was decided that the unordered approach could not be any good either. Therefore the hypothesis that the unordered approach in Maximum Parsimony optimization would yield realistic results was rejected.

Maximum Parsimony optimization, using the ordered approach yields realistic results.

Rejected. The chemical pathways that were constructed for this study were branching (see Figure 6 and 7) and it was therefore not possible to put all compounds in the pathway in one linear order. It was deemed very probable for pathways to be branching webs of chemical reactions, rather than linear transition-chains. If one uses the ordered approach for MP optimization, one assigns a certain (linear) order to all components in a pathway. As this would not be representing a real-life pathway, the hypothesis that Maximum Parsimony optimization using an ordered approach would yield realistic results was rejected.

Maximum Parsimony optimization, using a symmetric stepmatrix yields realistic results.

Accepted. The symmetric stepmatrix MP-approach performed best during the simulation-phase of this study. The results of the optimization were what they should be. The symmetric stepmatrix makes sense biologically, even though losses and gains are assigned equal costs. Subsequent to this Simulation's outcome the symmetric stepmatrix was used for the optimization of real data over the actual trees. Once again, the the outcomes are seem to be realistic, as the optimizations and the pathway were congruent and the character states that were calculated to be ancestral occur in other plant-taxa as well.
Within the Mono- and Sesquiterpenes Germacrene-D and δ-Cadinene are compounds that are present in the floral fragrance of some roses (e.g. Gang, 2005), α-Pinene and β-Myrcene are present in *Juniperus* (e.g. Adams et al., 1980) and in ylang-ylang (i.e. Benini et al, 2012b). Germacrene-D is also present in Ylang-ylang, as is trans-β-Caryophyllene (i.e. Benini et al., 2012a), which in turn is also present in the

floral fragrance of *Duguetia cadaverica* (e.g. Teichert et al., 2012). The fact that these components are present in so many other species, even species from different plant families, increases the likelihood of them being ancestral. However, several Monoterpenes and Sesquiterpenes that were not found to be ancestral states occur in these other species as well (Gang, 2005; Benini et al, 2012a; Benini et al, 2012b; Teichert et al., 2012).
The Chorismate-derived components are a bit harder to evaluate, as other studies seem to focus on the Terpenes. On the other hand, a possibly ancestral state as reconstructed in this study was 'None' of the Chorismate-derived components. It could also be that this simply is the best reconstruction, something that would explain the absence of these compounds in other studies.

Although Scenarios might be very interesting from a biological perspective (the combination of components in a fragrance could have a different effect on pollinators than the individual components). It seems almost impossible to test for their aptitude. The vast amount of VOCs that are present in one scenario plays a big role in this; if a species' fragrance comprises 20 known compounds and this species is found to be ancestral, this does not necessarily mean that all 20 components, present in its fragrance, are ancestral too.

Based on the outcomes of the simulation and, in line with this, the results of the analyses with the real data and the actual tree, it was decided that Maximum Parsimony optimization with a symmetric stepmatrix does indeed yield realistic results and thus this hypothesis was accepted.

Maximum Parsimony optimization, using an asymmetric stepmatrix yields realistic results.

Rejected. The asymmetric approach yields the most derived character state as the (possible) ancestral state, a result that makes no sense in view of the biosynthetic pathway. If a chemical pathway imposes the creation of compounds C out of A to go through the creation of compound B, it is highly unlikely that the ancestral state would be C, while A and B evolved later in time. Also, it does not make sense to suggest that the loss of compounds is a sign of evolutionary advancement, since this would mean that ancestors had more genes, to code for the higher complexity in their chemical composition (Adams et al., 1980). However, it does make sense taking the criterion of parsimony into consideration, as the second option involves fewer steps, and thus this is the option chosen during an optimization that is based on maximum parsimony. The higher cost of the manufacture of a chemical compound pushes this compound back in the tree. As this is not a realistic result the hypothesis that Maximum Parsimony optimization, using an asymmetric stepmatrix yields realistic results was rejected.

Maximum Likelihood optimization, using a symmetrical model yields realistic results.

Accepted. As already discussed in the Simulation's outcomes, this model yielded realistic results, but only if the rates were not higher than 3.0. When the rates of change of the characters exceed this threshold, the characters are calculated to have a probability of 0.5 to be present in all ancestors (if the Mk1-model is used). Rates are influenced by branch lengths (Lewis, 2001), and since the trees in this study were cladograms, the outcome of the study could be something completely different if the calculations were to be performed over phylograms (i.e. including branch lengths).
For the real data, ML-optimization was performed with the components that were present in the scenarios, found to be ancestral via MP-optimization. This combined approach works pretty well, as the output is realistic and the output is reduced (not all components were present in the possibly ancestral scenarios), which increases the user-friendliness of the approach. The hypothesis that ML-optimization, using a symmetric model yields realistic results is accepted, but it is suggested to always perform this optimization over the components form the scenarios that were reconstructed as being ancestral during MP-optimization.

Maximum Likelihood optimization, using an asymmetrical model yields realistic results.

Postponed. The asymmetrical, 2 parameter Markov k-state model assigns a bias to the forward rate over the backward rate (or the other way around). This bias can be estimated by Mesquite, or it can be prescribed manually. The latter is preferred in this case, as Mesquite sometimes defines the forward rate to be higher than the backward rate, which makes as little sense (biochemically) as the inverted stepmatrix. However, due to a lack of knowledge on what the bias of the backward rate over the forward rate should be, manual input of the bias was not possible during this study and this hypothesis could not be tested.

Best Tree

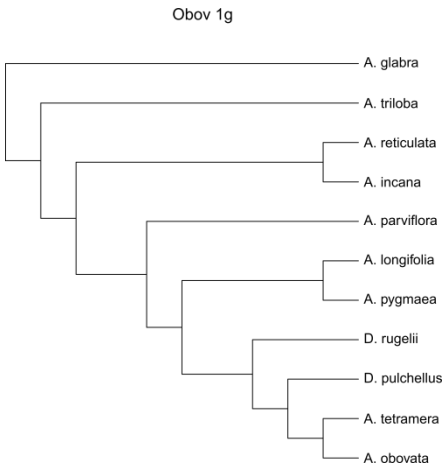


Figure 24 One of the resolved trees that were derived from the polytomious tree with the first *Asimina* obovata placement, Obov 1g, was designated as the best tree in this study.

Optimizing volatile organic compounds over trees can help resolve polytomies.

Accepted. Another objective of this study was to find out whether VOC-data could help find the best placement of *Asimina obovata* and maybe even help resolve the polytomies in trees. This was done by looking at the pScores (parsimony scores) of the optimizations of the volatile organic compounds over all trees (see also Table 13).

As for the placement of *Asimina obovata*, it seems obvious that the second placement is the least probable to be the right placement. All Obov2-trees (resolved versions of Obov2+) always had high pScores, while a pScore should be low.
It was difficult to decide whether the first or third placement should be preferred, as both the Obov1-trees and the Obov3-trees sometimes had the lowest pScores, while at the same time there could be trees with high pScores. The best trees out of these two subsets seem to be Obov 1g and Obov 3g, respectively. Because Obov 1g (Figure 24) had (one of) the lowest pScores most often, this tree was designated as the (most probably) best tree.

Although the way in which the pScores, obtained via optimizations of different characters, affects the distribution of highest and lowest pScore-sums over the different trees, there seems to be a trend. Based on this apparent trend, the hypothesis that optimizations of volatile organic compounds over trees can help resolve polytomies was accepted.

Overall

Overall there are too many weak assumptions to really attach value to the acceptance of the hypotheses above. It is important that these assumptions have better support, something that could be achieved simply by doing more research on for instance chemical pathways in plants, as these play an important part in the stepmatrix formation. Furthermore it is suggested to look into the connection between pollinators and floral, as well as vegetative VOCs. This could be done through GC-AED analyses, for instance.

If the components that play an important role in the communication of plants with pollinators are known, the optimization of scenarios (complete fragrances) becomes a lot more interesting. It is recommended that the components present in scenarios that are found to be (possibly) ancestral via MP-optimization are optimized using maximum likelihood (with a symmetrical Mk1-model).
Optimization of a chemical pathway as a polymorphic character with all components in this pathway as character states seems to work fine, but again the identification of important VOCs would increase this approach's credibility. Polymorphies will occur less frequent, making it easier for programs such as Mesquite to work with the data.

Literature Cited

Adams, R.P., Von Rudloff, E., Zanoni, T.A., Hogge, L. (1980) The Terpenoids of an Ancestral/Advanced Species Pair of *Juniperus*. *Biochemical Systematics and Ecology,* **8**, 35-37.

Agrawal, A.A., Fishbein, M. (2006) Plant defense syndromes. *Ecology,* **87**, 132-149.

Armbruster, W.S., M.E. Edwards, et al. (1994) Floral character displacement generates assemblage structure of Western-Australian triggerplants (*Stylidium*). *Ecology*, **75**, 315-329.

Armbruster, W., Muchhala, N. (2009) Associations between floral specialization and species diversity: cause, effect, or correlation? *Evolutionary Ecology*, **23,** 159-179.

Ashman, T.-L. (2009) Sniffing out patterns of sexual dimorphism in floral scent. *Functional Ecology,* **23,** 852-862.

Azuma, H., Toyota, M., Asakawa, Y., Yamaoka, R., Garcia-Franco, J.G., Dieringer, G., Thien, L.B., Kawano, S. (1997) Chemical divergence in floral scents of *Magnolia* and allied genera (Magnoliaceae). *Plant Species Biology*, **12**, 69-83.

Azuma, H., Thien, L.B., Kawano, S. (1999) Molecular phylogeny of *Magnolia* (Magnoliaceae) inferred from cpDNA seuences and evolutionary divergence of the floral scents. *Journal of Plant Research*, **3**, 291-306.

Bakkali, F., Averbeck, S., Averbeck, D., Idaomar, M. (2008) Biological effects of essential oils: a review. *Food and Chemical Toxicology*, **46,** 446-475.

Barkman, T.J., Beaman, J.H., Gage, D.A. (1997) Floral fragrance variation in *Cypripedium*: implications for evolutionary and ecological studies. *Phytochemistry*, **44**, 875-882.

Barkman, T.J. (2001) Character coding of secondary chemical variation for use in phylogenetic analyses. *Biochemical Systematics and Ecology*, **29,** 1-20.

Begerow, J., Jermann, E., Keles, T., Ranft, U., Dunemann, L. (1995) Passive sampling for volatile organic compounds (VOCs) in air at environmentally relevant concentration levels. *Fresenius' Journal of Analytical Chemistry*, **351,** 549-554.

Benini, C., Mahy, G., Bizoux, J.-P., Wathelet, J.-P., du Jardin, P., Brostaux, Y., Fauconnier, M.-L. (2012a) Comparative Chemical and Molecular Variability of *Cananga odorata* (Lam.) Hook.f. & Thomson forma *genuina* (Ylang-Ylang) in the Western Indian Ocean Islands: Implication for Valorization. *Chemistry & Biodiversity*, **9**, 1389-1402.

Benini, C., Ringuet, M., Wathelet, J.-P., Lognay, G., du Jardin, P., Fauconnier, M.-L. (2012b). Variations in the essential oils from ylang-ylang (*Cananga odorata* [Lam.] Hook f. & Thomson forma *genuina*) in the Western Indian Ocean islands. *Flavour and Fragrance Journal*, **27,** 356-366.

Bolick, M.R. (1983). A cladistic analysis of the Ambrosiinae Less. and Engelmanniinae Stuessy. *Advances in Cladistics, Vol. 2* (ed. by N.I. Platnick and V.A. Funk, pp 125-141. Columbia University Press, New York.

Bollback, J.P. (2006) SIMMAP: Stochastic character mapping of discrete traits on phylogenies. *BMC Bioinformatics*, **7,** 88.

Bosart, L.W. (1936). Perfumes and Other Odors. *Journal of the American Oil Chemists' Society,* **13**, 63-67.

Chatrou, L.W., Escribano, M.P., Virual, M.A., Maas, J.W., Richardson, J.E., Hormaza, J.I. (2009). Microsatellite flanking regions provide a new source of data for plant species-level phylogenetics. *Molecular Phylogenetics and Evolution*, **53,** 726-733.

Chatrou, L.W., Pirie, M.D., Erkens, R.H.J., Couvreur, T.L.P., Neubig, K.M., Abbott, J.R., Mols, J.B., Maas, J.W., Saunders, R.M.K., Chase, M.W. (2012). A new subfamilial and tribal classification of the

pantropical flowering plant family Annonaceae informed by molecular phylogenetics. *Botanical Journal of the Linnean Society*, **169,** 5-40.

Chess, S. K. R., Raguso, R. A., Lebuhn, G. (2008). Geographic divergence in floral morphology and scent in Linanthus dichotomus (Polemoniaceae). *American Journal of Botany*, **95,** 1652-1659.

Coley, P.D., Aide, M.T. (1991) Comparison of herbivory and plant defenses in temperate and tropical broad-leaved forests. *Plant-Animal Interactions: Evolutionary Ecology in Tropical and Temperate Regions* (ed. by P.W. Price, T.M. Lewinsohn, G.W. Fernandes, W.W. Benson), pp. 25-49. Wiley, New York.

Courtois, E.A., Paine, C.E.T., Blanieres, P.-A., Stien, D., Bessiere, J.-M., Houel, E., Baraloto, C., Chave, J. (2009) Diversity of the Volatile Organic Compounds Emitted by 55 Species of Tropical Trees: a Survey in French Guiana. *Journal of Chemical Ecology*, **35,** 1349-1362.

Croteau, R., Kutchan, T.M., Lewis, N.G. (2000) Natural Products (Secondary Metabolites). *Biochemistry & Molecular Biology of Plants* (ed. by B. Buchanan, W. Gruissem, R. Jones, pp. 1250-1318. American Society of Plant Physiologists, Rockville.

Davies, T.J., Barraclough T.G., Chase, M.W., Soltis, P.S., Soltis, D.E., Savolainen, V. (2004) Darwin's abominable mystery: insights from a supertree of the angiosperms. *Proceedings of the National Academy of Sciences of the United States of America*, **101,** 1904-1909.

Dobson, H.E.M. (1994). Floral Volatiles in Insect Biology. *Insect-Plant Interactions Volume V* (ed. by E. A. Bernays), pp. 47-82. CRC Press, Boca Raton.

Dobson, H.E.M. (2006). Relationship between Floral Fragrance Composition and Type of Pollinator. *Biology of Floral Scent* (ed. by N. Dudareva and E. Pichersky), pp 147-198. CRC Press, Boca Raton.

Dodd, M.E., Silvertown J., Chase, M.W. (1999) Phylogenetic analysis of trait evolution and species diversity variation among angiosperm families. *Evolution,* **53**, 732-744.

Doyle, J. A., Sauquet, H., Scharaschkin, T., Le Thomas, A. (2004) Phylogeny, molecular and fossil dating, and biogeographic history of Annonaceae and Myristicaceae (Magnoliales). *International Journal of Plant Sciences,* **165,** S55-S67.

Dudareva, N., Pichersky E. (2000) Biochemical and Molecular Genetic Aspects of Floral Scents*. Plant Physiology,* **122**, 627-634.

Faegri, K., Van der Pijl, L. (1966) *Principles of Pollination Ecology*. Pergamon Press, Oxford.

Farris, J.S. (1983) The logical basis of phylogenetic analysis. *Advances in Cladistics proceedings of the second meeting of the Willi Hennig Society* (ed. by N.I. Platnick and V.A. Funk), pp 1-36. Columbia University Press, New York.

Feild, T.S., Arens N.C. (2005) Form, function and environments of the early angiosperms: merging extant phylogeny and ecophysiology with fossils. *New Phytologist* **,166,** 383-408.

Felsenstein, J. (1981) Evolutionary trees from DNA sequences: a maximum likelihood approach. *Journal of Molecular Evolution*, **17**, 368-376.

Gang, D.R. (2005) Evolution of flavors and scents. *Annual Review of Plant Biology*, **56**, 301-325.

Gaut, B., Lewis, P. (1995) Succes of maximum likelihood in the four taxon case. *Molecular Biology and Evolution*, **12,** 152-162.

