Genetic Food Diagnostics

Approaches and Limitations of Species Level Diagnostics in Flowering Plants

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

Fakultät für Chemie und Biowissenschaften

Karlsruher Institut für Technologie (KIT) - Universitätsbereich

genehmigte

DISSERTATION

von

Dipl. Biologe

Thomas Horn

aus

77709 Wolfach

Dekan: Prof. Dr. Peter Roesky

Referent: Prof. Dr. Peter Nick

Korreferent: Prof. Dr. Horst Taraschewski

Tag der mündlichen Prüfung: 17.04.2014

Parts of this work are derived from the following publications:

Horn T, Völker J, Rühle M, Häser A, Jürges G, Nick P; 2013; Genetic authentication by RFLP versus ARMS? The case of Moldavian Dragonhead (*Dracocephalum moldavica* L.). European Food Research and Technology, doi 10.1007/s00217-013-2089-4

Horn T, Barth A, Rühle M, Häser A, Jürges G, Nick P; 2012; Molecular Diagnostics of Lemon Myrtle (*Backhousia citriodora* versus *Leptospermum citratum*). European Food Research and Technology, doi 10.1007/s00217-012-1688-9

Also included are works from the following teaching projects:

RAPD Analysis and SCAR design in the TCM complex *Clematis* Armandii Caulis (chuān mù tōng), F2 Plant Evolution, 2011

Effects of highly fragmented DNA on PCR, F3, Lidija Krebs, 2012

I. Acknowledgement

"Nothing is permanent except change"
Heraclitus of Ephesus

Entering adolescence – approximately 24 years ago – many aspects of life pretty much escaped my understanding. After a period of turmoil and subsequent experience of a life as laborer lacking an education, I realized that I did not want to settle for this kind of life. I wanted to change. With this work I would like to thank all people that ever bothered trying to explain the world to me, that allowed me to find my way and nurtured my desire to change.

In particular I would like to thank my family, my friends and colleagues for their support. A huge amount of gratitude goes to Peter Nick who offered me the opportunity to work in the field of biodiversity. I would also like to thank Hannes König, CEO of Phytocomm (Kehl, Germany) for the continuous financial support regarding our work in the area of Traditional Chinese Medicine.

For constant support in plant cultivation and patience during the establishment of the plant database I would like to thank the staff of the KIT botanical garden, in particular Angelika Piernitzki and Joachim Daumann. Last but not least, for technical support, I thank Ernst Heene and for administrative support Renate Herberger-Biester and Jochen Krüger.

Thomas Horn

II. Zusammenfassung

In der vorliegenden Arbeit wurden genetische Methoden zur taxonomischen Identifizierung von pflanzlichen Komponenten aus kommerziellen Produkten evaluiert.

Eine grundlegende Voraussetzung für die genetische Analyse mittels PCR basierter Methoden ist die erfolgreiche Isolierung von qualitativ hochwertiger DNA. Neben den inhibierenden Eigenschaften von sekundären Pflanzenstoffen auf die DNA Polymerase spielt die DNA Fragmentierung eine entscheidende Rolle bei der gängigen Methodik. Da vor allem bei der Extraktion aus prozessiertem Pflanzenmaterial mit hoher DNA Fragmentierung zu rechnen ist haben wir den Effekt selbiger auf die Amplifizierung von genetischen Markern untersucht. Die Ergebnisse weisen darauf hin, dass selbst bei starker Fragmentierung eine Amplifikation entsprechender Marker möglich ist, solange eine ausreichende Menge reiner DNA extrahiert werden kann.

Im nächsten Schritt haben wir anhand zweier Produkte Methoden getestet, um bestimmte taxonomische Einheiten zu identifizieren. Im ersten Fall thematisierten wir taxonomisch unklare Beschriftungen auf pflanzlichen Produkten. Unter der Komponente ‚Lemon Myrte' findet man in der Literatur zwei verschiedene Arten, *Backhousia citriodora* und *Leptospermum citratum*, welche wir mittels differentiellem Restriktionsverdau (PCR-RFLP) des rbcLa Markers voneinander unterscheiden konnten. Im zweiten Fall ging es um Verwechslungen von Produkt Komponenten, welche fatale medizinische Folgen mit sich bringen können. Hier haben wir mittels ‚Random Amplified Polymorphic DNA' (RAPD) einen spezifischen Marker für eine Art der Gattung *Aristolochia* entwickelt, welche in der Literatur als Verwechslungsart des pharmazeutischen Produkts Clematidis Armandii Caulis erwähnt wird. Die Spezifität des Markers konnte anhand der verwendeten Akzessionen gezeigt werden, wobei eine ausreichend robuste Validierung der Art- bzw. Gattungsspezifität aufgrund fehlender Referenzpflanzen nicht durchgeführt werden konnte.

Eine weitere grundlegende Voraussetzung für unsere Arbeit sind ausreichend korrekt identifizierte Referenzpflanzen. Zur Verbesserung der Organisation von entsprechenden Pflanzen Akzessionen des botanischen Gartens und zur Erweiterung der Verwaltung von notwendigen Hintergrundinformationen wurde eine interaktive Datenbank Anwendung erstellt. Im Vordergrund stand hier eine konsistente Datengrundlage zu schaffen, eine gemeinsame Verwaltung zuzulassen und die Anbindung der Datensätze an öffentlich zugängliche Hintergrundinformationen wie

taxonomischer Status, traditionelle und zeitgenössische Verwendung, Verbreitung, molekulare Daten, usw. zu realisieren.

Für eine tiefergehende Evaluierung von aktuell verwendeten Methoden zur genetischen Identifizierung von Pflanzenmaterial haben wir uns anschließend mit der Gattung *Dracocephalum* näher beschäftigt. Da der Moldawische Drachenkopf (*D. moldavica*) Anwendung als Tee findet haben wir in der ersten Phase PCR-RFLP dazu verwendet Verwechslungskandidaten (*Melissa*, *Nepeta*) in entsprechenden Produkten aufzuspüren. Für den Nachweis von Drachenkopf selbst haben wir hier zum ersten Mal das ‚Amplified Refractory Mutation System' (ARMS) für die Entwicklung eines diagnostischen Primer angewendet. Im Gegensatz zu PCR-RFLP erlaubt dieser Ansatz den direkten Nachweis (ohne zusätzlichen Verdau) und erhöht die Robustheit der Methode durch die Verwendung von universellen Primer als PCR positiv Kontrolle.

Mit dem Ziel von der gattungs- zur artspezifischen Diagnostik vorzudringen und verschiedene gängige genetische Marker auf deren Tauglichkeit zu untersuchen haben wir in der zweiten Phase insgesamt 55 Saatgut Akzessionen von 10 *Dracocephalum* Arten über das internationale Pflanzenaustausch Programm (IPEN), das Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK) und einen kommerziellen Anbieter erworben. Dabei bestand der erste Teil darin, die Identität der Akzessionen morphologisch zu überprüfen, wobei zwei Akzessionen identifiziert wurden, die einer anderen Gattung angehören, und eine weitere, welche zu einer anderen *Dracocephalum* Art zu zählen ist. Da ein Teil der Pflanzen nicht nach Blütenmerkmalen bestimmt werden konnte und wir regelmäßig auf Saatgut als Referenzquelle zurückgreifen, haben wir Saatgut Eigenschaften mit Hilfe des Programms SmartGrain auf Artspezifität untersucht. Zusätzlich zu diesen zwei morphologischen Ansätzen haben wir mittels ‚Tubulin based Polymorphismus' (TBP), Sequenz basiertem ‚Multidimensional Scaling' (MDS) und ‚Neighbor-Joining' (NJ) drei molekulare Methoden zur Identitätsprüfung angewendet. Für die Sequenz basierten Methoden wurden Daten von 4 Markern erhoben (rbcLa, matK-KIM, psbA-trnH und ITS45). Saatgut Charakter erwiesen sich als hilfreich bei der Identifizierung von fehlerhaft ausgezeichneten Akzessionen, sind aber stellenweise nur von limitiertem Nutzen. TBP Muster zeigten gutes Potential für die Anwendung bei der Authentifizierung von Referenz Arten. Zusätzlich wurden intraspezifische Variationen in *D. ruyschiana* gefunden, welche auf eine genetische Isolierung innerhalb der Art hinweisen könnten. Sequenz basiertes MDS mit den Markern psbA-trnH und ITS45 zeigte eine deutliche Abgrenzung der *Dracocephalum* Arten von anderen Gattungen, wobei *Schizonepeta* und *Hyssopus* die nächst näher verwandten Gattungen

sind. Mit Hilfe des NJ konnten abschließend alle ermittelten Fehlbestimmungen bestätigt und durch die Verwendung von Referenzsequenzen (z.B. GenBank) andere Identitäten gefestigt werden. Amplifikation und Sequenzierung gestaltete sich mit den Markern rbcLa und psbA-trnH am erfolgreichsten und matK-KIM sowie ITS45 erwiesen sich stelleweise als problematisch.