Gershenzon, J., Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nature Chemical Biology*, **3**, 408-414.

Givnish, T.J., Sytsma, K.J. (1997) Homoplasy in molecular vs. morphological data: the likelihood of correct phylogenetic inference. *Molecular Evolution and Adaptive Radiation* (ed. By T.J. Givnish and K.J. Sytsma), pp 55-102. Cambridge University Press, Cambridge.

Gols, R., Witjes, L.M.A., Van Loon, J.J.A., Posthumus, M.A., Dicke, M., Harvey, J.A. (2008) The effect of direct and indirect defenses in two wild brassicaceous plant species on a specialist herbivore and its gregarious endoparasitoid. *Entomologia Experimentalis et Applicata*, **128,** 99-108.

Goodrich, K.R. (2012). Floral scent in Annonaceae. *Botanical Journal of the Linnean Society*, **169,** 262-279.

Goodrich, K.R., Raguso, R.A. (2009). The olfactory component of floral display in Asimina and Deeringothamnus (Annonaceae). *New Phytologist*, **183,** 457-469.

Gottsberger, G. (1989). Comments on flower evolution and beetle pollination in the genera *Annona* and *Rollinia* (Annonaceae). *Plant Systematics and Evolution*, **167,** 189-194.

Gottsberger, G. (1999). Pollination and evolution in neotropical Annonaceae. *Plant Species Biology,* **14,** 143-152.

Harner, T., Pozo, K., Gouin, T., Macdonald, A.M., Hung, H., Cainey, J., Peters, A. (2006) Global pilot study for persistent organic pollutants (POPs) using PUF disk passive air samplers. *Environmental Pollution*, **144,** 445–452.

Heil, M. (2004) Direct defense or ecological costs: responses of herbivorous beetles to volatiles released by wild lima bean (*Phaseolus lunatus*). *Journal of Chemical Ecology,* **30,** 1289-1295.

Hickey, L.J., West R.M., Dawson, M.R., Choi, D.K. (1983) Arctic terrestrial biota: paleomagnetic evidence of age disparity with mid-Northern latitudes during the Late Cretaceous and Early Tertiary. *Science*, **221,** 1153-1156.

Hill, R.S., Scriven, L.J. (1995) The angiosperm-dominated woody vegetation of Antarctica: a review. *Review of Palaeobotany and Palynology*, **86**, 175-198.

Huelsenbeck, J.P., Nielsen, R., Bollback, J. (2003) Stochastic Mapping of Morphological Characters. *Systematic Biology*, **52**, 131-158.

Humphries, C.J., Richardson, P.M. (1980). Hennig's methods and phytochemistry. *Chemosystematics: principles and practice* (ed. by F.A. Bisby, J.G. Vaughan, C.A. Wright), pp. 353-378. Academic Press, London.

Jakobsen, H.B., Olsen, C.E. (1994) Influence of climatic factors on emission of flower volatiles in situ. *Planta*, **192,** 365-371.

Janzen, D.H. (1970) Herbivores and the number of tree species in tropical forests. *The American Naturalist*, **104,** 501-528.

Jiao, Y., Wickett, N.J., Ayyampalayam, S., Chanderbali, A.S., Landherr, L., Ralph, P.E., Tomsho, L.P., Hu, Y., Liang, H., Soltis, P.S., Soltis, D.E., Clifton, S.W., Schlarbaum, S.E., Schuster, S.C., Ma, H., Leebens-Mack, J., dePamphilis, C.W. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature*, **473,** 97-100.

Jürgens, A., Webber, A.C., Gottsberger, G. (2000) Floral scent compounds of Amazonian Annonaceae species pollinated by small beetles and thrips. *Phytochemistry*, **55**, 551-558.

Jürgens, A., Dötterl, S., Meve, U. (2006) The chemical nature of fetid floral odours in stapeliads (Apocynaceae-Asclepiadoideae-Ceropegieae). *New Phytologist*, **172**, 452-468.

Kaiser, R. (1993) *The Scent of Orchids: Olfactory and Chemical Investigations*. Elsevier, Amsterdam.

Kay, K.M., Sargent, R.D. (2009) The role of animal pollination in plant speciation: integrating ecology, geography, and genetics. *Annual Review of Ecology, Evolution, and Systematics*, **40,** 637-656.

Knudsen, J. T. (1999). Floral scent differentiation among coflowering, sympatric species of *Geonoma* (Arecaceae). *Plant Species Biology*, **14,** 137-142.

Knudsen, J.T., Eriksson, R., Gershenzon, J., Ståhl, B. (2006) Diversity and distribution of floral scent. *The Botanical Review*, **72,** 1-120.

Köllner, T.G., Held, M., Lenk, C., Hiltpold, I., Turlings, T.C.J., Gershenzon, J., Degenhardt, J. (2008) A maize (*E*)-β-caryophyllene synthase implicated in indirect defense responses against herbivores is not expressed in most American maize varieties. *Plant Cell***, 20,** 482-494.

Labandeira, C.C., Dilcher D.L., Davis, D.R., Wagner, D.L. (1994). Ninety-seven million years of angiosperm-insect association: paleobiological insights into the meaning of coevolution. *Proceedings of the National Academy of Sciences of the United States of America,* **91,** 12278-12282.

Levin, R.A., McDade, L.A., Raguso, R.A. (2003) The Systematic Utility of Floral and Vegetative Fragrance in Two Genera of Nyctaginaceae. *Systematic Biology*, **52,** 334-351.

Lewinsohn, T.M., Roslin, T. (2008) Four ways towards tropical herbivore megadiversity. *Ecology Letters,* **11**, 398-416.

Lewis, P.O. (2001) A likelihood approach to estimating phylogeny from discrete morphological character data. *Systematic Biology*, **50,** 913-925.

Lewis, P., Holder, M., Holsinger, K. (2005) Polytomies and Bayesian phylogenetic inference. *Systematic Biology*, **54,** 241-253.

Maas, P.J.M., Westra L.Y.Th., Chatrou, L.W (2003) *Duguetia. Flora Neotropica Monograph 88*. New York, The New York Botanical Garden.

Ma, L., Zeng, Y., Sun, Y., Wu, Z., Liu, M. (1998) Sudies on the aroma volatile constituents of Ylang-Ylang flowers by gas chromatography and gas chromatography/mass spectrometry. *Sepu*, **6**, 11-18.

Maddison, W.P., Maddison, D.R. (2010) Mesquite: a modular system for evolutionary analysis. Version 2.74 *http://mesquiteproject.org*.

Majetic, C.J., Raguso, R A., Ashman, T.-L. (2009) Sources of floral scent variation: can environment define floral scent phenotype? *Plant Signaling and Behavior*, **4,** 129-131.

Matich, A.J., Bunn, B.J., Comeskey, D.J., Hunt, M.B., Rowan, D.D. (2007) Chiralty and biosynthesis of lilac compounds in *Actinidia arguta* flowers. *Phytochemistry*, **68**, 1746-1751.

Millar, J.G., Sims, J.J. (1998) Preparation, cleanup and preliminary fractionation of extracts. *Methods in Chemical Ecology, Vol. 1* (ed. by J.G. Millar, K.F. Haynes), pp 1-37. Kluwer Academic Publishers, Boston.

Miresmailli, S., Bradbury, R, Isman, M.B. (2006) Comparative toxicity of *Rosmarinus officinalis* L. essential oil and blends of its major constituents against *Tetranychus urticae* Koch (Acari: Tetranychidae) on two different host plants. *Pest Management Science,* **62,** 366-371.

Miyake, T., Yafuso M. (2003) Floral scents affect reproductive success in fly-pollinated *Alocasia odora* (Araceae). *American Journal of Botany*, **90,** 370-376.

Muhlemann, J.K., Waelti, M.O., Widmer, A., Schiestl, F.P. (2006) Postpollination changes in floral odor in *Silene latifolia*: adaptive mechanisms for seed-predator avoidance? *Journal of Chemical Ecology*, **32**, 1855-1860.

Nandi, O.I., Chase, M.W., Endress, P.K. (1998) A combined cladistic analysis of angiosperms using *rbc*L and non-molecular data sets. *Annuals of the Missouri Botanical Garden*, **85**, 137-212.

Nielsen, J.K., Jakobsen, H.B., Friis, P., Hansen, K., Møller, J., Olsen, C.E. (1995) Asynchronous rhythms in the emission of volatiles from *Hesperis matronalis* flowers. *Phytochemistry*, **38**, 847-851.

Nielsen, R. (2002) Mapping mutations on phylogenies. *Systematic Biology*, **51**, 729-739.

Novotny, V., Drozd, P., Miller, S.E., Kulfan, M., Janda, M., Basset, Y., Weiblen, G.D. (2006) Why are there so many species of herbivorous insects in tropical rainforests? *Science*, **313,** 1115-1118.

Pawliszyn, J., Pawliszyn, B. and Pawliszyn, M. (1997) Solid Phase Microextraction (SPME). *The Chemical Educator*, **2,** 1-7.

Pellmyr, O., Thien, L.B. (1986) Insect Reproduction and Floral Fragrances: Keys to the Evolution of the Angiosperms? *Taxon*, **35,** 76-85.

Pennisi, E. (2005) What determines species diversity? *Science*, **309,** 90.

Pichersky, E. and Gershenzon, J. (2002) The formation and function of plant volatiles: perfumes for pollinator attraction and defense. *Current Opinion in Plant Biology*, **5,** 237-243.

Pichersky, E., Noel, J.P., Dudareva, N. (2006) Biosynthesis of Plant Volatiles: Nature's Diversity and Ingenuity. *Science*, **311,** 808-811.

Pripdeevech, P. (2011). Analysis of odor constituents of *Melodorum fruticosum* flowers by solid-phase microextraction-gas chromatography-mass spectrometry. *Chemistry of Natural Compounds,* **47**, 292-294.

Proctor, M.C.F., Yeo P., Lack, A. (1996) *The Natural History of Pollination*. Harper Collins, London.

Qi, X., Bakht, S., Leggett, M., Maxwell, C., Melton, R., Osbourn, A. (2004) A gene cluster for secondary metabolism in oat: implications for the evolution of metabolic diversity in plants. *Proceedings of the National Academy of Sciences of the United States of America*, **101**, 8233-8238.

Raguso, R.A. (2001) Floral scent, olfaction, and scent-driven foraging behavior. *Cognitive Ecology of Pollination* (ed. by L. Chittka, J. D. Thomson, pp. 83-105. Cambridge University Press, Cambridge.

Raguso, R.A. (2004) Flowers as sensory billboards: progress towards an integrated understanding of floral advertisement. *Current Opinion in Plant Biology*, **7**, 434-440.

Rainer, H. (2007) Monographic studies in the genus *Annona* L. (Annonaceae): Inclusion of the genus *Rollinia* A.St.-Hil. *Annalen des Naturhistorischen Museums in Wien. Serie B, Botanik und Zoologie. Vienna*, **108**, 191-205.

Ratnayake, R.M.C.S., Gunatilleke, I.A.U.N., Wijesundara, D.S.A., Saunders, R.M.K. (2007) Pollination Ecology and Breeding System of *Xylopia championii* (Annonaceae): Curculionid Beetle Pollination, Promoted by Floral Scents and Elevated Floral Temperatures. *International Journal of Plant Sciences*, **168,** 1255-1268.

Richardson, J.E., Chatrou, L.W., Mols, J.B., Erkens, R.H.J., Pirie, M.D. (2004) Historical biogeography of two cosmopolitan families of flowering plants: Annonaceae and Rhamnaceae. *Philosophical Transactions of the Royal Society, London B*, **359**, 1495-1508.

Richardson, P.M. (1983) Flavonoids and phylogenetic systematics. *Advances in Cladistics, Vol. 2* (ed. by N.I. Platnick and V.A. Funk), pp. 115-123. Columbia University Press, New York.

Ricklefs, R.E., Renner, S.S. (2000) Evolutionary flexibility and flowering plant familial diversity: a comment on Dodd, Silvertown, and Chase. *Evolution*, **54**, 1061-1065.

Salzmann, C.C., Nardella, A.M., Cozzolino, S., Schiestl, F.P. (2007) Variability in Floral Scent in Rewarding and Deceptive Orchids: The Signature of Pollinator-imposed Selection? *Annals of Botany*, **100**, 757-765.

Saunders, R.M.K. (2010) Floral evolution in the Annonaceae: hypotheses of homeotic mutations and functional convergence. *Biological Reviews*, **85,** 571-591.

Schiestl, F.P. (2010) The evolution of floral scent and insect chemical communication. *Ecology Letters*, **13**, 643-656.

Silberbauer-Gottsberger, I., Gottsberger, G., Webber, A.C. (2003) Morphological and Functional Flower Characteristics of New and Old World Annonaceae with Respect to Their Mode of Pollination. *Taxon*, **52,** 701-718.

Smith, J.F. (2001) High species diversity in fleshy-fruited tropical understory plants. *The American Naturalist*, **157**, 646-653.

Sneath, P.H., Sokal, R.R. (1973) *Principles of Numerical Taxonomy*. W.H. Freeman and Co, San Francisco.

Strauss, S.Y., Whittall, J.B. (2007) Non-pollinator agents of selection on floral traits. *Ecology and Evolution of Flowers* (ed. by L.D. Harder, S.C.H. Barrett), pp. 120-138.Oxford University Press, Oxford.

Suchet, C., Dormont, L., Schatz, B., Giurfa, M., Simon, V., Raynaud, C., Chave, J. (2011) Floral scent variation in two *Antirrhinum* majus subspecies influences the choice of naïve bumblebees. *Behavioral Ecology and Sociobiology*, **65**, 1015-1027.

Takabayashi, J., Dicke, M., Posthumus, M.A. (1994) Volatile herbivore-induced terpenoids in plant-mite interactions: Variation caused by biotic and abiotic factors. *Journal of Chemical Ecology*, **20**, 1329-1354.

Teichert, H. (2008) Pollination Biology of cantharophilous and melittophilous Annonaceae and Cyclanthaceae in French Guiana. *Ph.D. Thesis*. Universität Ulm.

Teichert, H., Dötterl, S., Zimma, B., Ayasse, M., Gottsberger, G. (2009) Perfume-collecting male euglossine bees as pollinators of a basal angiosperm: the case of *Unonopsis stipitata* (Annonaceae). *Plant Biology*, **11**, 29-37.

Teichert, H., Dötterl, S., Gottsberger, G. (2011) Heterodichogamy and nitidulid beetle pollination in *Anaxagorea prinoides* , an early divergent Annonaceae. *Plant Systematics and Evolution*, **291**, 25-33.

Teichert, H., Dötterl, S., Frame, D., Kirejtshuk, A., Gottsberger, G. (2012) A novel pollination mode, saprocantharophily, in *Duguetia cadaverica* (Annonaceae): A stinkhorn (Phallales) flower mimic. *Flora*, **207**, 522-529.

Van der Pijl, L. (1960) Ecological Aspects of Flower Evolution. I. Phyletic Evolution. *Evolution*, **14**, 403-416.

Van Heusden, E.C.H. (1992) *Flowers of Annonaceae: morphology, classification, and evolution*. Rijksherbarium / Hortus Botanicus, Leiden.

Van Schie, C.C.N., Haring, M.A., Schuurink, R.C. (2006) Regulation of terpenoid and benzenoid production in flowers. *Current Opinion in Plant Biology,* **9**, 203-208.

Von Frisch, K. (1919) Über den Geruchssinn der Bienen un seine blütenbiologische Bedeutung. *Zoologische Jahrbücher*, **37**, 1-238.

Waelti, M.O., Muhlemann, J.K., Widmer, A., Schiestl, F.P. (2008) Floral odour and reproductive isolation in two species of *Silene*. *Journal of Evolutionary Biology*, **21**, 111-121.

Wang, J., Dudareva, N., Bhakta, S., Raguso, R.A., Pichersky, E. (1997) Floral scent production in *Clarkia breweri* (Onagraceae) II. Localization and developmental modulation of the enzyme *S*-adenosyl-L-methionine: (iso)eugenol *O*-methyltransferase and phenylpropanoid emission. *Plant Physiology,* **114**, 213-221.

Williams, N.H., Whitten, W.M. (1999) Molecular phylogeny and floral fragrances of male euglossine bee-pollinated orchids: a study of *Stanhopea* (Orchidaceae). *Plant Species Biology*, **14**, 129-136.

Yuba, A., Yazaki, K., Tabata, M., Honda, G., Croteau, R. (1996) cDNA cloning, characterization, and functional expression of 4S-(-)-limonene synthase from *Perilla frutescens*. *Archives of Biochemistry and Biophysics*, **332**, 280-287.

Appendices

Chemical Data

	INDOLE	ETHANOL	VINEGAR	ETHYL_ACETATE	PHENYL_ACETALDEHYDE	2-PHENYLETHANOL	ACETONE	BENZYL_ALCOHOL	BENZALDEHYDE	METHYL_BENZOATE	METHYL_SALICYLATE	β-MYRCENE	*cis*-OCIMENE	*trans*-β-OCIMENE	α-TERPINENE	LIMONENE	α-PINENE	CAMPHENE	SABINENE	δ-3-CARENE	1,8-CINEOLE	LINALOOL	LILAC_ALCOHOL_ISOMER	LILAC_ALDEHYDE_ISOMER	α-FARNESENE_ISOMER	β-FARNESENE_ISOMER	α-ZINGIBIRENE_ISOMER	β-ELEMENE	α-HUMULENE	GERMACRENE_D	*trans*-β-CARYOPHYLLENE	γ-MUUROLENE	δ-CADINENE	6-METHYL-5-HEPTEN-2-ON
Asimina_parviflora	0	1	1	1	0	0	1	0	0	0	0	1	1	1	0	0	1	0	0	0	0	0	0	0	1	1	1	1	1	1	1	1	1	1
Asimina_triloba	0	1	1	1	0	0	0	0	0	0	0	1	0	0	1	0	1	1	0	1	0	1	0	0	0	1	0	0	1	1	1	1	1	0
Asimina_tetramera	1	1	0	1	0	0	0	0	0	0	1	1	1	1	0	1	1	1	0	0	0	1	0	0	0	1	0	0	1	1	1	0	0	1
Asimina_pygmaea	0	1	0	1	0	0	0	0	0	1	0	1	0	1	0	1	1	0	0	0	1	0	0	0	0	1	0	1	1	1	1	1	1	0
Asimina_reticulata	0	0	0	0	0	1	0	1	1	0	0	1	0	1	0	1	1	0	0	0	0	1	0	0	1	1	1	1	1	1	1	0	0	0
Asimina_obovata	0	0	0	1	0	0	0	0	1	0	0	1	1	1	0	1	1	1	1	0	1	1	1	1	1	0	0	1	1	1	1	1	0	0
Asimina_incana	1	0	0	0	1	1	0	1	1	1	0	1	1	1	0	0	1	0	0	0	1	1	1	1	1	1	0	1	1	1	1	1	1	0
Deeringothamnus_pulchellus	0	0	0	0	1	0	0	0	1	1	0	0	1	1	0	0	0	0	0	0	0	1	0	0	0	1	0	0	1	1	1	0	0	0
Deeringothamnus_rugelii	0	0	0	1	0	0	0	0	0	0	0	1	1	1	0	1	0	0	0	0	0	1	0	0	1	0	0	0	1	1	1	0	1	0
Annona_glabra	0	0	0	0	0	0	0	0	0	0	0	1	1	1	1	1	1	0	1	0	1	1	0	0	0	0	0	0	0	0	0	0	1	0
Asimina_longifolia	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0	1	0	0	0	0	1	1	1	0	1	0	1	0	1	1	1	0	0

Table I Binary presentation of identified chemical components. Compounds, whose names are colored red, yellow or orange are part of the Chorismate-pathway. Blue components are Monoterpenes, green components are Sesquiterpenes. The white VOCs are in the Terpenes-pathway as well.

Polymorphic Stepmatrices

Table II Polymorphic stepmatrices. The name of each stepmatrix is given in the upper left corner of every matrix. From left to right, top down: PolyChor1, PolyChor2, PolyTerp1, PolyTerp2.

PolyChor1	INDOLE	ETHANOL	ACETATE	ETHYL-ACETATE	PHENYLACETALDEHYDE	2-PHENYLETHANOL	ACETONE	BENZYLALCOHOL	BENZALDEHYDE	METHYL-BENZOATE	METHYLSALICYLATE	NONE
INDOLE	0	6	4	7	9	10	15	13	12	14	17	5
ETHANOL	6	0	2	1	15	16	21	19	18	20	23	11
ACETATE	4	2	0	3	13	14	19	17	16	18	21	9
ETHYL-ACETATE	7	1	3	0	16	17	22	20	19	21	24	12
PHENYLACETALDEHYDE	9	15	13	16	0	1	8	6	5	7	10	4
2-PHENYLETHANOL	10	16	14	17	1	0	9	7	6	8	11	5
ACETONE	15	21	19	22	8	9	0	10	9	11	14	10
BENZYLALCOHOL	13	19	17	20	6	7	10	0	1	3	6	8
BENZALDEHYDE	12	18	16	19	5	6	9	1	0	2	5	7
METHYL-BENZOATE	14	20	18	21	7	8	11	3	2	0	5	9
METHYLSALICYLATE	17	23	21	24	10	11	14	6	5	5	0	12
NONE	5	11	9	12	4	5	10	8	7	9	12	0

PolyChor2	INDOLE	ETHANOL	ACETATE	ETHYL-ACETATE	PHENYLACETALDEHYDE	2-PHENYLETHANOL	ACETONE	BENZYLALCOHOL	BENZALDEHYDE	METHYL-BENZOATE	METHYLSALICYLATE	NONE
INDOLE	0	6	4	7	9	11	15	13	12	14	17	5
ETHANOL	6	0	2	1	15	17	21	19	18	20	23	11
ACETATE	4	2	0	3	13	15	19	17	16	18	21	9
ETHYL-ACETATE	7	1	3	0	16	18	22	20	19	21	24	12
PHENYLACETALDEHYDE	9	15	13	16	0	2	8	6	5	7	10	5
2-PHENYLETHANOL	11	17	15	18	2	0	10	8	7	9	12	6
ACETONE	15	21	19	22	8	10	0	10	9	11	14	10
BENZYLALCOHOL	13	19	17	20	6	8	10	0	1	3	6	8
BENZALDEHYDE	12	18	16	19	5	7	9	1	0	2	5	7
METHYL-BENZOATE	14	20	18	21	7	9	11	3	2	0	5	9
METHYLSALICYLATE	17	23	21	24	10	12	14	6	5	5	0	12
NONE	5	11	9	12	5	6	10	8	7	9	12	0

PolyTerp1	β-MYRCENE	*cis*-β-OCIMENE	*trans*-β-OCIMENE	α-TERPINENE	LIMONENE	α-PINENE	CAMPHENE	SABINENE	δ-3-CARENE	1,8-CINEOLE	LINALOOL	LILAC_ALCOHOL_ISOMER	LILAC_ALDEHYDE_ISOME	α-FARNESENE_ISOMER	β-FARNESENE_ISOMER	α-ZINGIBIRENE_ISOMER	β-ELEMENE	α-HUMULENE	GERMACRENE_D	*trans*-β-CARYOPHYLLENE	γ-MUUROLENE	δ-CADINENE	6-METHYL-5-HEPTEN-2-O
β-MYRCENE	0	2	2	4	2	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
***cis*-OCIMENE**	2	0	2	4	2	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
***trans*-β-OCIMENE**	2	2	0	4	2	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
α-TERPINENE	4	4	4	0	4	4	4	4	4	5	4	8	7	4	4	4	4	4	4	4	4	4	13
LIMONENE	2	2	2	4	0	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
α-PINENE	2	2	2	4	2	0	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
CAMPHENE	2	2	2	4	2	2	0	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
SABINENE	2	2	2	4	2	2	2	0	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
δ-3-CARENE	2	2	2	4	2	2	2	2	0	3	2	6	5	3	3	3	3	3	3	3	3	3	12
1,8-CINEOLE	3	3	3	5	3	3	3	3	3	0	3	7	6	4	4	4	4	4	4	4	4	4	13
LINALOOL	2	2	2	4	2	2	2	2	2	2	0	4	3	3	3	3	3	3	3	3	3	3	12
LILAC_ALCOHOL_ISOMER	6	6	6	8	6	6	6	6	6	7	4	0	1	7	7	7	7	7	7	7	7	7	16
LILAC_ALDEHYDE_ISOME	5	5	5	7	5	5	5	5	5	6	3	1	0	6	6	6	6	6	6	6	6	6	15
α-FARNESENE_ISOMER	3	3	3	4	3	3	3	3	3	4	3	7	6	0	2	2	2	2	2	2	2	2	11
β-FARNESENE_ISOMER	3	3	3	4	3	3	3	3	3	4	3	7	6	2	0	2	2	2	2	2	2	2	11
α-ZINGIBIRENE_ISOMER	3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	0	2	2	2	2	2	2	11
β-ELEMENE	3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	0	2	2	2	2	2	11
α-HUMULENE	3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	0	2	2	2	2	11
GERMACRENE_D	3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	0	2	2	2	11
***trans*-β-CARYOPHYLLENE**	3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	2	0	2	2	11
γ-MUUROLENE	3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	2	2	0	2	11
δ-CADINENE	3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	2	2	2	0	11
6-METHYL-5-HEPTEN-2-ONE	12	12	12	13	12	12	12	12	12	13	12	16	15	11	11	11	11	11	11	11	11	11	0

PolyTerp2	β-MYRCENE	*cis*-β-OCIMENE	*trans*-β-OCIMENE	α-TERPINENE	LIMONENE	α-PINENE	CAMPHENE	SABINENE	δ-3-CARENE	1,8-CINEOLE	LINALOOL	LILAC_ALCOHOL_ISOMER	LILAC_ALDEHYDE_ISOME	α-FARNESENE_ISOMER	β-FARNESENE_ISOMER	α-ZINGIBIRENE_ISOMER	β-ELEMENE	α-HUMULENE	GERMACRENE_D	*trans*-β-CARYOPHYLLENE	γ-MUUROLENE	δ-CADINENE	6-METHYL-5-HEPTEN-2-O
β-MYRCENE	0	2	2	4	4	4	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
***cis*-OCIMENE**	2	0	2	4	4	4	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
***trans*-β-OCIMENE**	2	2	0	4	4	4	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
α-TERPINENE	4	4	4	0	4	4	4	4	4	4	4	8	7	4	4	4	4	4	4	4	4	4	13
LIMONENE	4	4	4	4	0	4	4	4	4	4	4	8	7	4	4	4	4	4	4	4	4	4	13
α-PINENE	4	4	4	4	4	0	4	4	4	4	4	8	7	4	4	4	4	4	4	4	4	4	13
CAMPHENE	2	2	2	4	4	4	0	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
SABINENE	2	2	2	4	4	4	2	0	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
δ-3-CARENE	2	2	2	4	4	4	2	2	0	3	2	6	5	3	3	3	3	3	3	3	3	3	12
1,8-CINEOLE	3	3	3	4	4	4	3	3	3	0	3	7	6	4	4	4	4	4	4	4	4	4	13
LINALOOL	2	2	2	4	4	4	2	2	2	2	0	4	3	3	3	3	3	3	3	3	3	3	12
LILAC_ALCOHOL_ISOMER	6	6	6	8	8	8	6	6	6	7	4	0	1	7	7	7	7	7	7	7	7	7	16
LILAC_ALDEHYDE_ISOMER	5	5	5	7	7	7	5	5	5	6	3	1	0	6	6	6	6	6	6	6	6	6	15
α-FARNESENE_ISOMER	3	3	3	4	4	4	3	3	3	4	3	7	6	0	2	2	2	2	2	2	2	2	11
β-FARNESENE_ISOMER	3	3	3	4	4	4	3	3	3	4	3	7	6	2	0	2	2	2	2	2	2	2	11
α-ZINGIBIRENE_ISOMER	3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	0	2	2	2	2	2	2	11
β-ELEMENE	3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	0	2	2	2	2	2	11
α-HUMULENE	3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	0	2	2	2	2	11
GERMACRENE_D	3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	0	2	2	2	11
***trans*-β-CARYOPHYLLENE**	3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	2	0	2	2	11
γ-MUUROLENE	3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	2	2	0	2	11
δ-CADINENE	3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	2	2	2	0	11
6-METHYL-5-HEPTEN-2-ONE	12	12	12	13	13	13	12	12	12	13	12	16	15	11	11	11	11	11	11	11	11	11	0

Scenario Stepmatrices

Table III Scenario stepmatrices. The name of the stepmatrix is given in the upper left corner of each matrix. From left to right, top down: ScenChor1, ScenChor2, ScenMono1, ScenMono2, ScenSesqui1, ScenTerp1, ScenTerp2.

ScenChor1	A. parviflora	A. triloba	A. tetramera	A. pygmaea	A. reticulata	A. obovata	A. incana	D. pulchellus	D. rugelii	A. glabra	A. longifolia
A. parviflora	0	10	12	15	24	9	14	24	6	18	18
A. triloba	10	0	12	9	22	7	19	22	0	12	12
A. tetramera	12	12	0	5	20	5	15	18	12	24	24
A. pygmaea	15	9	5	0	17	2	10	13	9	24	24
A. reticulata	24	22	20	17	0	15	7	4	22	10	10
A. obovata	9	7	5	2	15	0	12	15	7	19	19
A. incana	14	19	15	10	7	12	0	7	19	17	17
D. pulchellus	24	22	18	13	4	15	7	0	22	10	10
D. rugelii	6	0	12	9	22	7	19	22	0	12	12
A. glabra	18	12	24	24	10	19	17	10	12	0	0
A. longifolia	18	12	24	24	10	19	17	10	12	0	0

ScenChor2	A. parviflora	A. triloba	A. tetramera	A. pygmaea	A. reticulata	A. obovata	A. incana	D. pulchellus	D. rugelii	A. glabra	A. longifolia
A. parviflora	0	10	12	15	25	9	15	25	6	18	18
A. triloba	10	0	12	9	23	7	20	23	0	12	12
A. tetramera	12	12	0	5	21	5	16	19	12	24	24
A. pygmaea	15	9	5	0	18	2	11	14	9	24	24
A. reticulata	25	23	21	18	0	16	7	4	23	11	11
A. obovata	9	7	5	2	16	0	13	16	7	19	19
A. incana	15	20	16	11	7	13	0	7	20	18	18
D. pulchellus	25	23	19	14	4	16	7	0	23	11	11
D. rugelii	6	0	12	9	23	7	20	23	0	12	12
A. glabra	18	12	24	24	11	19	18	11	12	0	0
A. longifolia	18	12	24	24	11	19	18	11	12	0	0

ScenMono1	A. parviflora	A. triloba	A. tetramera	A. pygmaea	A. reticulata	A. obovata	A. incana	D. pulchellus	D. rugelii	A. glabra	A. longifolia
A. parviflora	0	6	3	4	3	11	7	3	4	7	6
A. triloba	6	0	6	9	6	13	12	8	8	8	9
A. tetramera	3	6	0	5	2	7	8	4	2	6	7
A. pygmaea	4	9	5	0	3	8	7	7	5	5	10
A. reticulata	3	6	2	3	0	9	8	4	2	6	7
A. obovata	11	13	7	8	9	0	3	11	9	7	6
A. incana	7	12	8	7	8	3	0	8	8	8	3
D. pulchellus	3	8	4	7	4	11	8	0	2	8	7
D. rugelii	4	8	2	5	2	9	8	2	0	6	7
A. glabra	7	8	6	5	6	7	8	8	6	0	11
A. longifolia	6	9	7	10	7	6	3	7	7	11	0

ScenMono2	A. parviflora	A. triloba	A. tetramera	A. pygmaea	A. reticulata	A. obovata	A. incana	D. pulchellus	D. rugelii	A. glabra	A. longifolia
A. parviflora	0	10	12	15	25	9	15	25	6	18	18
A. triloba	10	0	12	9	23	7	20	23	0	12	12
A. tetramera	12	12	0	5	21	5	16	19	12	24	24
A. pygmaea	15	9	5	0	18	2	11	14	9	24	24
A. reticulata	25	23	21	18	0	16	7	4	23	11	11
A. obovata	9	7	5	2	16	0	13	16	7	19	19
A. incana	15	20	16	11	7	13	0	7	20	18	18
D. pulchellus	25	23	19	14	4	16	7	0	23	11	11
D. rugelii	6	0	12	9	23	7	20	23	0	12	12
A. glabra	18	12	24	24	11	19	18	11	12	0	0
A. longifolia	18	12	24	24	11	19	18	11	12	0	0