Anhand des taxonomisch korrigierten Datensatzes wurde dann eine Marker Evaluation durchgeführt. Zum einen wurde das Programm TaxonGap verwendet, welches Separabilität und Heterogenität der Arten in Bezug auf die Ähnlichkeit der Marker Informationen darstellt. Hier zeigte sich, dass rbcLa nicht und matK-KIM nur beschränkt zur Unterscheidung zwischen Arten der Gattung *Dracocephalum* geeignet ist. ITS45 und psbA-trnH auf der anderen Seite zeigten deutlich bessere Werte und lassen eine Unterscheidung der untersuchten Arten zu. Anschließend wurden mit den zuletzt genannten Markern ein DNA Barcoding Effizienz Test durchgeführt, wobei das Programm R mit dem Packet SPIDER verwendet wurde. Hierbei zeigte sich psbA-trnH mit 100 % Effizienz bei der Verwendung des nextCloseMatch Algorithmus am geeignetsten als DNA Barcoding Marker zur Proben Identifizierung. Bei der Untersuchung von ITS45 konnte eine Art nicht identifiziert werden. Abschließend wurde mit beiden Markern eine ‚Sliding Window' Analyse durchgeführt welche das Vorkommen diagnostischer Nucleotide visualisiert. Sowohl psbA-trnH als auch ITS45 bieten für die meisten aber nicht alle untersuchten Arten diagnostische Positionen an. Um unser Vorhaben von der gattungs- zur artspezifischen Diagnostik vorzudringen umzusetzen wurde zu Letzt ein ARMS primer spezifisch für *D. moldavica* entwickelt und in einer ‚Multiplex PCR' getestet. Wie in den vorangegangenen Fällen konnten wir auch hier das gewünschte Ziel erreichen und eine bestimmte taxonomische Einheit mittels PCR Diagnostik nachweisen.

Index

III. Introduction

As consumers we usually trust that a product label describes the content appropriately. We also trust that as long as the product can be obtained without prescription it can be consumed like any other common food. This trust is enforced by strict regulations allowing only certain products to enter the market as foods while others (e.g. pharmaceuticals) are subject to different regulations. In the next instance producers are responsible for safety and quality of their products while confronted with various scenarios, e.g. if demands are high and resources limited quality decreases due to frequent supplementation of the limited resource with a cheaper and qualitatively inferior one [Bijoy 2007, Guo 2011]. Another aspect is scientific knowledge. The label of a plant based food product usually contains a common name rather than a scientific one and does not include further information on chemical components. As long as precise information is known by (and available from) the producer meaningful quality controls are possible on a genetic and biochemical basis. The European novel food regulation was introduced to prevent exotic products to rush into the market without sound scientific evidence of compliance to high standards comparable to already established products.

This thesis follows the premise that genetic food diagnostics is a viable tool to supplement current quality control standards by introducing genetic background information on involved taxonomic groups – establishing taxonomic certainty – and by providing quick and cheap yet reliable assays to authenticate products – placing the resource into a taxonomic context. Furthermore the acquired information can be used to assess biodiversity and plan conservation strategies which are especially important for the sustainable use of resources that are collected from natural populations [Guo 2011].

A. Food and Medicine

With autarky gone millions of people depend on a market delivering food and other products from around the world. Heterogeneity in educational background among consumers as well as producers necessitates the establishment of standards by legislative organs. Although many things appear to be common sense – e.g. nothing should be sold that could harm consumers – it usually takes experts from different fields to assess risks and deliver standards that appropriately protect consumers. The European Single Market depends on such standards providing confidence for consumers and transparency for producers.

In the following paragraphs a basic introduction to European food law is given and interfaces with genetic food diagnostics are highlighted.

1. Food Safety and Consumer Protection

“Food” in terms of the European regulation EC N° 178 / 2002 means any substance or product, whether processed, partially processed or unprocessed, intended to be, or reasonably expected to be ingested by humans and does not include medicinal products within the meaning of Council Directives 65 / 65 / EEC (21) and 92 / 73 / EEC (22) as well as residues and contaminants (Article 2 - Definition of "food"). This however does not mean that there are no food products that contain substances that have physiological effects which can be considered medicinal. In December 2006, as consequence of increased demand for health products, the “regulation on nutrition and health claims made on foods” was adopted and enabled vendors to use nutritional claims (e.g. “low fat” or “high fiber”) as well as health claims (e.g. “supporting the immune system”) which indicates the existing overlap between food and medicinal products.

One of the general objectives of European food law is the right of consumers to safe food and to accurate and honest information (EC N° 178 / 2002). The limitation of accuracy can be derived from another included directive which states that food labels must be easy to understand. Although it is stressed that measures relating to food safety must be underpinned by strong science, the area of food law relating to consumer information is excluded (Directive 2000 / 13 / EC). The subsequent compromise is that there are no scientific names of the source organisms on labels but vernacular names which are scientifically inaccurate and prone to being misunderstood. For example the common name ‘Lemon Myrtle’ covers two completely different species, *Backhousia citriodora* and *Leptospermum citratum* – a synonym for *L. petersonii subsp. petersonii*. These two species not only differ with respect to secondary compounds but also in dose-dependent toxicity. Using simple but accurate genetic diagnostics (Figure 2) producers could differentiate between these two species and use a more appropriate name for their product.

Article 14 of regulation EC N° 178 / 2002 introduces definitions and formulations to protect consumers from food that is either a health risk or unacceptable. This includes foods that are contaminated whether by extraneous matter or otherwise, or through putrefaction, deterioration or decay (Article 14). Adulteration of herbal products with surrogates or misidentified species are contaminations that can pose a serious health risk to consumers obligating vendors of respective products to act preventively by application of suitable tests.

2. The Novel Food Regulation

According to the Novel Food Regulation (NFR) the European Commission considers foods and food ingredients that have not been used for human consumption to a significant degree in the EU before 15 May 1997 as novel foods and novel food ingredients. Like “normal” food the respective products have to be safe for consumers and properly labeled. A company that wants to introduce a novel food product must apply to an authority of an EU country for authorization, presenting the scientific information and safety assessment report. Several examples and the impact of the NFR on developing countries have been discussed by Hermann in 2009.