ScenSesqui1	A. parviflora	A. triloba	A. tetramera	A. pygmaea	A. reticulata	A. obovata	A. incana	D. pulchellus	D. rugelii	A. glabra	A. longifolia
A. parviflora	0	13	5	12	12	13	11	15	14	18	14
A. triloba	13	0	12	1	5	2	2	2	3	5	3
A. tetramera	5	12	0	13	13	13	14	10	13	15	13
A. pygmaea	12	1	13	0	4	3	1	3	4	6	2
A. reticulata	12	5	13	4	0	3	3	3	4	8	4
A. obovata	13	2	13	3	3	0	2	4	3	7	3
A. incana	11	2	14	1	3	2	0	4	3	7	3
D. pulchellus	15	2	10	3	3	4	4	0	3	5	3
D. rugelii	14	3	13	4	4	3	3	3	0	4	6
A. glabra	18	5	15	6	8	7	7	5	4	0	6
A. longifolia	14	3	13	2	4	3	3	3	6	6	0

ScenTerp1	A. parviflora	A. triloba	A. tetramera	A. pygmaea	A. reticulata	A. obovata	A. incana	D. pulchellus	D. rugelii	A. glabra	A. longifolia
A. parviflora	0	19	8	16	15	24	18	18	18	25	20
A. triloba	19	0	18	10	11	15	14	10	11	13	12
A. tetramera	8	18	0	18	15	20	22	14	15	21	20
A. pygmaea	16	10	18	0	7	11	8	10	9	11	12
A. reticulata	15	11	15	7	0	12	11	7	6	14	11
A. obovata	24	15	20	11	12	0	5	15	12	14	9
A. incana	18	14	22	8	11	5	0	12	11	15	6
D. pulchellus	18	10	14	10	7	15	12	0	5	13	10
D. rugelii	18	11	15	9	6	12	11	5	0	10	13
A. glabra	25	13	21	11	14	14	15	13	10	0	17
A. longifolia	20	12	20	12	11	9	6	10	13	17	0

ScenTerp2	A. parviflora	A. triloba	A. tetramera	A. pygmaea	A. reticulata	A. obovata	A. incana	D. pulchellus	D. rugelii	A. glabra	A. longifolia
A. parviflora	0	19	10	17	16	25	18	19	19	26	20
A. triloba	19	0	19	11	12	16	14	11	13	14	12
A. tetramera	10	19	0	18	15	20	23	16	16	21	21
A. pygmaea	17	11	18	0	7	11	9	11	10	11	13
A. reticulata	16	12	15	7	0	12	12	9	7	14	12
A. obovata	25	16	20	11	12	0	6	17	13	14	10
A. incana	18	14	23	9	12	6	0	13	13	16	6
D. pulchellus	19	11	16	11	9	17	13	0	6	15	11
D. rugelii	19	13	16	10	7	13	13	6	0	11	15
A. glabra	26	14	21	11	14	14	16	15	11	0	18
A. longifolia	20	12	21	13	12	10	6	11	15	18	0

R-script Phangorn

```
setwd("Drive\\Working_Directory")

###     Loading libraries       ###
library(ape)
library(phangorn)


###     Trees                   ###
Obov                    <-              read.nexus("Trees.nex")

#   vectors for names
ScenNames               <-              c("Aparv", "Atril", "Atetr", "Apygm", "Areti", "Aobov", "Ainca",
                                        "Dpulc", "Druge", "Aglab", "Along")
SpecNames               <-              attr(Obov, "TipLabel")

###     Data per Character      ###
#   Monoterpenes
vMono                   <-              ScenNames
Mono                    <-              matrix(vMono)
row.names(Mono)         <-              SpecNames
#   Sesquiterpenes
vSesqui                 <-              ScenNames
Sesqui                  <-              matrix(vSesqui)
row.names(Sesqui)       <-              SpecNames
#   All terpenes
vTerp                   <-              ScenNames
Terp                    <-              matrix(vTerp)
row.names(Terp)         <-              SpecNames
#   Chorismate
vChor                   <-              ScenNames
Chor                    <-              matrix(vChor)
row.names(Chor)         <-              SpecNames

#   changing class from matrix to phyDat
phyMono                 <-              phyDat(Mono, type = "USER", levels = ScenNames)
phySesqui               <-              phyDat(Sesqui, type = "USER", levels = ScenNames)
phyTerp                 <-              phyDat(Terp, type = "USER", levels = ScenNames)
phyChor                 <-              phyDat(Chor, type = "USER", levels = ScenNames)

###     Stepmatrices            ###
#   Monoterpenes
StepMono1vector         <-              (#ScenMono1 as a vector)
StepMono1               <-              matrix(StepMono1vector, byrow=T, nrow=11)
row.names(StepMono1)    <-              ScenNames
colnames(StepMono1)     <-              ScenNames
StepMono1

StepMono2vector         <-              c(#ScenMono2 as a vector)
StepMono2               <-              matrix(StepMono2vector, byrow=T, nrow=11)
row.names(StepMono2)    <-              ScenNames
colnames(StepMono2)     <-              ScenNames
StepMono2

#   Sesquiterpenes
StepSesqui1vector       <-              c(#ScenSesqui1 as a vector)
StepSesqui1             <-              matrix(StepSesqui1vector, byrow=T, nrow=11)
row.names(StepSesqui1)  <-              ScenNames
colnames(StepSesqui1)   <-              ScenNames
StepSesqui1

#   Terpenes
StepTerp1vector         <-              c(#ScenTerp1 as a vector)
StepTerp1               <-              matrix(StepTerp1vector, byrow=T, nrow=11)
row.names(StepTerp1)    <-              ScenNames
colnames(StepTerp1)     <-              ScenNames
StepTerp1
```

```
StepTerp2vector         <-          c(#ScenTerp2 as a vector)
StepTerp2               <-          matrix(StepTerp2vector, byrow=T, nrow=11)
row.names(StepTerp2)    <-          ScenNames
colnames(StepTerp2)     <-          ScenNames
StepTerp2

#   Chorismate
StepChor1vector         <-          c(#ScenChor1 as a vector)
StepChor1               <-          matrix(StepChor1vector, byrow=T, nrow=11)
row.names(StepChor1)    <-          ScenNames
colnames(StepChor1)     <-          ScenNames
StepChor1

StepChor2vector         <-          c(#ScenChor2 as a vector)
StepChor2               <-          matrix(StepChor2vector, byrow=T, nrow=11)
row.names(StepChor2)    <-          ScenNames
colnames(StepChor2)     <-          ScenNames
StepChor2


###     Optimization            ###
Treename                <-          gsub("\tTREE ", "",  names(Obov))

MPObov_Mono1            <-          sankoff(Obov, phyMono, cost=StepMono1, site="site")
MPObov_Mono2            <-          sankoff(Obov, phyMono, cost=StepMono2, site="site")
MPObov_Sesqui1          <-          sankoff(Obov, phySesqui, cost=StepSesqui1, site="site")
MPObov_Terp1            <-          sankoff(Obov, phyTerp, cost=StepTerp1, site="site")
MPObov_Terp2            <-          sankoff(Obov, phyTerp, cost=StepTerp2, site="site")
MPObov_Chor1            <-          sankoff(Obov, phyChor, cost=StepChor1, site="site")
MPObov_Chor2            <-          sankoff(Obov, phyChor, cost=StepChor2, site="site")

# Summarize Results
Obov_results            <-          data.frame(tree = treename, Obov_Mono_1 = MPObov_Mono1,
                                    Obov_Mono_2 = MPObov_Mono2, Obov_Sesqui =
                                    MPObov_Sesqui1, Obov_Terp_1 = MPObov_Terp1,
                                    Obov_Terp_2 = MPObov_Terp2, Obov_Chor_1 =
                                    MPObov_Chor1, Obov_Chor_2 = MPObov_Chor2)

###     Write Results           ###
write.table(Obov_results, "Drive\\Working_Directory\\Obov_results.txt", row.names = FALSE)
```

Mesquite, Nexus-file

```
#NEXUS

BEGIN TAXA;
TITLE Taxa;
DIMENSIONS NTAX=11;
TAXLABELS	A._parviflora A._triloba A._tetramera A._pygmaea A._reticulata A._obovata A._incana
D._pulchellus D._rugelii A._glabra A._longifolia;
END;


BEGIN CHARACTERS;
TITLE  'Bin_All';
DIMENSIONS  NCHAR=34;
FORMAT DATATYPE = STANDARD GAP = - MISSING = ? SYMBOLS = "  0 1";
CHARSTATELABELS 1 'INDOLE^n', 2 'ETHANOL^n', 3 'VINEGAR^n', 4 'ETHYL_ACETATE^n', 5
'PHENYL_ACETALDEHYDE ^n', 6 '2-PHENYLETHANOL^n', 7 'ACETONE^n', 8 'BENZYL_ALCOHOL^n', 9
'BENZALDEHYDE^n', 10 'METHYL_BENZOATE^n', 11 'METHYL_SALICYLATE^n', 12 '?-MYRCENE^n', 13
'cis-OCIMENE^n', 14 'trans-?-OCIMENE^n', 15 '?-TERPINENE^n', 16 'LIMONENE^n', 17 '?-PINENE^n',
18 'CAMPHENE ^n', 19 'SABINENE ^n', 20 '?-3-CARENE^n', 21 '1,8-CINEOLE^n', 22 'LINALOOL ^n', 23
'LILAC_ALCOHOL_ISOMER^n', 24 'LILAC_ALDEHYDE_ISOMER^n', 25 '?-FARNESENE_ISOMER^n', 26 '?-
FARNESENE_ISOMER^n', 27 '?-ZINGIBIRENE_ISOMER^n', 28 '?-ELEMENE  ^n', 29 '?-HUMULENE  ^n',
30 'GERMACRENE_D  ^n', 31 'trans-?-CARYOPHYLLENE^n', 32 '?-MUUROLENE^n', 33 '?-CADINENE^n',
34 '6-METHYL-5-HEPTEN-2-ONE^n';
	MATRIX
	A._parviflora	0111001000011100100000001111111111
	A._triloba	0111000000010010110101000100111110
	A._tetramera	1101000000111101110001000100111001
	A._pygmaea	0101000001010101100010000101111110
	A._reticulata	0000010110010101100001001111111000
	A._obovata	0001000010011101111011111001111100
	A._incana	1000110111011100100011111101111110
	D._pulchellus	0000100011001100000001000100111000
	D._rugelii	0001000000011101000001001000111010
	A._glabra	0000000000011111101011000000000010
	A._longifolia	0000000000011000100001110101011100;
END;

BEGIN CHARACTERS;
TITLE  'Chorismate_all';
DIMENSIONS  NCHAR=3;
FORMAT DATATYPE = STANDARD GAP = - MISSING = ? SYMBOLS = "  0 1 2 3 4 5 6 7 8 9 A B";
CHARSTATELABELS 1 'Chorismate_pathway' /  'INDOLE^n' 'ETHANOL^n' ACETATE 'ETHYL_ACETATE^n'
'PHENYL_ACETALDEHYDE ^n' '2-PHENYLETHANOL^n' 'ACETONE^n' 'BENZYL_ALCOHOL^n'
'BENZALDEHYDE^n' 'METHYL_BENZOATE^n' 'METHYL_SALICYLATE^n' none, 2 'trans-Cinnamate' /
'INDOLE^n' 'ETHANOL^n' ACETATE 'ETHYL_ACETATE^n' 'PHENYL_ACETALDEHYDE ^n' '2-
PHENYLETHANOL^n' 'ACETONE^n' 'BENZYL_ALCOHOL^n' 'BENZALDEHYDE^n' 'METHYL_BENZOATE^n'
'METHYL_SALICYLATE^n' none, 3 Indole /  'INDOLE^n' 'ETHANOL^n' ACETATE 'ETHYL_ACETATE^n'
'PHENYL_ACETALDEHYDE ^n' '2-PHENYLETHANOL^n' 'ACETONE^n' 'BENZYL_ALCOHOL^n'
'BENZALDEHYDE^n' 'METHYL_BENZOATE^n' 'METHYL_SALICYLATE^n' none;
	MATRIX
	A._parviflora	(1 2 3 6)6(1 2 3)
	A._triloba	(1 2 3)B(1 2 3)
	A._tetramera	(0 1 3 A)A(0 1 3)
	A._pygmaea	(1 3 9)9(1 3)
	A._reticulata	(5 7 8)(7 8)B
	A._obovata	(3 8)83
	A._incana	(0 4 5 7 8 9)(7 8 9)0
	D._pulchellus	(4 8 9)(8 9)B
	D._rugelii	3B3
	A._glabra	BBB
	A._longifolia	BBB;
END;

BEGIN CHARACTERS;
TITLE  'Terpenes_all';
```

```
DIMENSIONS  NCHAR=2;
FORMAT DATATYPE = STANDARD GAP = - MISSING = ? SYMBOLS = "  0 1 2 3 4 5 6 7 8 9 A B C D E F G
H J K M N P";
CHARSTATELABELS 1 Monoterpenes /  'b-MYRCENE^n' 'cis-OCIMENE^n' 'trans-?-OCIMENE^n' 'a-
TERPINENE^n' 'LIMONENE^n' 'a-PINENE^n' 'CAMPHENE ^n' 'SABINENE ^n' 'd-3-CARENE^n' '1,8-
CINEOLE^n' 'LINALOOL ^n' 'LILAC_ALCOHOL_ISOMER^n' 'LILAC_ALDEHYDE_ISOMER^n' 'a-
FARNESENE_ISOMER^n' 'b-FARNESENE_ISOMER^n' 'a-ZINGIBIRENE_ISOMER^n' 'b-ELEMENE  ^n' 'a-
HUMULENE  ^n' 'GERMACRENE_D  ^n' 'trans-b-CARYOPHYLLENE^n' 'g-MUUROLENE^n' 'd-CADINENE^n'
'3-BUTEN-2-ONE^n' '6-METHYL-5-HEPTEN-2-ONE^n', 2 Sesquiterpenes /  'b-MYRCENE^n' 'cis-
OCIMENE^n' 'trans-?-OCIMENE^n' 'a-TERPINENE^n' 'LIMONENE^n' 'a-PINENE^n' 'CAMPHENE ^n'
'SABINENE ^n' 'd-3-CARENE^n' '1,8-CINEOLE^n' 'LINALOOL ^n' 'LILAC_ALCOHOL_ISOMER^n'
'LILAC_ALDEHYDE_ISOMER^n' 'a-FARNESENE_ISOMER^n' 'b-FARNESENE_ISOMER^n' 'a-
ZINGIBIRENE_ISOMER^n' 'b-ELEMENE  ^n' 'a-HUMULENE  ^n' 'GERMACRENE_D  ^n' 'trans-b-
CARYOPHYLLENE^n' 'g-MUUROLENE^n' 'd-CADINENE^n' '6-METHYL-5-HEPTEN-2-ONE^n' ;
        MATRIX
        A._parviflora     (0 1 2 5)(D E F G H J K M N P)
        A._triloba        (0 3 5 6 8 A)(E H J K M N)
        A._tetramera      (0 1 2 4 5 6 A)(E H J K P)
        A._pygmaea        (0 2 4 5 9)(E G H J K M N)
        A._reticulata     (0 2 4 5 A)(D E F G H J K)
        A._obovata        (0 1 2 4 5 6 7 9 A B C)(D G H J K M)
        A._incana         (0 1 2 5 9 A B C)(D E G H J K M N)
        D._pulchellus     (1 2 A)(E H J K)
        D._rugelii        (0 1 2 4 A)(D H J K N)
        A._glabra         (0 1 2 3 4 5 7 9 A)N
        A._longifolia     (0 1 5 A B C)(E G J K M);
END;
BEGIN CHARACTERS;
TITLE  'Scenario_Terpenes';
DIMENSIONS  NCHAR=3;
FORMAT DATATYPE = STANDARD GAP = - MISSING = ? SYMBOLS = "  0 1 2 3 4 5 6 7 8 9 A";
CHARSTATELABELS 1 ScenMono /  Aparv Atril Atetr Apygm Areti Aobov Ainca Dpulc Druge Aglab Along, 2
ScenSesqui /  Aparv Atril Atetr Apygm Areti Aobov Ainca Dpulc Druge Aglab Along, 3 ScenTerp /  Aparv
Atril Atetr Apygm Areti Aobov Ainca Dpulc Druge Aglab Along ;
        MATRIX
        A._parviflora     000
        A._triloba        111
        A._tetramera      222
        A._pygmaea        333
        A._reticulata     444
        A._obovata        555
        A._incana         666
        D._pulchellus     777
        D._rugelii        888
        A._glabra         999
        A._longifolia     AAA;
END;

BEGIN CHARACTERS;
TITLE  'Scenario_Chorismate';
DIMENSIONS  NCHAR=1;
FORMAT DATATYPE = STANDARD GAP = - MISSING = ? SYMBOLS = "  0 1 2 3 4 5 6 7 8 9";
CHARSTATELABELS 1 ScenChor /  Aparv Atril Atetr Apygm Areti Aobov Ainca Dpulc Druge Aglab Along ;
        MATRIX
        A._parviflora     0
        A._triloba        1
        A._tetramera      2
        A._pygmaea        3
        A._reticulata     4
        A._obovata        5
        A._incana         6
        D._pulchellus     7
        D._rugelii        8
        A._glabra         9
        A._longifolia     9;
END;
```

```
BEGIN TREES;
	Title 'Trees from "Atriloba.nex"';
	LINK Taxa = Taxa;
	TRANSLATE
		1	A._parviflora,
		2	A._triloba,
		3	A._tetramera,
		4	A._pygmaea,
		5	A._reticulata,
		6	 A._obovata,
		7	A._incana,
		8	D._pulchellus,
		9	D._rugelii,
		10	A._glabra,
		11	A._longifolia;
	TREE Obov1+	= ((((((4,11),(6,3),9,8),1),(7,5)),2),10);
	TREE Obov1a	= (((((((4,11),(6,3)),(9,8)),1),(7,5)),2),10);
	TREE Obov1b	= (((((((4,11),9),((6,3),8)),1),(7,5)),2),10);
	TREE Obov1c	= (((((((4,11),8),((6,3),9)),1),(7,5)),2),10);
	TREE Obov1d	 = (((((((4,11),(6,3)),9),8),1),(7,5)),2),10);
	TREE Obov1e	= (((((((4,11),(6,3)),8),9),1),(7,5)),2),10);
	TREE Obov1f	= (((((((6,3),9),8),(4,11)),1),(7,5)),2),10);
	TREE Obov1g	= (((((((6,3),8),9),(4,11)),1),(7,5)),2),10);
	TREE Obov1h	= (((((((9,8),(4,11)),(6,3)),1),(7,5)),2),10);
	TREE Obov1i	= (((((((9,8),(6,3)),(4,11)),1),(7,5)),2),10);
	TREE Obov2+	= (((((((4,11),3,9,8),6),1),(7,5)),2),10);
	TREE Obov2a	= (((((((4,11),3),(9,8)),6),1),(7,5)),2),10);
	TREE Obov2b	= (((((((4,11),9),(3,8)),6),1),(7,5)),2),10);
	TREE Obov2c	= (((((((4,11),8),(3,9)),6),1),(7,5)),2),10);
	TREE Obov2d	= (((((((((4,11),3),9),8),6),1),(7,5)),2),10);
	TREE Obov2e	= (((((((((4,11),3),8),9),6),1),(7,5)),2),10);
	TREE Obov2f	= (((((((((4,11),9),8),3),6),1),(7,5)),2),10);
	TREE Obov2g	= (((((((((4,11),9),3),8),6),1),(7,5)),2),10);
	TREE Obov2h	= (((((((((4,11),8),3),9),6),1),(7,5)),2),10);
	TREE Obov2i	= (((((((((4,11),8),9),3),6),1),(7,5)),2),10);
	TREE Obov3+	= ((((((4,11),3,9,8),1),((7,6),5)),2),10);
	TREE Obov3a	= (((((((4,11),3),(9,8)),1),((7,6),5)),2),10);
	TREE Obov3b	= (((((((4,11),9),(8,3)),1),((7,6),5)),2),10);
	TREE Obov3c	= (((((((4,11),8),(3,9)),1),((7,6),5)),2),10);
	TREE Obov3d	= (((((((((4,11),8),3),9),1),((7,6),5)),2),10);
	TREE Obov3e	= (((((((((4,11),8),9),3),1),((7,6),5)),2),10);
	TREE Obov3f	= (((((((((4,11),9),3),8),1),((7,6),5)),2),10);
	TREE Obov3g	= (((((((((4,11),9),8),3),1),((7,6),5)),2),10);
	TREE Obov3h	= (((((((((4,11),3),8),9),1),((7,6),5)),2),10);
	TREE Obov3i	= (((((((((4,11),3),9),8),1),((7,6),5)),2),10);
END;
```

BEGIN ASSUMPTIONS;

USERTYPE Terpenes1 (Stepmatrix) =

23

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H	J	K	M	N	P
0	2	2	4	2	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	0	2	4	2	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	0	4	2	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
4	4	4	0	4	4	4	4	4	5	4	8	7	4	4	4	4	4	4	4	4	4	13
2	2	2	4	0	2	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	2	4	2	0	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	2	4	2	2	0	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	2	4	2	2	2	0	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	2	4	2	2	2	2	0	3	2	6	5	3	3	3	3	3	3	3	3	3	12
3	3	3	5	3	3	3	3	3	0	3	7	6	4	4	4	4	4	4	4	4	4	13
2	2	2	4	2	2	2	2	2	2	0	4	3	3	3	3	3	3	3	3	3	3	12
6	6	6	8	6	6	6	6	6	7	4	0	1	7	7	7	7	7	7	7	7	7	16
5	5	5	7	5	5	5	5	5	6	3	1	0	6	6	6	6	6	6	6	6	6	15
3	3	3	4	3	3	3	3	3	4	3	7	6	0	2	2	2	2	2	2	2	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	0	2	2	2	2	2	2	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	0	2	2	2	2	2	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	0	2	2	2	2	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	0	2	2	2	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	0	2	2	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	2	0	2	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	2	2	0	2	11
3	3	3	4	3	3	3	3	3	4	3	7	6	2	2	2	2	2	2	2	2	0	11
12	12	12	13	12	12	12	12	12	13	12	16	15	11	11	11	11	11	11	11	11	11	0

;

USERTYPE Terpenes2 (Stepmatrix) =

23

0	1	2	3	4	5	6	7	8	9	A	B	C	D	E	F	G	H	J	K	M	N	P
0	2	2	4	4	4	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	0	2	4	4	4	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	0	4	4	4	2	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
4	4	4	0	4	4	4	4	4	4	4	8	7	4	4	4	4	4	4	4	4	4	13
4	4	4	4	0	4	4	4	4	4	4	8	7	4	4	4	4	4	4	4	4	4	13
4	4	4	4	4	0	4	4	4	4	4	8	7	4	4	4	4	4	4	4	4	4	13
2	2	2	4	4	4	0	2	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	2	4	4	4	2	0	2	3	2	6	5	3	3	3	3	3	3	3	3	3	12
2	2	2	4	4	4	2	2	0	3	2	6	5	3	3	3	3	3	3	3	3	3	12
3	3	3	4	4	4	3	3	3	0	3	7	6	4	4	4	4	4	4	4	4	4	13
2	2	2	4	4	4	2	2	2	2	0	4	3	3	3	3	3	3	3	3	3	3	12
6	6	6	8	8	8	6	6	6	7	4	0	1	7	7	7	7	7	7	7	7	7	16
5	5	5	7	7	7	5	5	5	6	3	1	0	6	6	6	6	6	6	6	6	6	15
3	3	3	4	4	4	3	3	3	4	3	7	6	0	2	2	2	2	2	2	2	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	0	2	2	2	2	2	2	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	0	2	2	2	2	2	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	0	2	2	2	2	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	0	2	2	2	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	0	2	2	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	2	0	2	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	2	2	0	2	11
3	3	3	4	4	4	3	3	3	4	3	7	6	2	2	2	2	2	2	2	2	0	11
12	12	12	13	13	13	12	12	12	13	12	16	15	11	11	11	11	11	11	11	11	11	0

;

```
	USERTYPE Chorismate1 (Stepmatrix) =
		11
	0	1	2	3	4	5	6	7	8	9	A
	0	6	4	7	9	10	15	13	12	14	17
	6	0	2	1	15	16	21	19	18	20	23
	4	2	0	3	13	14	19	17	16	18	21
	7	1	3	0	16	17	22	20	19	21	24
	9	15	13	16	0	1	8	6	5	7	10
	10	16	14	17	1	0	9	7	6	8	11
	15	21	19	22	8	9	0	10	9	11	14
	13	19	17	20	6	7	10	0	1	3	6
	12	18	16	19	5	6	9	1	0	2	5
	14	20	18	21	7	8	11	3	2	0	5
	17	23	21	24	10	11	14	6	5	5	0
;
	USERTYPE Chorismate2 (Stepmatrix) =
		11
	0	1	2	3	4	5	6	7	8	9	A
	0	6	4	7	9	11	15	13	12	14	17
	6	0	2	1	15	17	21	19	18	20	23
	4	2	0	3	13	15	19	17	16	18	21
	7	1	3	0	16	18	22	20	19	21	24
	9	15	13	16	0	2	8	6	5	7	10
	11	17	15	18	2	0	10	8	7	9	12
	15	21	19	22	8	10	0	10	9	11	14
	13	19	17	20	6	8	10	0	1	3	6
	12	18	16	19	5	7	9	1	0	2	5
	14	20	18	21	7	9	11	3	2	0	5
	17	23	21	24	10	12	14	6	5	5	0
;
	USERTYPE ScenChor1 (Stepmatrix) =
		10
	0	1	2	3	4	5	6	7	8	9
	0	10	12	15	24	9	14	24	6	18
	10	0	12	9	22	7	19	22	0	12
	12	12	0	5	20	5	15	18	12	24
	15	9	5	0	17	2	10	13	9	24
	24	22	20	17	0	15	7	4	22	10
	9	7	5	2	15	0	12	15	7	19
	14	19	15	10	7	12	0	7	19	17
	24	22	18	13	4	15	7	0	22	10
	6	0	12	9	22	7	19	22	0	12
	18	12	24	24	10	19	17	10	12	0
;
	USERTYPE ScenChor2 (Stepmatrix) =
		10
	0	1	2	3	4	5	6	7	8	9
	0	10	12	15	25	9	15	25	6	18
	10	0	12	9	23	7	20	23	0	12
	12	12	0	5	21	5	16	19	12	24
	15	9	5	0	18	2	11	14	9	24
	25	23	21	18	0	16	7	4	23	11
	9	7	5	2	16	0	13	16	7	19
	15	20	16	11	7	13	0	7	20	18
	25	23	19	14	4	16	7	0	23	11
	6	0	12	9	23	7	20	23	0	12
	18	12	24	24	11	19	18	11	12	0
;
	USERTYPE ScenMono1 (Stepmatrix) =
		11
	0	1	2	3	4	5	6	7	8	9	A
	0	6	3	4	3	11	7	3	4	7	6
	6	0	6	9	6	13	12	8	8	8	9
	3	6	0	5	2	7	8	4	2	6	7
	4	9	5	0	3	8	7	7	5	5	10
	3	6	2	3	0	9	8	4	2	6	7
	11	13	7	8	9	0	3	11	9	7	6
```

```
        7       12      8       7       8       3       0       8       8       8       3
        3       8       4       7       4       11      8       0       2       8       7
        4       8       2       5       2       9       8       2       0       6       7
        7       8       6       5       6       7       8       8       6       0       11
        6       9       7       10      7       6       3       7       7       11      0
;
        USERTYPE ScenMono2 (Stepmatrix) =
                 11
        0       1       2       3       4       5       6       7       8       9       A
        0       6       5       5       4       12      7       4       5       8       6
        6       0       7       10      7       14      12      9       10      9       9
        5       7       0       5       2       7       9       6       3       6       8
        5       10      5       0       3       8       8       8       6       5       11
        4       7       2       3       0       9       9       6       3       6       8
        12      14      7       8       9       0       4       13      10      7       7
        7       12      9       8       9       4       0       9       10      9       3
        4       9       6       8       6       13      9       0       3       10      8
        5       10      3       6       3       10      10      3       0       7       9
        8       9       6       5       6       7       9       10      7       0       12
        6       9       8       11      8       7       3       8       9       12      0
;
        USERTYPE ScenSesqui (Stepmatrix) =
                 11
        0       1       2       3       4       5       6       7       8       9       A
        0       13      5       12      12      13      11      15      14      18      14
        13      0       12      1       5       2       2       2       3       5       3
        5       12      0       13      13      13      14      10      13      15      13
        12      1       13      0       4       3       1       3       4       6       2
        12      5       13      4       0       3       3       3       4       8       4
        13      2       13      3       3       0       2       4       3       7       3
        11      2       14      1       3       2       0       4       3       7       3
        15      2       10      3       3       4       4       0       3       5       3
        14      3       13      4       4       3       3       3       0       4       6
        18      5       15      6       8       7       7       5       4       0       6
        14      3       13      2       4       3       3       3       6       6       0
;
        USERTYPE ScenTerp1 (Stepmatrix) =
                 11
        0       1       2       3       4       5       6       7       8       9       A
        0       19      8       16      15      24      18      18      18      25      20
        19      0       18      10      11      15      14      10      11      13      12
        8       18      0       18      15      20      22      14      15      21      20
        16      10      18      0       7       11      8       10      9       11      12
        15      11      15      7       0       12      11      7       6       14      11
        24      15      20      11      12      0       5       15      12      14      9
        18      14      22      8       11      5       0       12      11      15      6
        18      10      14      10      7       15      12      0       5       13      10
        18      11      15      9       6       12      11      5       0       10      13
        25      13      21      11      14      14      15      13      10      0       17
        20      12      20      12      11      9       6       10      13      17      0
;
        USERTYPE ScenTerp2 (Stepmatrix) =
                 11
        0       1       2       3       4       5       6       7       8       9       A
        0       19      10      17      16      25      18      19      19      26      20
        19      0       19      11      12      16      14      11      13      14      12
        10      19      0       18      15      20      23      16      16      21      21
        17      11      18      0       7       11      9       11      10      11      13
        16      12      15      7       0       12      12      9       7       14      12
        25      16      20      11      12      0       6       17      13      14      10
        18      14      23      9       12      6       0       13      13      16      6
        19      11      16      11      9       17      13      0       6       15      11
        19      13      16      10      7       13      13      6       0       11      15
        26      14      21      11      14      14      16      15      11      0       18
        20      12      21      13      12      10      6       11      15      18      0
;
END;
```