3. “Novel” Medicines

There are many traditional, complementary and alternative medicines that have gained public and scientific interest over the last decades [Fabricant 2001, WHO 2002]. One reason – exemplarily derived from Traditional Chinese Medicine (TCM) – is the appeal of using natural resources in a holistic way – considering the whole to support the whole – as contrast to western medicine where isolated compounds are prescribed to treat localized symptoms. An additional reason however is the potential risk of different traditions when they are not well understood and placed into another context without considering necessary adaptations. For example, in Traditional Chinese Medicine several components (e.g. species) are used to form a medicinal preparation. In this preparation each component implements a specific function and can only be used in combination for the intended purpose. The single component does not have a fixed effect *per se*. Instead its function is defined with respect to all other components. Such a system cannot be explained focusing on a single substance. Thus it is necessary to consult the traditional context when separating one of the components (i.e. chemical or taxon). An almost trivial aspect – naming of plants as part of a medicinal formulation – can have dramatic consequences and is highlighted by an incidence, where more than 100 women in Belgium lost their kidneys after a TCM based slimming regime [Vanherweghem 1993]. The preparation was supposed to contain han fangji (*Stephania tetrandrae*), but was later suspected to contain the nephrotoxic aristolochic acid, most presumably derived from the morphologically similar and almost synonymous guan fangji (*Aristolochia fangchi*). Although the explanations for this incidence remained controversial, as a response to this and similar incidences, the FDA in 2001 proclaimed a ban on a large number of TCM drugs.

Besides clear defined medicines with restricted access, other products containing plants with medicinal potential are freely available. Although these products have been cleared for sale according to their long history of safe use and through elaborate testing with extracts of the source plant, adulterations remain a constant threat.

4. Genetics vs. Biochemistry

While biochemical assays are able to deliver direct evidence for the presence of certain compounds genetic assays present the taxonomic context of the used organisms. Especially in situations where one is confronted with rather new or unexplored groups and thus little knowledge about chemical composition exists, genetic research is a viable choice for identification and exploration.

Biochemical tests are prone to natural variations (i.e. developmental and environmental) in concentration of respective components thus cannot easily detect the origin species used for a certain product. Additionally, for many plants only few of the components and their physiological effects have been adequately characterized. It has been estimated that from 250'000 species of higher plants only 6 % have been screened for biological activity and about 15 % have been evaluated phytochemically [Verpoorte 1998]. Tests that authenticate a plant product merely by using an arbitrary component (e.g. the most abundant) as reference are not delivering real proof of authenticity. Other closely related plants may possess significant amounts of the same component along with others. It is common that chemical constituents are not restricted to a single species (e.g. Rhein in *Rheum* species [Xiao 1984]). A close approximation of authenticity using biochemical analysis can only be achieved by studying the complex compositions of a species and comparison to others. This restricts a precise determination to methods that are far more laborious than a genetic approach (e.g. DNA Barcoding) which uses information stored in one type of molecule present in all forms of cellular life. Additionally, since such efforts are mandatory for the introduction of new products into the European Market small businesses are being penalized [Hermann 2009].

Elucidating proximity of genetic relations benefits the discovery of alternative sources and the development of cultivars. Furthermore genetic information can be used to assess biodiversity and plan conservation strategies which are especially important for the sustainable use of resources that are collected from natural populations [Guo 2011]. While genetic approaches only require small amounts of tissue which can be derived from any part of the organism, biochemical analyses

need to be performed either considering the whole organism or the respective part used for a particular product.

Simple and cost efficient tests based on traces of genetic information can be used to screen small samples for the presence of adulteration thus indicating the need for further examination or the disposal of the respective batch. Additionally, the use of the correct species – authentication – can be diagnosed in parallel. In any case the correct identification of organisms used in products is the most crucial step and should be taken prior to any other scientific test and safety assessment. Since expertise determining species the classical way is shrinking alternative approaches are needed. Methodologies of genetic food diagnostics represent an alternative way to determine the identity of reference material as well as processed food products.

B. Diagnostics Toolbox

For it is the primary goal of this thesis to evaluate the usefulness of genetic food diagnostics we will first establish the methodological background. To verify the presence of DNA from a certain taxonomic group contained within a sample derived from a commercial product, either as authentication measure or to detect adulteration, the methodology used needs to be robust, particularly in regard to DNA fragmentation and purity. Furthermore it should be easy to apply while at the same time being highly reliable. To be able to detect different taxonomic groups without the need of laborious adjustments the methodology should be universally applicable.

1. DNA Amplification and Marker Systems

The polymerase chain reaction (PCR) [Mullis 1986] is a very robust and reproducible technique able to amplify DNA fragments from as little as a few copies of the target region. Products containing at least some traces of DNA from the originating organism should be accessible to genetic diagnostics by PCR. Methods available can be grouped in those who facilitate a pattern based on the whole genome (pattern based diagnostics, PBD), specific to a certain taxonomic group, and those that rely on the specificity of only a fraction of the genome (sequence based diagnostics, SBD).

PBD is either based on the sequence of arbitrary primers e.g. Random Amplified Polymorphic DNA (RAPD) [Williams 1990], Arbitrary Primed PCR, DNA Amplification Fingerprinting (DAF) [Caetano-Anollés 1991], on primers that target certain DNA regions like simple sequence repeats or regions in between, on degenerated primers designed to amplify introns of gene families like the plant β-

tubulin family (Tubulin based Polymorphism [Bardini 2004]) or on the amplification of specific subpopulation of a genome that has been digested with restriction enzymes (Amplified fragment length Polymorphism, AFLP) [Vos 1995]. All of these approaches produce patterns that are characteristic for the respective genome, enabling scientist to detect variations between species, populations and even individuals. The sensitivity of the produced patterns in respect of DNA quality (i.e. level of fragmentation) prevents the application of DNA fingerprinting using processed materials.

SBD on the other hand only requires the respective region of DNA to be accessible for analysis and is therefore preferred diagnosing these materials. The success of such an endeavor still strongly depends on the purity of extracted DNA and the presence of enough template fragments containing the region of interest. DNA extractions from plants that are high in secondary metabolites and commercial products that usually contain certain parts rich in those metabolites frequently fail as templates for PCR due to remaining traces (e.g. polysaccharides, polyphenols) which tend to interfere with PCR [Varma 2007].

2. Sequence Based Diagnostics

a) Developing New Markers

Although PBD is not directly applicable using degraded DNA it is of use in developing SBD primers - also called sequence characterized amplified region (SCAR) primers [Paran 1993]. Pure high molecular weight DNA of respective organisms is used to produce patterns based on restriction fragment length polymorphism (RFLP) [Smith 1970, Botstein 1980] or PCR based DNA fingerprinting techniques (e.g. RAPD, DAF and AFLP). The pattern is compared, a taxonomically unique fragment is isolated and the included sequence information is used to design specific primers.

b) Using Available Markers

Another approach is to develop specific primers based on sequence information that is already available through public databases (e.g. GenBank [Benson 2008], BOLD [Ratnasingham 2007]) or by generating sequence information using known taxon specific or universal primers. Comparing such data specific differences coherent with taxonomy can be used to design an assay that visualizes the respective genetic difference:

PCR based restriction fragment length polymorphism

PCR based restriction fragment length polymorphism (PCR-RFLP) is based on differential marker restriction patterns identified by restriction analysis of sequence data derived from a respective marker. PCR amplification, restriction digestion and subsequent electrophoretic separation of restriction fragments produce an artificial pattern which can be used to distinguish between different taxonomic groups.

Multiplex PCR

A slightly faster approach compared to PCR-RFLP is to use a primer mixture that contains not only primers for the amplification of the marker region but also one or several taxon specific primers (e.g. Amplified Refractory Mutation System; ARMS [Newton 1989]) that anneal within the marker region. Due to the simultaneous amplification of multiple fragments by different primer combinations the approach is called multiplex PCR.

Both approaches can be used to detect single nucleotide polymorphisms in the marker region. In the case of PCR-RFLP coding regions are preferred as target because insertions and deletions frequently occurring in non-coding markers would subsequently complicate the interpretation of the results. Both approaches strongly rely on sufficient sequence information of the marker region, i.e. sampling of individuals from different populations with respect to natural distribution and cultivation. Since only small differences are used to determine which taxon is present in the sample, it has to be ensured that these differences are characteristic for the respective taxon. How wide or narrow the circle of congeneric has to be drawn needs to be decided on a case by case basis.

c) Using DNA Barcoding

During the last decade, DNA Barcoding [Hebert 2003] has been developed as new approach to address genetic identity in numerous organisms. DNA Barcoding is based on sequence diversity in short standardized gene regions of taxonomically pre-defined groups. It is used to identify species in food, wildlife trade, environmental monitoring projects, the validation of processed wood and lumber products and forensic investigations.

Many scientists have employed the formerly described PCR based methods as a mean to authenticate certain plant products [Chiou 2007, Gao 2010, Horn 2012]. The range in quality however is diverse. Critical aspects are geographic, intra-generic and -specific sampling as well as the choice of marker region. While most of these approaches paid little attention to sampling

efficiency and marker evaluation DNA Barcoding and its refinement in the past 10 years provides a valuable framework for the evaluation of these aspects.

DNA Barcoding basically consists of two parts: Firstly, the determination of a useful genetic region – the barcoding region – and secondly, the accumulation of sequence information from reference organisms to establish a species barcode database. Using the database sequences from unknown specimens can be assigned to a species. The proposed mitochondrial DNA Barcoding marker, cytochrome c oxidase I (COI), so far [Nov 2013], has been used to generate 285'061 public DNA barcode index numbers for animals (84 % from *Arthropoda*). Plant barcodes on the other hand are still not integrated into the BIN system of the Barcoding Life Web Site [Nov 2013, www.boldsystems.org]. With COI showing limited variation in land plants [Fazekas 2008, for an exception see Cho 2004] the chloroplast genome which has been used extensively for phylogenetic studies [Olmstead 1992, Kondo 1996, Shaw 2005, Sen 2011] was targeted in the search for a suitable barcoding region. Currently the coding region of the large subunit of ribulose-bisphosphate carboxylase (rbcL) and maturase K (matK) are considered the primary barcoding markers [CBOL plant working group 2009] while the inter-genic spacer between the psbA and trnH genes [Kress 2007, Hollingsworth 2011] as well as the nuclear internal transcribed spacer regions [Chen 2010, China Plant BOL Group 2011] are suggested to be complementing barcoding markers due to their increased variability.

The application of DNA Barcoding in food diagnostics [Stoeckle 2011, Wallace 2012] is based on the amplification of a standardized region of the genome, the retrieval of sequence information and subsequent comparison with a DNA barcode database. Identification success relies on the presence of a representative sequence from individuals of the species that have been pre-determined morphologically. Due to the previously mentioned difficulties in finding a suitable single locus barcode and the use of other non-standard markers in combination, data availability of potential regions varies between different taxonomic groups and needs to be supplemented before reliable diagnostics can be performed.

> *"If we study a system at an inappropriate scale, we may not detect its actual dynamics and patterns but may instead identify patterns that are artifacts of scale. Because we are clever at devising explanations of what we see, we may think we understand the system when we have not even observed it correctly."*
>
> Wiens, 1989

The evaluation of the efficacy of a single or multi locus barcode is based on divergence rates in terms of genetic distance within and between species. The initial way to test the efficacy of a marker was to build a phylogenetic tree and to evaluate the clusters [Hebert 2003]. More recently algorithms that directly work with distance data and even apply distance threshold criteria [Meier 2006] are being used to determine the identification efficacy of markers. Another measure that has been used to test markers is the number of monophyletic species resolved in a gene tree. However, limited resolution is one of the major issues in pant barcoding which also frequently leads to the appearance of non-monophyletic species in respective gene trees [Fazekas 2009]. Additionally, the occurrence of paraphyletic plant species is expected to be quite common [Rieseberg 1994] which makes monophyly unlikely to be an effective tool for species identification. The presence of the so called barcoding gap – low intraspecific and considerably higher interspecific variation - characterizes a useful barcoding marker [Meyer 2005]. Factors capable of influencing the barcoding gap and therefore a critical point to consider prior to any marker evaluation are geographical sampling [Bergsten 2012], status of taxonomy and the correct identity of used references.

3. Taxonomic Considerations

Since the early days of Carolus Linnaeus taxonomy is the discipline where different organisms are joined into groups by considering shared characteristics. The most basic group – the species – represents the only natural group. With passing time restrictions related to the imperfection of humankind were lifted by the invention of technologies and allowed observations to be more precise. Eventually DNA was discovered as the molecule that combines information of life history with that of development.

For some time now, genetic information has been used to evaluate classical taxonomy. But data on many groups is still patchy or completely missing. Increasing the resolution of our senses by applying respective techniques (e.g. microscopy, cytology, chemo taxonomy or genetics) inevitably leads to something new: either new evidence supporting the morphological species, evidence discovering an additional species or evidence suggesting the fusion of separate ones. Usually these determinations are based on markers that are highly invariable within the respective group but show sufficient variation in comparison to related groups.

> *“Thus it can be stated, as a general principle, that individual characters do not have a fixed, a priori importance; a character is only as important as it proves to be in each*

> *individual instance in marking a group that has been recognized on the basis of all the available information. The value of character is established a posteriori."*
>
> Cronquist, 1988

According to this general principle we consider genetic information as additional and hypothesize that correlation between genetic and all other information eventually will lead to individual characters – genetic markers – that can be used to place specimens into known taxonomic groups.

Finally if we want to assign specimens to the most basic taxon – the species – and assume that a species is a natural entity formed by respective processes, we need to pinpoint the responsible processes and find a way to detect inflicted change *ergo* differences between species on the genetic level. The most prominent species concept focuses on sexually reproducing organisms. In contrast to asexually reproducing organisms where individual genomes are always isolated and mutations are the only source of variation, in sexual organisms meiosis introduces stability (repair) and variation by allowing gene flow / exchange of information between the two genomes before gametes are formed. Thus interbreeding has a stabilizing effect on genomic integrity while the absence of interbreeding allows integrity loss to develop expressed by incompatibilities or the hybrid syndrome.

> *A species is a group of individuals fully fertile inter se, but barred from interbreeding with other similar groups by its physiological properties (producing either incompatibility of parents, or sterility of the hybrids, or both)*
>
> Lotsy modified by Dobzhansky, 1935

Later the second part was simplified to "reproductively isolated" and the term "natural populations" was introduced [Mayr 1940], constituting the biological species concept. Although an adequate concept for sexually reproducing organisms and closely related to the genetic criterion which we want to use to identify species, we have to consider at least one more operational species concept. The identity of reference material is exclusively determined using morphological characters.

> *Species are the smallest groups that are consistently and persistently distinct, and distinguishable by ordinary means.*
>
> Cronquist, 1978

The strong dependence on personal judgment and declining experience has been used as arguments for the transition to a DNA based species identification approach [Hebert 2003]. The frequent occurrence of other modes of reproduction (i.e. asexual) in the plant kingdom would suggest including another suitable operational species concept like the agamospecies concept (syn. microspecies, paraspecies, pseudospecies and semispecies). However, since respective cases are not represented in this study we will skip that for now.

The fact that DNA Barcoding and other sequence based approaches are built on existing taxonomic information and have shown to be rather successful together with the fact that commercial products containing biological material usually are designed based on the most common taxonomic information suggests developing methods for genetic food diagnostics on the same principles.

To conclude we keep in mind that genetic analysis has the potential to increase our ability to recognize groups in more details as long as they are genetically distinct. It also facilitates the discovery of evolutionary significance within these groups and thus broadens our understanding of biological diversity and its conservation for future generations. Using existing information to face new challenges has been the way of life for millions of years. Thus, it will most likely be reasonable to apply it in the field of genetic food diagnostics.