Polymorph, Chor1

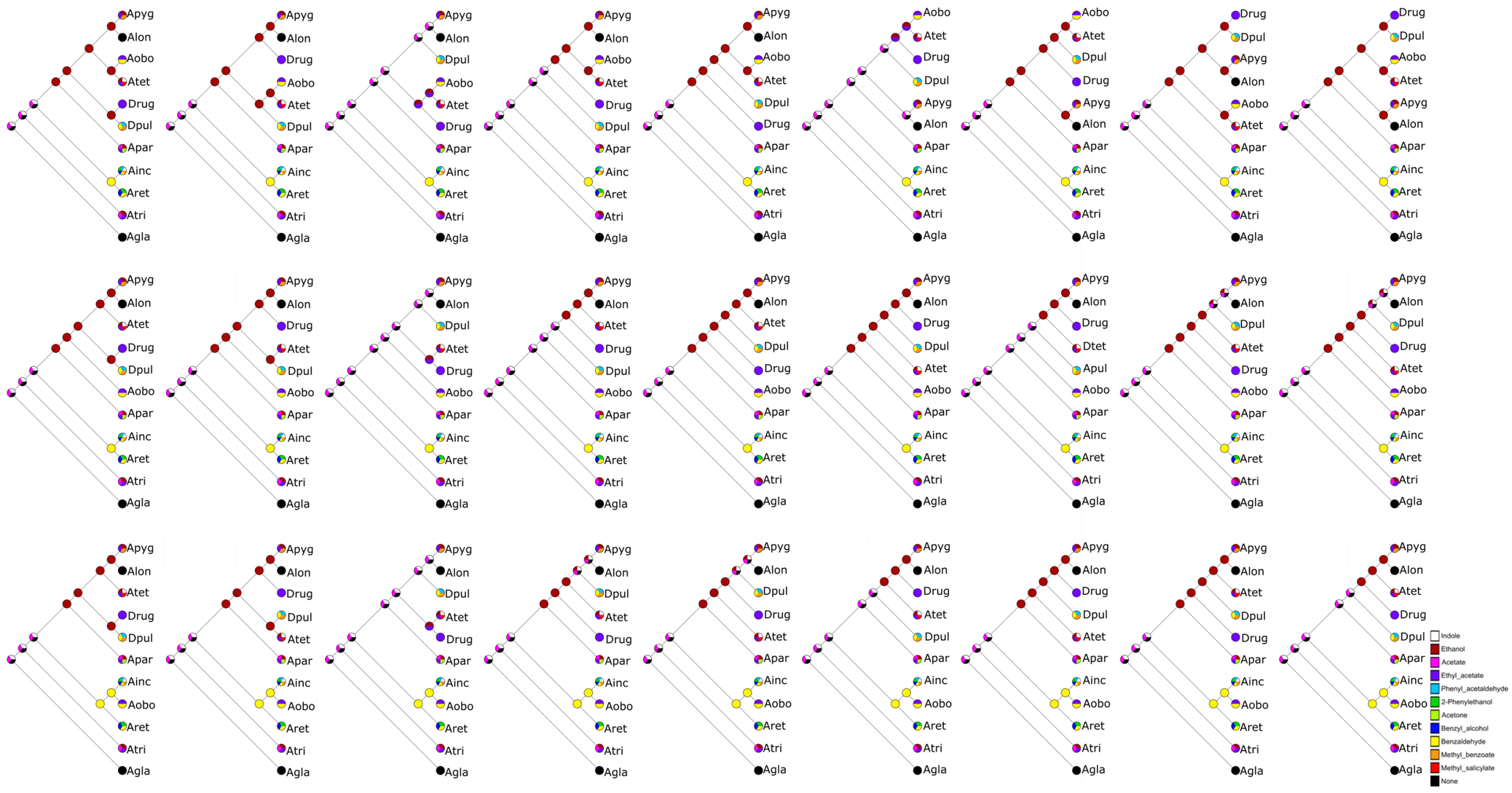


Figure I Optimization of VOCs from the Chorismate-pathway as polymorphic characters, with StepChor1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Scenario, Chor1

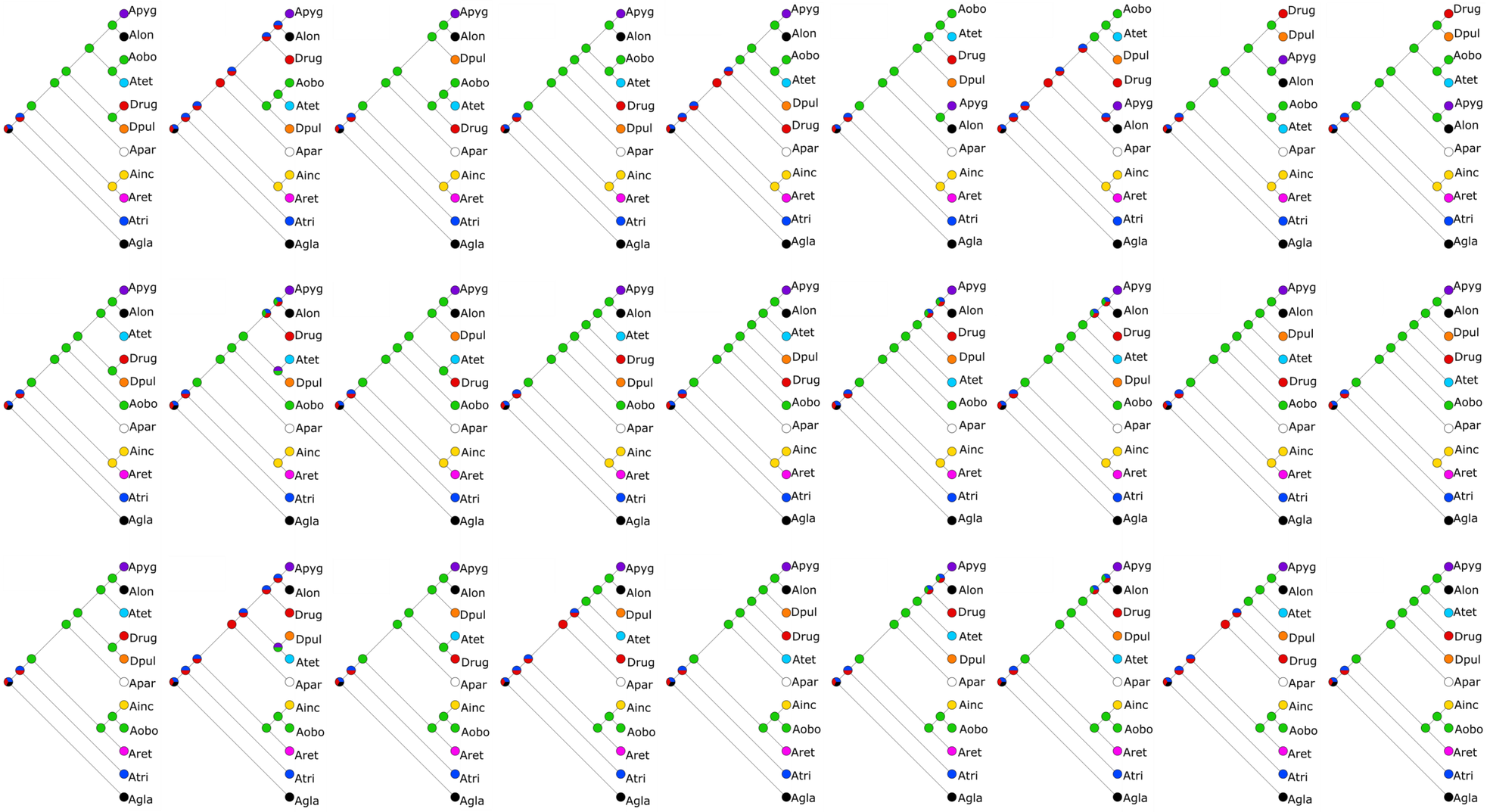


Figure II Optimization of the scenarios from the Chorismate-pathway, with ScenChor1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Scenario, Chor2

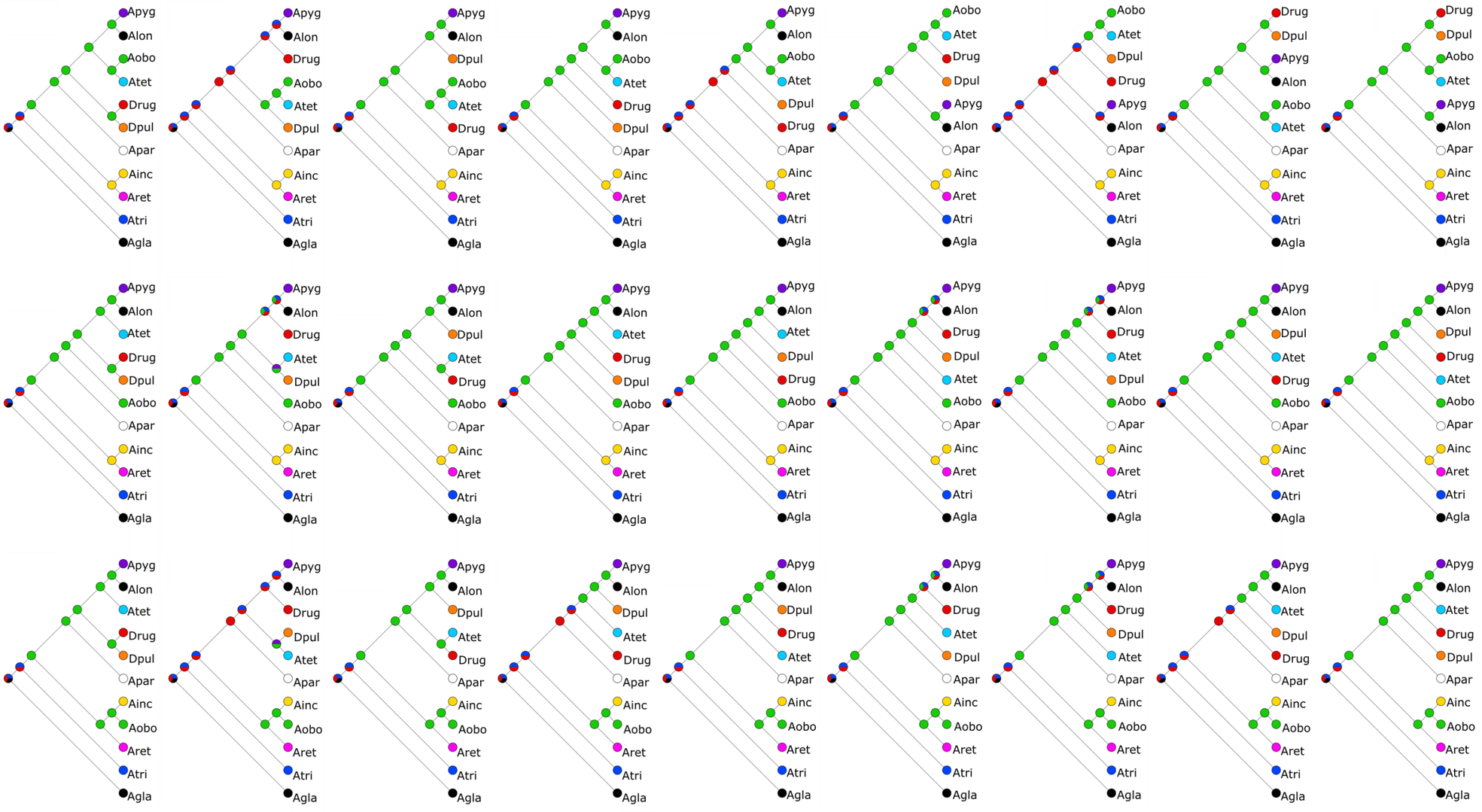


Figure III Optimization of the scenarios from the Chorismate-pathway, with ScenChor1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Polymorph, MonoTerp1

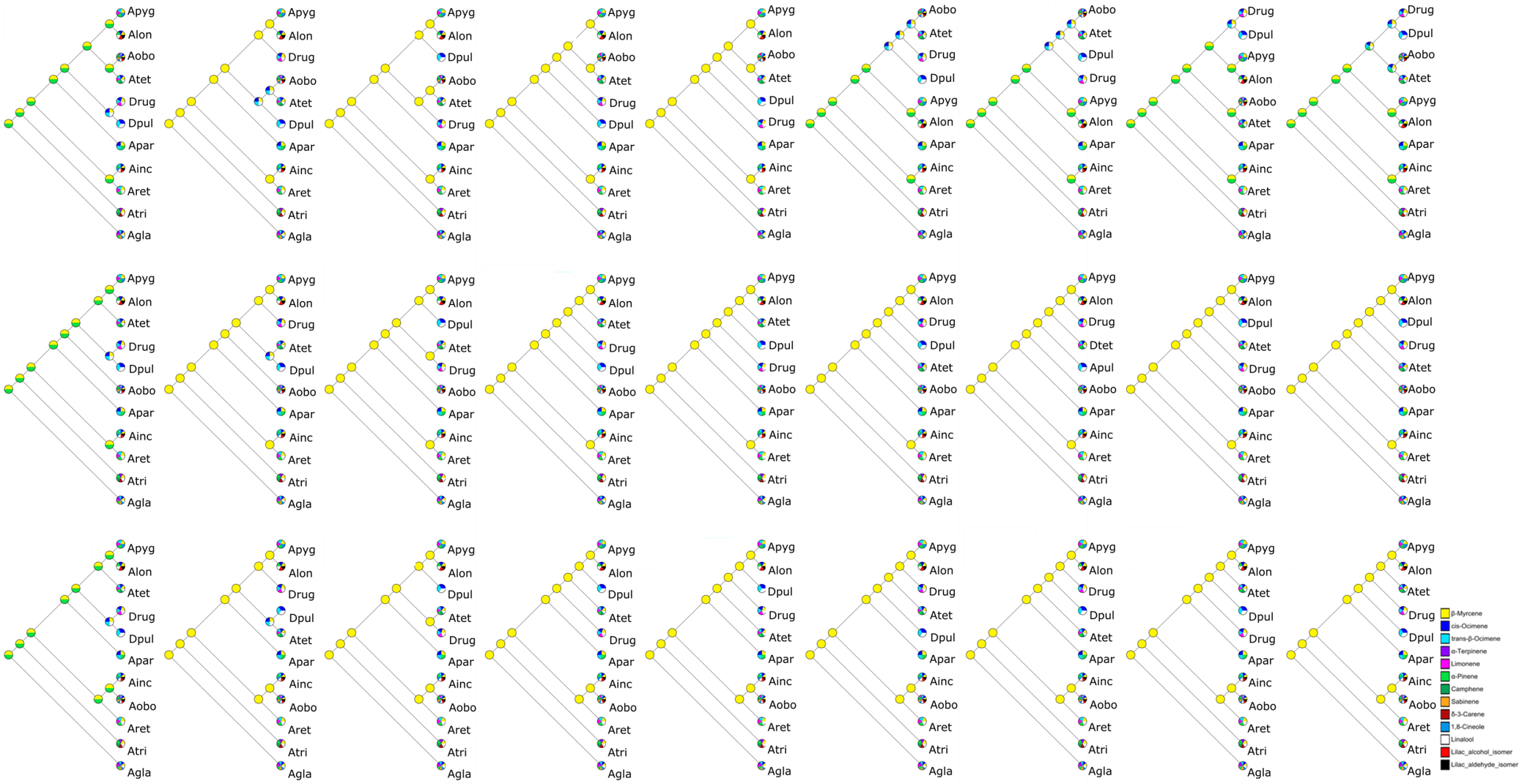


Figure IV Optimization of Monoterpenes as polymorphic characters, with StepMono1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Scenario, MonoTerp1