4. Limitations of Barcoding Markers in Higher Plants

Even before the proposal of using rbcL and matK as primary barcoding markers it was clear that both do not reach the discriminatory power of the well-established animal barcoding marker COI. While the latter exhibits divergence rates of more than 10 % between species, plastid genome markers like rbcL and matK merely reach 1 % [Kress 2005] and 2 %, respectively. Species identification efficacy values of less than 70 % are thus not uncommon in plants using the proposed standard barcoding markers. Other factors likely to lead to decreased identification success rates within certain plant taxa are hybridization [Zhu 2009], the role of polyploidy in speciation, the breeding system, very recent speciation, introgression and taxonomic bias [Hollingsworth 2011].

While also prone to the above mentioned factors, using other markers (e.g. ITS, psbA-trnH or newly developed) has been shown to increase the success of species identification considerably. Criteria for selecting DNA Barcoding markers (e.g. universal application) are only in part relevant to genetic food diagnostics. The consideration of alternatives which might involve group specific

adaptations seems necessary to identify respective groups with the highest possible success. Especially for extremely poisonous plants the effort of discovering new markers appears to be worthwhile.

5. Reference Plants

To establish authentication of products made from exotic plant species one is not only confronted with new taxonomic implications and limitations of molecular markers but also with a limited availability of reference material. For example screening 79 seed catalogues received from European botanical gardens in 2012 we only found accessions for 10 species of about 70 described of the genus *Dracocephalum*. For four of them a wild origin was known while all others were of unknown origin cultivated in the respective gardens. The most frequently offered species were those which are either naturally occurring in Europe or have been cultivated for some time while for most of the other species only a single accession was available. Other aspects lowering the value of reference material are the long time cultivation in botanical gardens (i.e. increased potential for hybridization events [Ye 2006, Aplin 2007], artificial selection and genetic bottleneck effects) and lack of information on geographic origin of the natural population that prohibits the necessary assessment of sufficient geographical sampling.

C. Case Study: *Dracocephalum* L.

The genus *Dracocephalum* combines several characteristics that make it an ideal model genus for our purposes. As representative of a family that contains many plants used in human cuisine (e.g. basil, rosemary, lavender, marjoram, thyme and peppermint) and traditional medicine (*Agastache*, *Ajuga*, *Dracocephalum*, *Lavandula*, *Leonurus*, *Melissa*, *Mentha*, *Ocimum*, *Rosmarinus*, *Salvia*, *Scutellaria*, *Stachys*, *Thymus*) – the *Lamiaceae* – it gives us the opportunity to explore a genetically rather unknown genus that on the other hand recently has been started to be explored for chemical constituents and pharmacological activities (Table 1 and Table 2). Other welcome challenges of the family are frequently occurring hybrids, e.g. *Sideritis* [Nuñez 1990], *Mentha* [Gobert 2002], *Prunella* [Fritsche 2000], *Phlomis* [Albaladejo 2007] and *Ipomopsis* [Campbell 2004], sister species, e.g. *Mentha* [Bunsawat 2004], *Monarda* [Prather 2002], *Otostegia* [Scheen 2009] and *Salvia* [Jenks 2011], and so called chemovars, e.g. *Thymus* [Alonso 1984], *Ocimum* [Yavari 2011] and *Perilla* [Meng 2009]. If discovered such cases will introduce new opportunities to develop and test genetic markers. *Dracocephalum* mostly distributed outside Europe represents many other genera used in traditional medicines for which reference plants are hard to obtain

thus forcing us to evaluate strategies to overcome these limitations. With one species already commercially exploited (i.e. *D. moldavica* tea and essential oils) we can evaluate genetic markers not only using reference material but also include commercial products. Finally at least two species of the genus present in Europe are of conservational interest (*D. ruyschiana* and *D. austriacum*). Informations about genetic markers resulting from this study could be used to evaluate different populations in more detail in the future.

1. Abstract

Dracocephalum moldavica L. ('Moldavian Dragonhead') is one of the few species of Dragonhead that is well known in Europe for its use as ornamental, spice and bee plant. More recently it has been re-discovered as a tea product. Due to its strong aromatic resemblance of 'Lemon Balm' (*Melissa officinalis* L.) and 'Catnip' (*Nepeta cataria* L.) as well as for the overwhelming morphological diversity within the affinity of the *Lamiaceae* (Labiatae) family adulterations are likely to occur.

DNA based diagnostics circumvents morphological variation and evades the need for specimens taken within a specific life stage. As proof of principal we demonstrate the diagnostic potential of PCR-RFLP and multiplex PCR to distinguishing between samples containing DNA from *Dracocephalum*, *Melissa* and *Nepeta*.

In the second part we investigate an increased sample of *Dracocephalum* accessions and species as the genus contains several species with medicinal potential. By combining information of morphology, DNA sequences and fingerprinting patterns we evaluate their identity. Finally we assess the diagnostic potential of four barcoding markers by comparison of intra- and interspecific genetic similarity using TaxonGap and by conducting algorithmic identification efficacy tests and sliding window analysis of diagnostic nucleotides using SPIDER to find the most useful marker for species level authentication.

Due to the increased genetic variation within psbA-trnH and ITS compared to standard barcoding markers rbcL and matK we conclude that the former markers are more useful for genetic diagnostics within the genus *Dracocephalum* than the proposed barcoding markers. While PCR-RFLP and ARMS based on rbcLa was used to differentiate between different genera, an ARMS primer based on psbA-trnH was successfully applied to differentiate between samples of *D. moldavica* and congenerics. Seed characters showed to be of moderate use verifying reference material. However, improving statistical analysis by improving the quality of the data set and using

clustering approaches might increase the resolution of seed characters in intra generic taxonomic identifications. Tubulin based Polymorphism (TBP) displayed useful patterns which in most cases appeared to be species specific. Additionally, we found two distinct TBP patterns among *D. ruyschiana* accessions appearing in two different psbA-trnH haplotypes. Here improvement of the analysis by using more sensitive separation and detection methods are recommended.

2. *Dracocephalum* L.

Dracocephalum L., a genus of the family Lamiaceae consists of about 70 species ['The Plant List', Version 1, 2013] predominantly found in alpine and semidry regions in temperate Asia. While the Flora of China [Zheng-Yi 1994] and the Flora of Russia [FOR, Vol. 20, 1968] describe 35 species, in North American (Flora of Missouri) and European (Rothmaler Exkursionsflora; Schmeil Fitschen 90. Aufl.) botanical literature only few of these species are described.

D. ruyschiana is distributed from southern Siberia (Lake Baikal) to the Scandinavian Peninsula (Global Biodiversity Information Facility) and Eastern Europe (Ukraine, Belarus) with southwestern most range on the Pester Plateau in south-west Serbia [Lazarević 2009] and *D. austriacum* from the eastern Pyrenees, across France, Italy, Switzerland, Austria, the Czech Republic (northern limit of the distribution range), the Slovak Republic, Hungary, Romania, Ukraine, and the Caucasus [Dostálek 2010]. Both are members of the section *Ruyschiana* [Budantsev 1987] and are considered to be endangered or at least of high conservational interest in Europe [Holub 2000, Council of European Communities 1992].

Members of the section Dracocephalum are native to Asia but are sometimes cultivated elsewhere. In Europe *D. rupestre* and *D. grandiflorum* are used as ornamental plants. *D. moldavica* and *D. foetidum* are two closely related annual herbs distributed over East Asia to Siberia [Měsíček 1973]. In the second half of the 16th century, *D. moldavica* was introduced into Europe and cultivated as ornamental, spice, medicinal and bee plant [Mansfeld 2001, Adam 2012]. In the late 20th century cultivation started in Eastern Europe. Because of its short vegetation period and adaptation to moderate climate it is used as substitute for *Melissa officinalis* L. – 'Lemon Balm' [Holm 1988]. Morphological and aromatic similarities can lead to adulteration of respective products. Joharchi et al. [2012] reported adulteration by *D. moldavica* (Iran: German Madwort, Badershoo) in Badranjbuyeh (aerial parts of *M. officinalis*) and Balangu-e-shirazi (seed of *Lallemantia iberica* Fisch. & C.A.Mey. or L. royleana Benth.). Nowadays *D. moldavica* is used in herbal teas with a strong aromatic resemblance of 'Lemon Balm'. For that it received its vernacular

names Moldavian Balm and Türkische Melisse. The main component of the essential oil responsible for the flavor of *D. moldavica* [Holm 1988, El-Baky 2008, Nikitina 2008, Shuge 2009] and *M. officinalis* [Carnat 1998, Taherpour 2011] is citral (geranial and neral).

Table 1: Activities and (medicinal) relevance of *D. moldavica* (* = dissertation)

Author(s), Year	Plant Source	Activity	(Medical) Relevance
Keyvan Dastmalchi, 2007, 2008*	Iran	Antioxidant	Alzheimer's Disease
Khazaeli et al., 2008	Iran	Antioxidant	UV protection
Najafi et al.,2008	Iran	Cardio protective	Reperfusion injury prevention
Martinez-Vazquez et al., 2012	Mexico	Sedative	CNS inhibition
Povilaityte et al., 2000 & 2001	Lithuania	Antioxidant	Food conservation

Table 2: Studies on components (Vol = Volatiles, Flav = Flavonoids) and indicated activity of other Dragonhead species. Geography indicates the place of study and if not indicated by "unclear" is also the origin of the plant specimen. TM indicates the mentioned use in a traditional medicinal system.

***1: Lee et al. 2007 referenced Shatar et al. 2000 stating *D. foetidum* to be one of the most popular essential oil-bearing plants. Shatar et al. 2000 however did not investigate *D. foetidum* but *D. moldavica*. Although plants of *D. foetidum* and *D. moldavica* are very similar they have been accepted as two distinct species.**

***2: References to *Dracocephalum* "chemotypes". * unclear in the Geography-column means there is no clear indication where the plant originated from.**

***3: Analyzed seed oil**

Species	Authors	Vol	Flav	Effects/Activity	Geography	TM
D. foetidum	Lee 2007[*1]	X		Anti-microbial	Mongolia	X
D. nutans	Misra 1987	X			India, Kashmir	
	Misra 1992[*2]	X			India, Kashmir	
D. argunense	Kim 2006			Anti-inflammatory	Korea, South	X
D. rupestre	Han 2008, 2012		X	Anti-oxidant	China, unclear	X
	Ren 2008		X		China, Shanxi	X
	Lou 2012		X	Anti-oxidant	China, unclear	
D. ruyschiana	Selenge 2013		X	Anti-oxidant	Mongolia	X
D. parviflorum	Coxworth 1965[*3]				Canada	
D. heterophyllum	Wang 2012		X		China, Qinghai	X
	Numonov 2013		X		Russia, unclear	
D. kotschyi	Faham 2008			apoptosis	Iran, Isfahan	X
D. multicaule	Oganesyan 2009		X		Armenia, unclear	

IV. Materials & Methods

A. Reference Plant Material

1. PCR: Effects of DNA Fragmentation

For DNA Extraction, the fragmentation assay and subsequent evaluation of barcoding marker performance one accession of *Rheum officinale* (ID: 1473) and one of *Iris domestica* (ID: 5939) was used. Both plants were cultivated and are maintained as living specimens in the botanical garden of the KIT.

2. PCR-RFLP: The 'Lemon Myrtle' Case

One specimen of each 'Lemon Myrtle' species was purchased from a commercial source (Rühlemanns, Horstedt, Germany), and their identity verified by the morphology of flowers and leaves based on taxonomic literature [Walsh 1996]. They are maintained as living specimens with ID 10 (Backhousia citriodora F.Muell) and 7596 (Leptospermum citratum (ex. J.F.Bailey & C.T.White) Challinor, Cheel & A.R.Penfold). Four commercial tea samples (Table 6) were used for validation of the assay and partially sorted and identified based on microscopic criteria developed in the respective publication (p. 1) prior to molecular analysis.

3. SCAR Development: The *Clematis* Complex

For the RAPD analysis and subsequent SCAR development the following 8 plant accessions were used: *Aristolochia longgangensis* (ID: 2514, received as *A. manshuriensis*, a1), *Akebia trifoliate* (ID: 2506, s1), *Akebia quinata* (ID: 2081/5498, s3), *Clematis tangutica* (ID: 2513, s9), *Clematis apiifolia* (ID: 2504/2611, t2), *Clematis heracleifolia* (ID: 2961/5590, t3), *Clematis integrifolia* (ID: 2962/4790, t5) and *Aristolochia sempervirens* (ID: 2549/NA, t12). All plants were cultivated and are (with the exception of *Aristolochia sempervirens*) maintained as living specimen or seeds in the botanical garden of the KIT.

4. Case Study: *Dracocephalum* L.

a) Product Authentication using PCR-RFLP and ARMS

For the product authentication by ARMS and PCR-RFLP two accessions of *D. moldavica* (IDs: 5862, 5863) were ordered from the Leibniz-Institut für Pflanzengenetik und Kulturpflanzen Forschung (IPK). Another accession of *D. moldavica* (ID: 5861) was ordered from a commercial vendor

(Rühlemanns, Horstedt, Germany). Accessions of *D. ruyschiana* (ID: 5156), *Melissa officinalis* (ID: 4639), *Nepeta cataria* (ID: 7575) and *N. nuda* (ID: 5395) were already available and are maintained as seeds or living specimens at the botanical garden of the KIT. The identity of the plants was confirmed by morphological determination prior to the molecular work.

b) Marker Evaluation and Species Level Diagnostics

For the *Dracocephalum* case study we screened 79 seed catalogues (Indices Seminum) received from European botanical gardens in 2012 and the database of the Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung (IPK). The search of the database yielded accessions of 7 *Dracocephalum* species while the index seminum search resulted in 9 species of which 7 were not found at IPK. In summary there were 53 accessions of 14 *Dracocephalum* species available and one accession of an unknown species. The most frequently offered species were those which are either naturally occurring (15 accessions of *D. ruyschiana*) or have been cultivated in Europe for some time (15 accessions of *D. moldavica*, 5 of *D. grandiflorum* and 2 of *D. rupestre*). For most of the species however only one or two accessions were available. The majority of seeds were derived from ex-situ cultivated plants with unknown wild origin. Seeds of four accessions were collected from natural habitats (*D.austriacum*, *D.parviflorum* and *D.ruyschiana*). A detailed list of all these accessions can be found in Table 5.

All accessions were cultivated and are maintained as seed or living specimens in the botanical garden of the KIT. The identity of the plants was evaluated and is described in detail below (see page 30).

B. Basic Studies

1. Interactive Plant Reference Database

To improve organizational issues of the plant collection harbored in the botanical garden of the KIT, the collection was transferred from an excel list to an online database management system. The objectives were firstly to develop a robust management system which increases data consistency. Secondly the system also should function as a platform serving current scientific information on the respective plant species by using web based repositories accessible via application programming interface (API) or other means. Finally, the design of the system should include the possibility to store generated data from local projects and preserve the history of cultivation within the garden. Since the botanical garden is a member of the international plant

exchange network (IPEN) the benefits of these improvements should greatly increase quality of data connected to accessions that will be introduced into the network.

A linux based operating system was chosen (Debian 2.6) and complemented with web server software (Zend Server CE 5.6 based on Apache 2.2, PHP 5.3 and MySQL). The backbone which provides the respective information in form of HTML pages was built on Zend Framework V. 1.11, an open source framework for developing web applications and services using PHP. The frontend is supported by JavaScript through integration of the Dojo Toolkit (V. 1.72), an open source modular JavaScript toolkit designed to ease the rapid development of cross-platform, JavaScript / Ajax-based applications and web sites.