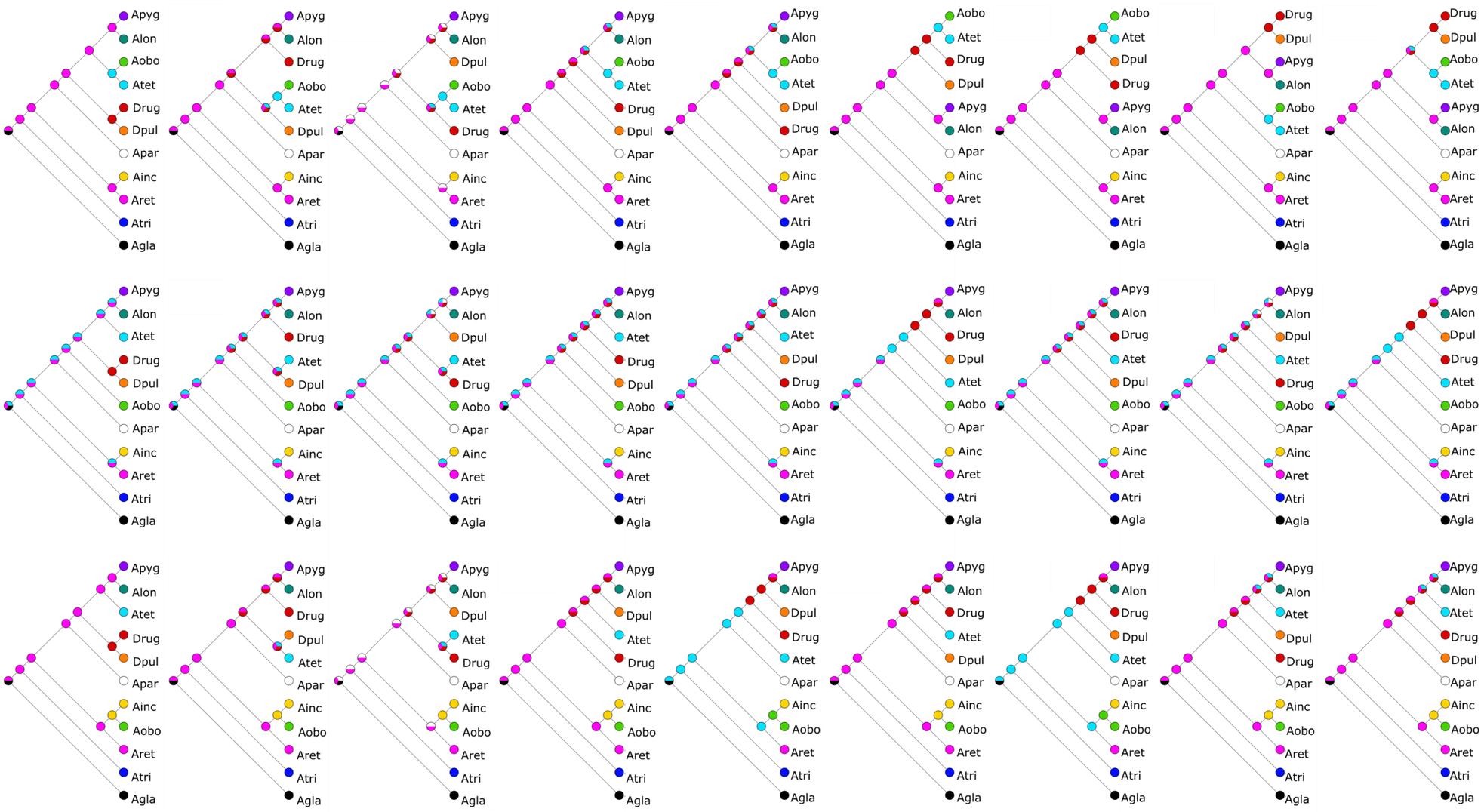


Figure V Optimization of the Monoterpene-scenarios, with ScenMono1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Scenario, MonoTerp2

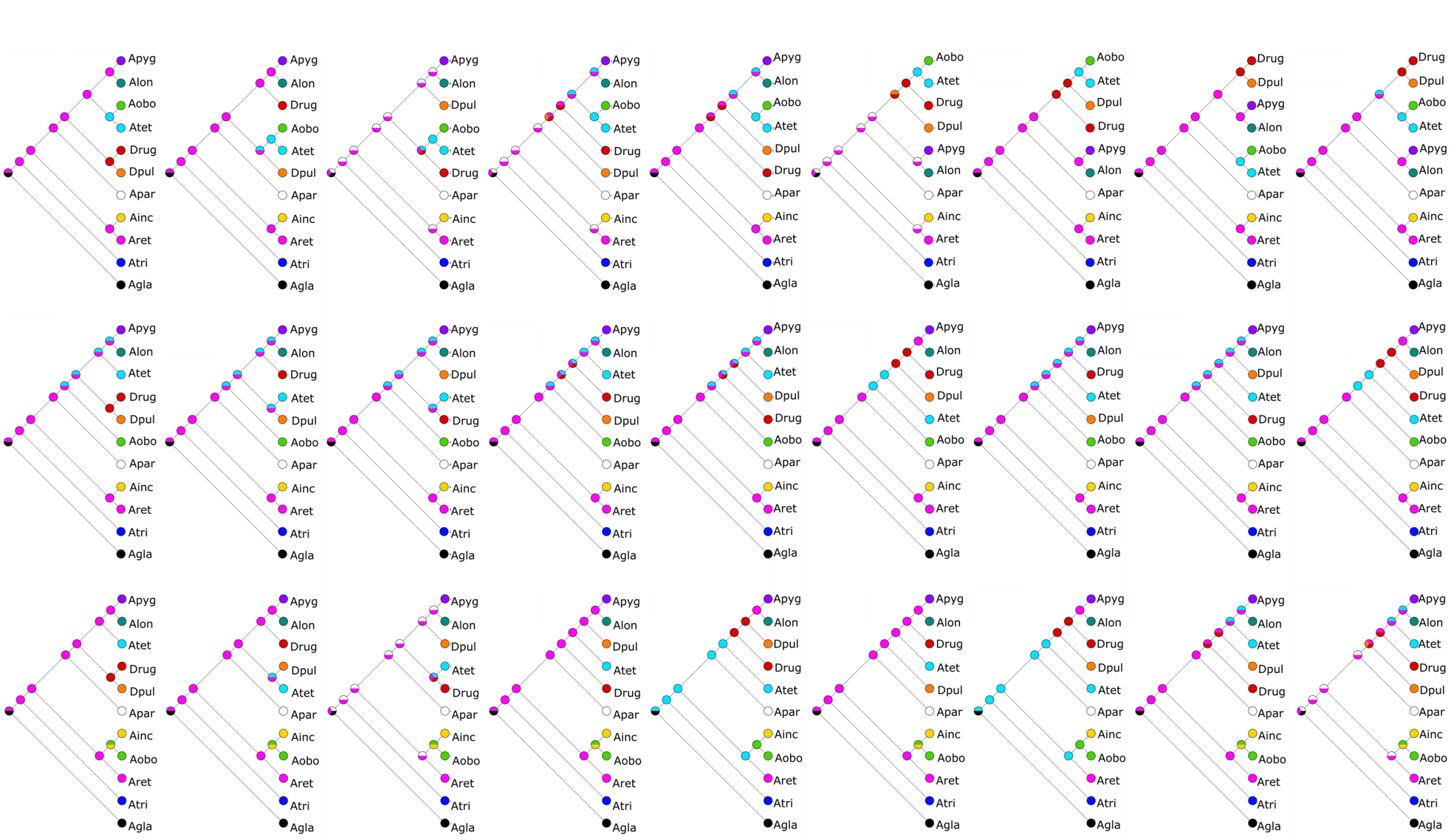


Figure VI Optimization of the Monoterpene-scenarios, with ScenMono2 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Polymorph, Sesqui1

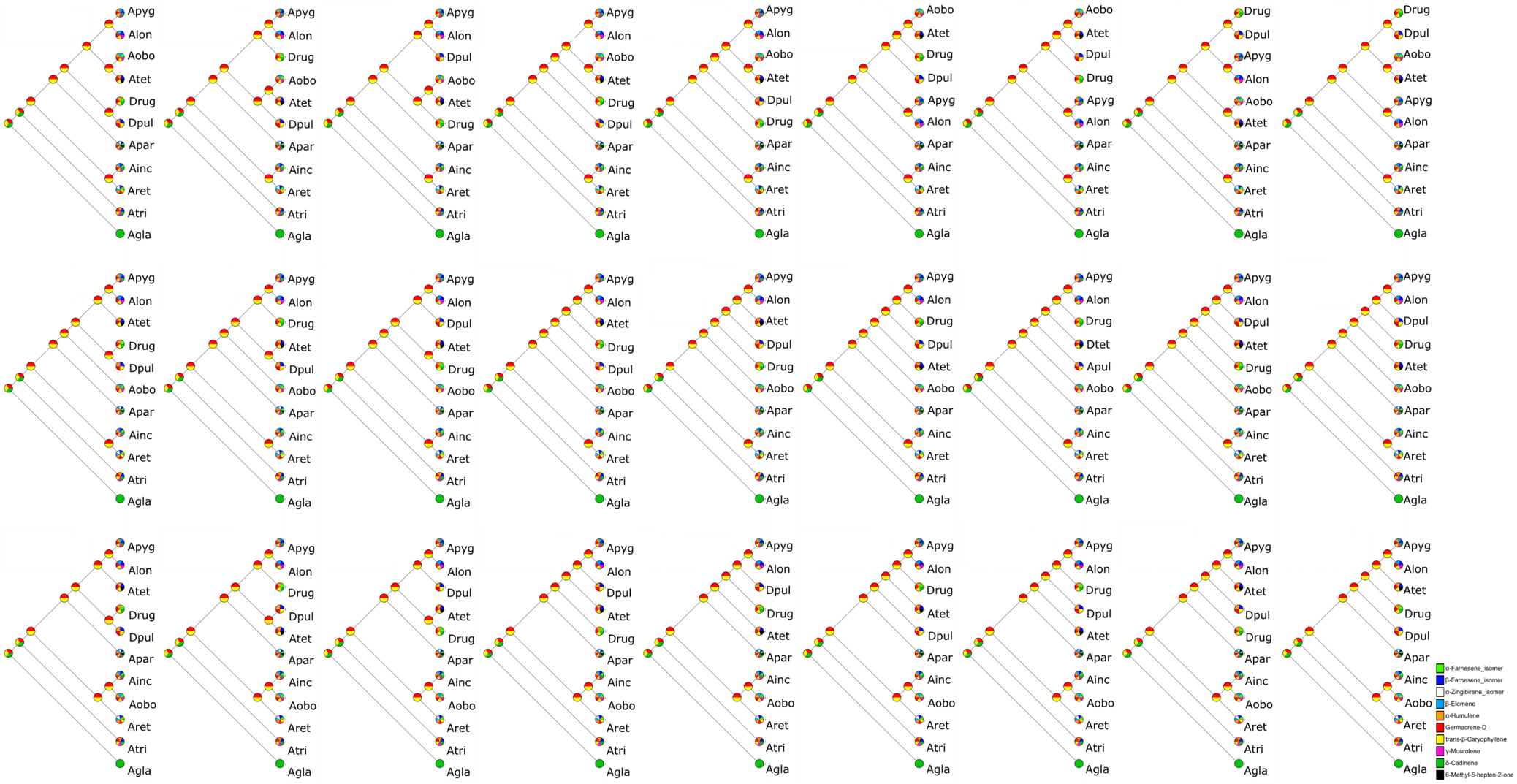


Figure VII Optimization of Sesquiterpenes as polymorphic characters, with StepSesqui1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Scenario, Sesqui1

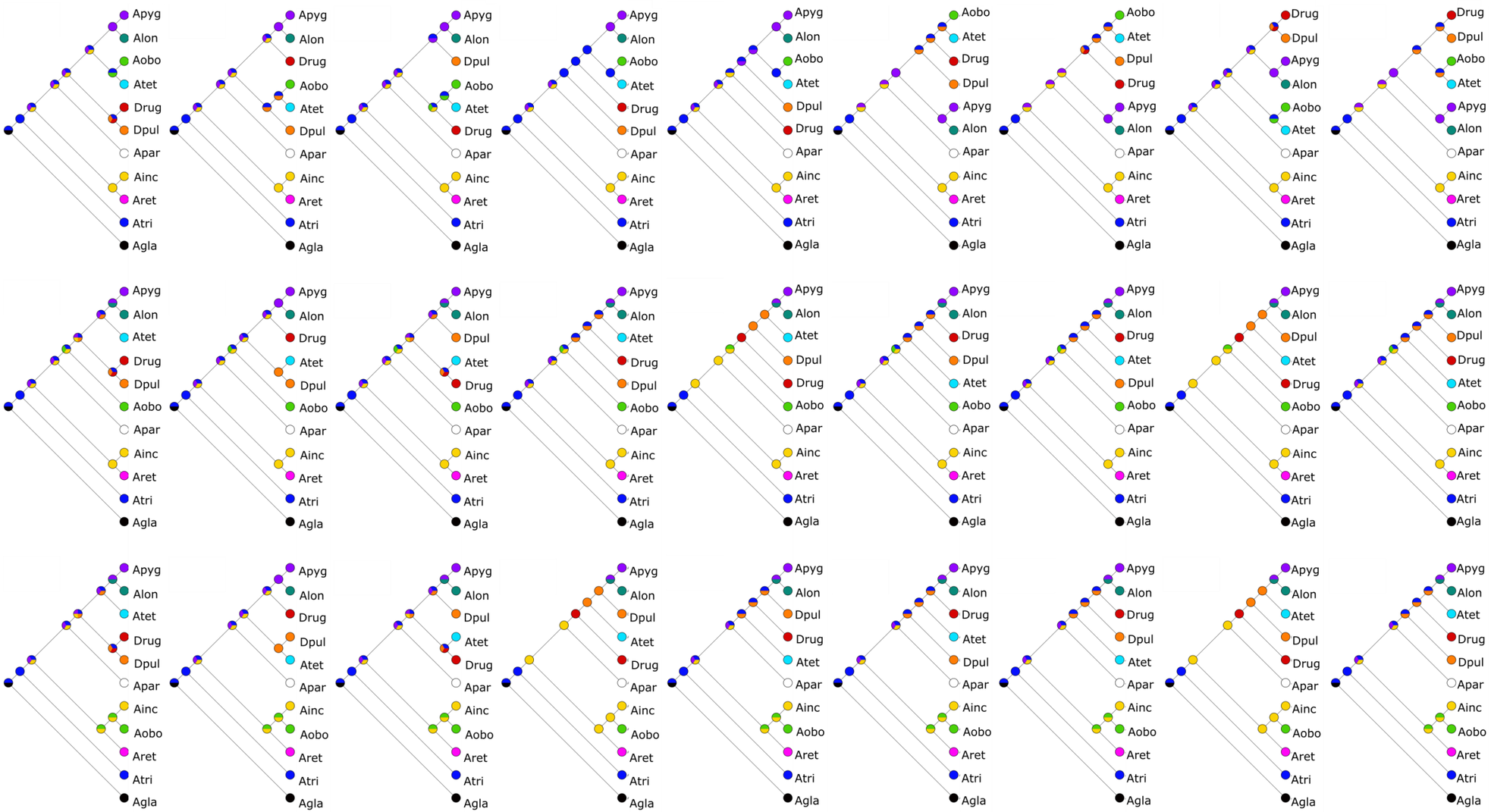


Figure VIII Optimization of the Sesquiterpene-scenarios, with ScenSesqui1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Scenario, Terp1

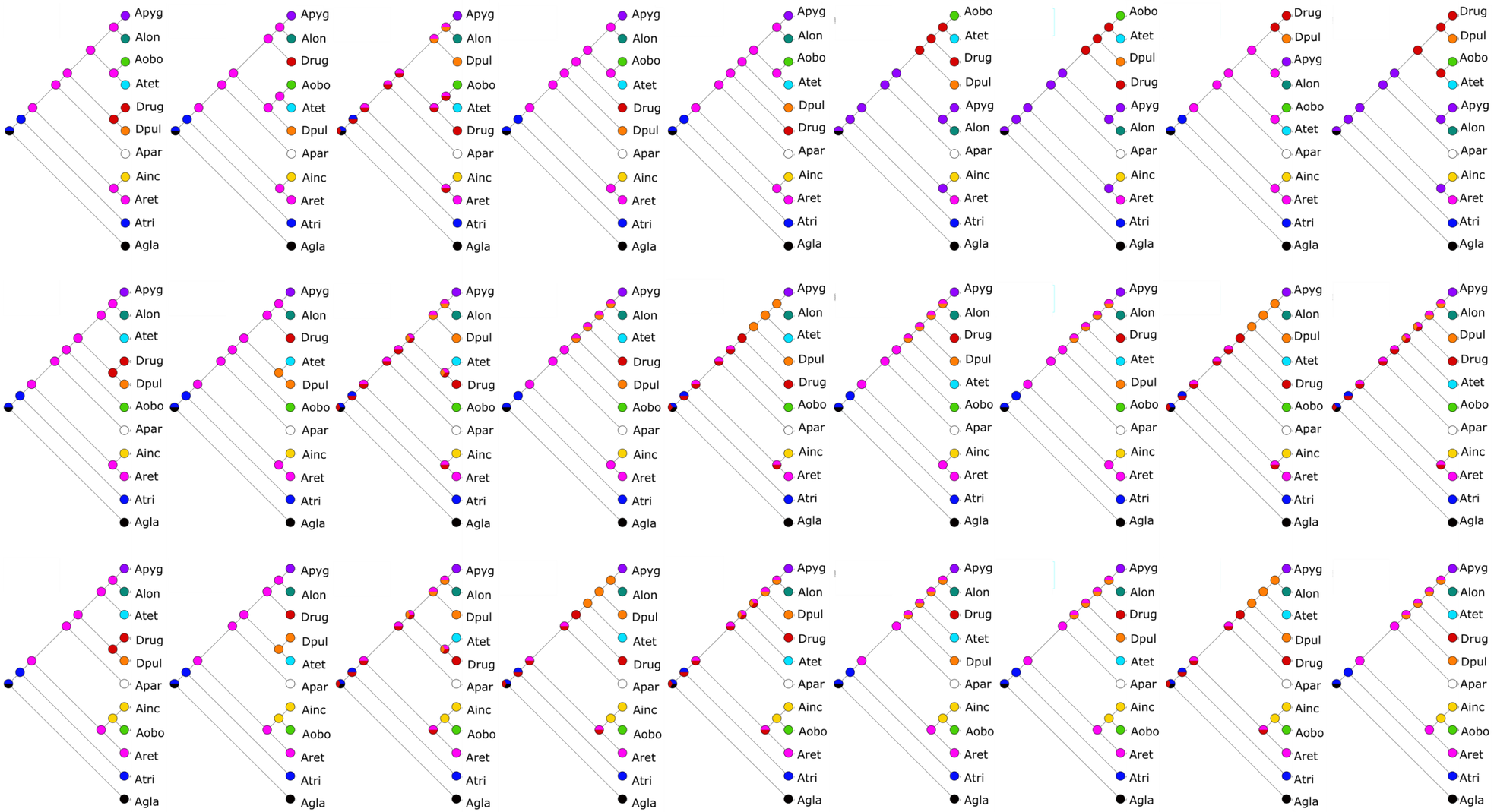


Figure IX Optimization of the Terpene-scenarios, with ScenTerp1 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