The main sources for external data integration are firstly 'The Plant List' (TPL, Version 1.1. Published on the Internet; http://www.theplantlist.org/), a widely accessible working list of known plant species which has been developed and disseminated as a direct response to the Global Strategy for Plant Conservation, adopted in 2002 by the 193 governments who are Parties to the Convention on Biological Diversity. The status of the species name of each of the local accessions is denoted according to TPL (i.e. Synonym, Accepted, Unresolved and Unknown). Additionally, all known synonyms of the respective species are accessible. Secondly we chose the Encyclopedia of Life [Wilson 2003] as data source for species images, descriptions, distributional data, synonyms and Wikipedia articles. Secondary data sources are the National Center for Biotechnology Information serving information regarding publications and molecular data and Barcoding of Life Data Systems (BOLD) serving information regarding barcoding data. More recently we added the 'Global Compendium of Weeds' (GCW) to our data sources. For each local accession the occurrence of the respective species name in the GCW list is checked. If found it is indicated by a direct link to the GCW page.

2. PCR: Effects of DNA Fragmentation

DNA was extracted from young leaves with a modified cetyl trimethyl ammonium bromide (CTAB) protocol using 2 % CTAB buffer (2 % w / v CTAB, 150 mM Tris, 30 mM EDTA, 1.4 M NaCl, 2 % v / v mercaptoethanol, 4 % w / v polyvinyl pyrolidone). Quality and quantity of the DNA was evaluated using a spectrophotometer (ND1000, Thermo Scientific) and by agarose gel electrophoresis.

Extracted DNA was diluted to a final concentration of 150 ng / µl (*Iris*) and 250 ng / µl (*Rheum*) respectively. The extraction was then divided into 8 samples and 1.5 µg DNA of each was analyzed in a 1 % (w / v) agarose gel to document the non-fragmented situation. Subsequently 7 of the

samples were incubated in an ultrasonic sound water bath (Sonorex, RK52, Bandelin Electronic, Berlin, Germany) for 6, 30, 60 seconds, 5, 10, 50 and 100 minutes respectively (2, 5, 10 and 20 minutes for *Rheum*). Temperature of the water was recorded after each treatment period. After fragmentation again 1.5 µg of DNA was analyzed on a 1% (w / v) agarose gel to document the results of the fragmentation procedure.

To test the effect of fragmentation on PCR amplification two different markers were selected. Firstly we used the official DNA Barcoding plastid marker rbcLa (~600 bp fragment) which performs very well in most land plants and secondly a marker that amplifies a considerably smaller fragment – ITS2 (~300 bp) – of the nuclear genome but whose performance in *Rheum* and *Iris* was unknown. PCR reactions were performed in 10 µl volume containing 7.8 µl ddH_2O, 1 µl reaction buffer (NEB, 2 mM MgCl), 200 µM dNTPs, 0.2 µM of each primer, 25 ng template DNA and 0.5 Units of polymerase (NEB Taq). For further details on primer sequences please refer to (Table 10). PCR products (5 µl) were separated for 20 minutes at 100 V in a 1.5 % (w / v) agarose gel.

Since all samples resulted in good amplification of the rbcLa marker a template dilution series was analyzed to find the threshold concentration of the most fragmented sample (100 minutes treatment). PCR reactions were repeated with template concentrations of 500 pg / µl and 50 pg / µl and separated by gel electrophoresis with previously described settings.

3. PCR-RFLP: The 'Lemon Myrtle' Case

As briefly mentioned in the introduction the term 'Lemon Myrtle' is ambiguously used for at least two different species. Thus our goal was to design a simple genetic assay to distinguish between these two species.

Extraction of genomic DNA: Fresh leaf material (preferably young leaves) was harvested from healthy plants. About 40 mg of the sample were transferred into a reaction tube (2 ml, Eppendorf) together with one stainless steel bead (diameter 5 mm), and shock frozen in liquid nitrogen. The frozen sample was then ground three times for 15 s at 20 Hz (Tissuelyser, Qiagen, Hildesheim, Germany). After each individual grinding step, the sample was returned into liquid nitrogen to ensure that the powder did not thaw during the extraction. Genomic DNA was extracted using a modified extraction method based on CTAB (Doyle 1987) using about 25 mg ground and frozen leaf tissue. The powder was complemented with 1 ml pre-warmed extraction buffer (3% w / v CTAB) containing 8 µl / ml mercaptoethanol, and incubated for 30 minutes at 55°C followed by a centrifugation to remove debris. Subsequently, the sample was digested with proteinase K (55°C,

30 min), and then mixed with 750 µl of chloroform:isoamylalcohol (24:1) and then spun down for 10 min (14000 g, 25°C), the aequous upper phase (containing the DNA) was transferred into a fresh reaction tube, the DNA was precipitated with 0.65 volumes of isopropanol, collected by centrifugation (10 min, 14000 g), washed with 70 % EtOH and dissolved in 50 µl ddH2O. The concentration of the eluted DNA was determined photometrically (NanoDrop ND-100, peqlab). The E260 / E280 of the extracted DNA was between 1.7 and 2.1. The quality of the DNA-extracts was controlled by electrophoresis on a 1.5-% agarose gel supplemented with 5% v / v of the fluorescent dye SYBR Safe (Invitrogen).

PCR-amplification and restriction digest of rbcL: A partial sequence of the large subunit of the ribulose-bisphosphate carboxylase gene (rbcL), rbcLa, was amplified by PCR in a 10 µl-reaction using 50 ng of genomic DNA as template and a reaction mix containing single-strength buffer (thermopol, NEB), 200 µM mixed dNTPs (NEB), 200 nM of each primer (rbcLa-f / rbcLa-rev, Invitrogen), 0.4 units Taq polymerase (NEB), and 10 mg / ml bovine serum albumine (Sigma-Aldrich, Deisenhofen, Germany). The amplified fragments were separated by electrophoresis in a 1.5 % agarose gel and their size verified using a 100-bp DNA ladder (NEB) after fluorescent staining with SYBR Safe (Invitrogen). Subsequently fragments were extracted from the gel using the NucleoSpin® Extract II kit (Machery-Nagel, Karlsruhe) following the protocol of the producer, and then sequenced (GATC Biotech, Konstanz). The sequences were verified by BLAST search, and alignment with related rbcL-sequences (ClustalX, http://www.clustal.org), and are deposited in GenBank under the accession numbers JN676919 (*Leptospermum citratum*), and JN676920 (*Backhousia citriodora*). To discriminate the two species, 6 µl of the rbcLa PCR reaction were digested overnight at 37° C in a 25 µl reaction volume consisting of 2.5 µl 10-fold enzyme buffer (NEB, No. 4), 2.0 µl SacII enzyme (NEB), and 14.5 µl bidestilled water. The digested amplificates were separated by electrophoresis in a 1.5 % agarose gel along with a 100-bp DNA ladder as size marker (NEB).

4. SCAR Development: The *Clematis* Complex

Eleven RAPD primers (50a-50k) were used to generate respective DNA fingerprints of all 8 plant accessions. The PCR reactions had a volume of 10 µl containing 5.95 µl ddH_2O, 1 µl reaction buffer (NEB, 2 mM MgCl), 200 µM dNTPs, 1.25 µM primer, 25 ng template DNA and 0.5 Units of polymerase (NEB Taq). The full volume of the PCR product was separated for 30 minutes at 100 V in a 1.5 % (w / v) agarose gel. Fragments specific for accession 2514 (a1) using primers 50e, 50f, 50h and 50i were extracted from gel using a commercial kit (Macherey Nagel, Düren, Germany)

and subsequently used as templates in another RAPD PCR. After verifying the sole amplification of the selected fragments by gel electrophoresis with previously described settings the PCR products were purified using the mentioned kit and sequenced (GATC, Konstanz, Germany). The sequence derived from fragment 50e of accession a1 (50e-a1) was used to design a primer pair that would amplify a 629 bp fragment. Primers were synthesized by Invitrogen and used in a 10 µl PCR containing 7.8 µl ddH_2O, 1 µl reaction buffer (NEB, 2 mM MgCl), 200 µM dNTPs, 0.2 µM of each primer, 25 ng template DNA and 0.5 Units of polymerase (NEB Taq). 2 µl of the resulting PCR product was separated for 30 minutes at 100 V on a 1.5 % (w / v) agarose gel. For detailed information on primer sequences please refer to Table 10.