Scenario, Terp2

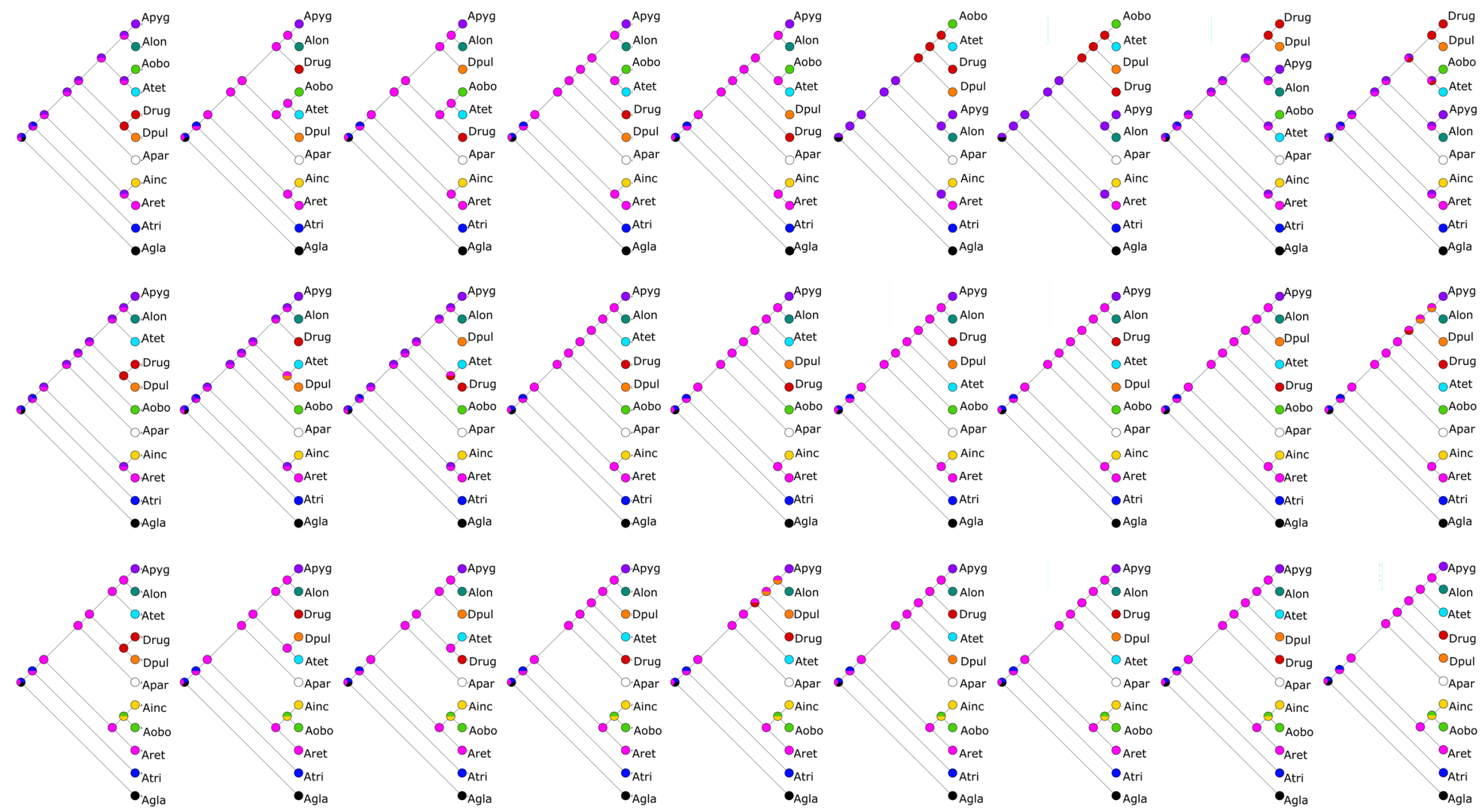


Figure X Optimization of the Terpene-scenarios, with ScenTerp2 as parsimony model, over the 27 resolved trees. The upper row shows the 9 possible resolved trees for the first placement of *Asimina abovata*, the second row does the same but for the second placement. The same applies to the third row.

pScores

Table IV Relative pScores per tree, per stepmatrix.

	POLYMORPH					SCENARIO						
Tree	Chor 1	Chor 2	Sesqui 1	Mono 1	Mono 2	Chor 1	Chor 2	Sesqui 1	Mono 1	Mono 2	Terp 1	Terp 2
Obov1+	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	0.96	0.96	1.00	1.00
Obov1a	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	0.92	0.91	0.98	0.97
Obov1b	1.00	1.00	1.00	1.00	1.00	0.89	0.89	1.00	0.96	0.96	1.00	1.00
Obov1c	0.99	0.99	1.00	1.00	1.00	0.92	0.92	1.00	0.96	0.96	1.00	1.00
Obov1d	0.99	0.99	1.00	1.00	1.00	0.92	0.92	1.00	0.96	0.96	1.00	1.00
Obov1e	1.00	1.00	1.00	1.00	1.00	0.89	0.89	1.00	0.96	0.96	1.00	1.00
Obov1f	0.99	0.99	1.00	1.00	1.00	0.92	0.92	0.98	0.92	0.93	0.97	0.97
Obov1g	1.00	1.00	1.00	1.00	1.00	0.89	0.89	0.98	0.92	0.93	0.97	0.97
Obov1h	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	0.92	0.91	0.98	0.97
Obov1i	1.00	1.00	1.00	1.00	1.00	0.92	0.92	0.98	0.92	0.91	0.97	0.97
Obov2+	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	1.00	1.00	1.00	1.00
Obov2a	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	0.96	0.95	0.98	0.97
Obov2b	1.00	1.00	1.00	1.00	1.00	0.92	0.92	0.96	1.00	1.00	0.99	1.00
Obov2c	0.99	0.99	1.00	1.00	1.00	0.92	0.92	1.00	1.00	1.00	1.00	1.00
Obov2d	0.99	0.99	1.00	1.00	1.00	0.92	0.92	1.00	1.00	1.00	1.00	1.00
Obov2e	1.00	1.00	1.00	1.00	1.00	0.92	0.92	0.98	1.00	1.00	0.99	1.00
Obov2f	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	0.96	0.96	1.00	1.00
Obov2g	0.99	0.99	1.00	1.00	1.00	0.92	0.92	1.00	1.00	1.00	1.00	1.00
Obov2h	1.00	1.00	1.00	1.00	1.00	0.92	0.92	0.98	1.00	1.00	0.99	1.00
Obov2i	1.00	1.00	1.00	1.00	1.00	0.92	0.92	1.00	0.96	0.96	1.00	1.00
Obov3+	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.91	0.94	0.95
Obov3a	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.84	0.86	0.92	0.92
Obov3b	1.00	1.00	1.00	1.00	1.00	0.97	0.97	0.96	0.88	0.91	0.93	0.95
Obov3c	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.91	0.94	0.95
Obov3d	1.00	1.00	1.00	1.00	1.00	0.97	0.97	0.98	0.88	0.91	0.93	0.95
Obov3e	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.86	0.89	0.94	0.95
Obov3f	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.91	0.94	0.95
Obov3g	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.86	0.89	0.94	0.95
Obov3h	1.00	1.00	1.00	1.00	1.00	0.97	0.97	0.98	0.88	0.91	0.93	0.95
Obov3i	0.99	0.99	1.00	1.00	1.00	1.00	1.00	1.00	0.88	0.91	0.94	0.95

Fieldwork

During a period of four weeks, plant material and floral fragrances from Annonaceae were collected in Costa Rica. The species sampled were as diverse as possible, in order to obtain a broad spectrum of floral fragrances within this particular plant family. Fieldwork was done together with supervisor Dr. Lars Chatrou (first three weeks) and fellow-student Rutger Wilschut BSc.

Fieldtrips

6-10 September: Osa Peninsula

Between the sixth and tenth of September, collections were made on the beautiful Osa Peninsula. A total of 23 plants were collected, 18 of which were members of the Annonaceae. Seven Annonaceae-collections had fragrant flowers, and were sampled for their floral fragrance: *Guatteria lucens* Standl., *Guatteria pudica* N.Zamora & Maas, *Anaxagorea dolichocarpa* Sprague & Sandw., *Annona muricata* .L., *Guatteria amplifolia* Triana & Planch. (2 verschillende bomen), *Xylopia aromatica* (Lam.) Mart..

12-16 September: One-day trips from Santo Domingo de Heredia

National Parks that were visited between the 12th and 16th of September 2011 were: Braulio Carrillo, Carara, El Rodeo, Cerro Turrubares and, once again, Braulio Carrillo. A total of 34 plants were collected, of which 17 were Annonaceae. From *Anaxagorea phaeocarpa* Mart. and *Guatteria aeruginosa* Standl., the floral fragrance was collected. These two collections were made in National Park Braulio Carillo.

17 September: Armed mugging

After dinner money was withdrawn, to pay for the upcoming fieldtrip, which was supposed to start the next morning. Unfortunately, the fieldtrip had to be postponed, as there was an encounter with a young man and his pistol, and passports etcetera were lost. The days that followed were dedicated to getting a new passport and visa, as these were needed to return to the home-country.

21 September: San Carlos

A yet undescribed *Mosannona*-species was once collected in San Carlos. As Rutger's research was all about the genus *Mosannona* and its distribution, a search for this species was started. Despite all the hard work, the species was not found, and neither were any flowering Annonaceae. In total, 2 collections were made, both from Annonaceae species.

26-29 September: Límon and surroundings

During the final week (without Dr. Lars Chatrou), the Caribbean coast was chosen to be the next research area. Nine species were collected, all Annonaceae. Regrettably, none had (fragrant) flowers.

Permits

For some reason the permits to export the floral fragrances was never granted. All floral fragrances, collected during the days as described above, are still in a freezer at the InBio institute, in Costa Rica.

Field Data

Table V Field data for flowering Annonaceae, collected in Costa Rica in September 2011.

Fragrance	Main Collector	Add collectors	Coll. Number	Day	Month	Year	Family	Genus	species	Author	Habitat
1	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	705	8	09	20 11	Annonaceae	Guatteria	lucens	Standl.	West-facing steep slope in meadow accompanying small river
2	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	706	8	09	20 11	Annonaceae	Guatteria	pudica	N.Zamora & Maas	
3	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	709	9	09	20 11	Annonaceae	Anaxagorea	dolichocarpa	Sprague & Sandw.	
4	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	710	9	09	20 11	Annonaceae	Annona	muricata		Roadside vegetation
5	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	711	9	09	20 11	Annonaceae	Guatteria	amplifolia	Triana & Planch.	
6	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	712	9	09	20 11	Annonaceae	Guatteria	amplifolia	Triana & Planch.	
7	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	715	10	09	20 11	Annonaceae	Xylopia	aromatica	(Lam.) Mart.	Roadside vegetation
8	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	717	12	09	20 11	Annonaceae	Anaxagorea	phaeocarpa	Mart.	
9	Chatrou, L.W.	N. Zamora; J.L. van de Poel; R.A. Wilschut	720	12	09	20 11	Annonaceae	Guatteria	aeruginosa		

Fragrance	Country	Major area	Minor area	Gazetteer	Coördinates			
1	Costa Rica	Prov. Puntarenas	cantón Osa			N		W
2	Costa Rica	Prov. Puntarenas	cantón Osa			N		W
3	Costa Rica	Prov. Puntarenas	cantón Osa			N		W
4	Costa Rica	Prov. Puntarenas	cantón Osa	Rincón de Osa, on the road to Agua Buena		N		W
5	Costa Rica	Prov. Puntarenas	cantón Osa			N		W
6	Costa Rica	Prov. Puntarenas	cantón Osa			N		W
7	Costa Rica					N		W
8	Costa Rica	Prov. Limón	cantón Pococí	Parque Nacional Braulio Carillo, Entrada Aguas above Río Costa Rica, Finca Bosque Lluvioso, property of INBio	10° 11 52"	N	83° 51' 24"	W
9	Costa Rica	Prov. Limón	cantón Pococí	Parque Nacional Braulio Carillo, Entrada Aguas above Río Costa Rica, Finca Bosque Lluvioso, property of INBio	10° 11 52"	N	83° 51' 24"	W

Fragrance	Habitus
1	Tree, 10 m tall, 30 cm in diam., flowers numerous, petals and sepals green, stamen connectives orange, stigmas yellow, immature fruits numerous, green with red tinges [check with pictures]. Flowers with fruity odour, with elements of citrus and peach
2	Tree, x m tall, with long branches wriggling through the vegetation, flowers [picture for description]
3	Tree, 10 m tall, 8 cm in diam., with distichous branching, several flowers on trunk and thicker branches, pedicel, bract and sepals covered with dense brown indument, petals yellow with brown tinge caused by indument, flowers at anthesis with smell of 'citroenmelisse' and 'bandages'. Later smell of artificial banana.
4	Tree, 15 m tall, 35 cm in diam., many flowers in all stages of development, pedicels and sepals green, outer petals greenish-yellow, inner petals creamy-yellow. Flowers at anthesis with smell of boiled milk mixed with lemon, visited by Cycophela sp. (Scarabeideae). Few mature fruits, green, uncollected
5	Woody plant with seven inclining stems of max. 5 cm in diam., originating from the base of the plant, reaching height of 8 m, young leaves 'slap' and reddish brown, immature petals green with orange-brown base, flowers at anthesis with green pedicel and sepals, dark yellow petals, smell with elements of pineapple (8 am) and banana (1 pm - 6 pm)
6	Tree, 12 m tall, 15 cm in diam., young leaves 'slap' and reddish brown, immature petals green with orange-brown base, flowers at anthesis with green pedicel and sepals, dark yellow petals, smell with elements of pineapple (8 am) and banana (1 pm - 6 pm)
7	Tree, 5 m, 10 cm in diam., petals bright white, outer petals with reddish-brown indument on outer side, immature fruits green, maturing dark red
8	Tree, 4 m tall, 4 cm in diam., few flowers and flower buds, pedicels green, bracts and sepals brown, petals creamy-yellow with brown indument. Immature fruits yellow with reddish-brown apex, flowers at anthesis with with smell with elements of banana and fermentation-alcohol
9	Tree, 17 m tall, 35 cm in diam., a multitude of flowers, [picture description from photo], flowers at anthesis with smell of spice bush (later), later with strong banana smell (14.30 h)