C. Case Study: *Dracocephalum* L.

1. Product Authentication using PCR-RFLP and ARMS

In 2010 a diploma thesis on Dragonhead diagnostics was finished describing morphological traits accessible in processed material. In addition to that we wanted to design a DNA based authentication approach. The basic idea was to use a universal primer pair as PCR positive control and a taxon specific ARMS primer to detect the presence of DNA from Dragonhead. Additionally, to detect putative adulterants from *Melissa* and *Nepeta* an approach based on differential restriction digestion patterns should be employed.

Using sequence data from the respective *Dracocephalum*, *Melissa* and *Nepeta* accessions (see page 24) a restriction site analysis was conducted. BamHI was chosen for digestion of the rbcLa fragment in *Melissa* and *Nepeta* and EcoRI for digestion of the rbcLa fragment in *Dracocephalum* and *Melissa*. The BamHI digestion indicates the presence of *Melissa* or *Nepeta* while EcoRI can be used to differentiate between *Melissa* and *Nepeta*. The predicted restriction patterns are displayed in Figure 5 A and C on page 42.

RbcLa fragments of respective accessions were amplified in 10 µl PCR reactions containing 5.95 µl ddH_2O, 1 µl reaction buffer (NEB, 2 mM MgCl), 200 µM dNTPs, 1.25 µM primer, 25 ng template DNA and 0.5 Units of polymerase (NEB Taq). 6 µl of the reaction volume subsequently was digested overnight at 37 °C in a 25 µl reaction volume consisting of 2.5 µl 10X enzyme buffer (NEB, no. 1), 2.0 µl BamHI or EcoRI enzyme (NEB) respectively, and 14.5 µl bidistilled water. The result of the digestion was separated by electrophoresis using a 1.5 % (w / v) agarose gel along with a 100-bp DNA ladder (NEB) as size marker.

The sequence data also was used to design an ARMS primer (DC4, Table 10). The primer was designed to anneal with its 3'-end at position 462 of the rbcLa fragment where a thymine is present in *Dracocephalum* and which is replaced by a guanine in *Melissa* and *Nepeta*. Additionally, a destabilization of the 3'-end (3rd position G to A) was introduced to increase differentiation power of the primer (for details on primer design refer to Figure 6 A on page 42). Together with the universal rbcLa primer pair the diagnostic primer DC4 was used in a multiplex PCR. Resulting fragments were separated by electrophoresis in a 1.5 % agarose gel along with a 100-bp DNA ladder (NEB) as size marker.

2. Taxonomic Verification

a) Morphological Determination

Plants that successfully developed generative organs were morphologically determined using one of the available determination keys (see Table 3, p. 43). Available morphological traits (i.e. leaf shape and size, habitus) of the remaining accessions were compared to descriptions and images found in respective literature and other online resources (e.g. Encyclopedia of Life, Biodiversity Heritage Library, Wikipedia).

b) Seed Morphology based Determination

Seeds of an accession were distributed over a white surface that had a scale attached to it. Images were captured using a digital camera (Canon EOS 550D) and analyzed using the program SmartGrain [Tanabata 2012] in the following way:

- The internal scale (pixel / mm) was set by measuring one centimeter of the captured scale
- Seed detection intensity was set to high
- Color segmentation method was used
- Accuracy of segmentation process was set to 1 (highest)
- Threshold for awn removal was deactivated
- Seed and Background color where determined using one or two representative seeds
- The area where the seeds were distributed on was selected and analyzed
- Detected seeds where checked and obvious errors (seed fragments, detection errors) removed
- Finally the measurements were exported

The measurement files were merged and edited with MS Excel. Additional columns for accession ID and taxonomic name (one for the putative name and one containing the corrections derived from the morphological determination) were added. Finally the data was exported in tab-delimited format for analysis with R (V. 3.0.2) using RStudio (V. 0.98.484). Data of seed length-with-ratio, circularity, area and perimeter length for the complete data set using pooled seed data of a species and for subsets of taxa with more than one accession available (*D. argunense* and *D. ruyschiana* from section Ruyschiana, *D. grandiflorum* and *D. rupestre*, and *D. moldavica*) were generated. The data was sorted by descending mean value and plotted in the form of horizontal boxplots. The height of each boxplot was set to be proportional to the square-root of the number of observations for the respective group. Correlation graphs including robust regression lines of circularity and length-width ratio were generated using subsets of accessions from *D. ruyschiana* and *D. argunense*.

c) Tubulin based Polymorphism

Tubulin based Polymorphism (TBP) is a DNA fingerprinting technique based on intron length variation in the subtypes of β-tubulin. Results can be used to assess relationships between samples by pattern similarity. We wanted to know if TBP patterns of our accessions correlate with the determined morphological species and evaluate the usefulness of TBP as species identification tool.

Samples of representative *Dracocephalum* species were used as template in 10 µl PCR reactions containing ddH_2O, single-strength reaction buffer, 2 mM MgCl, 200 µM dNTPs, 0.2 µM of each primer, 150 ng template DNA and 0.1 Units of NEB Taq polymerase. For details on primer sequences see Table 10 (p. 91). The complete reaction volume was subsequently separated for 30 minutes at 100 V in a 1.5 % (w / v) agarose gel.

d) Genetic Distance Analysis

Sequence data of the respective markers was aligned using Muscle in MEGA5 [Edgar 2004, Tamura 2011]. Missing data was indicated by using the question mark symbol and primer sequences were removed. For each species (corrected according to morphological determination and seed data) and marker, unique sequences were determined. Sequence data of species that were represented by only one accession (singletons) was blasted (NCBI blast) and representatives of each genus among the matches were included in a dataset to create a Multidimensional Scaling (MDS) plot with R (V. 3.0.2, R Core Team 2013) using RStudio (Version 0.98.484). Two of these representatives

which according to the MDS analysis were closest to the *Dracocephalum* group were included in the final alignment together with all unique *Dracocephalum* sequences, all sequences isolated from accessions whose classification was challenged, sequence data from our *Nepeta*, *Melissa* and *Lallemantia* accessions (if available) and external sequence data as putative references to our accession (see Table 8, p. 89 and Table 9, p. 90 for details on included sequences).

For each marker the evolutionary history was inferred using the Neighbor-Joining method [Saitou 1987]. Clusters were evaluated with a bootstrap test [Felsenstein 1985] using 1000 replicates. If a cluster was supported by 50 % or more it was indicated next to the respective node. Trees were drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The evolutionary distances were computed using the Kimura 2-parameter method [Kimura 1980] and are in the units of the number of base differences per site. Ambiguous positions were removed for each sequence pair (pairwise deletion).

The resulting trees were used to determine if predetermined and corrected classifications were represented by respective clusters in the tree.

3. Marker Evaluation

The following marker evaluations are based on the classifications derived after morphological determination, seed character and TBP pattern comparison and tree based analysis (Table 3, p. 43). To be able to establish species specific diagnostics, available markers need to be evaluated within the taxonomic affinity of the respective group. ARMS and PCR-RFLP depend on single or very few nucleotide differences that – from the diagnostic point of view - are supposed to be group specific – e.g. species specific. The species specificity within a certain range can only be achieved when sufficient correct identified accessions of the respective range are included into a marker evaluation.

We analyzed parts of the plastid markers rbcL (rbcLa), matK (matK-KIM) and psbA-trnH as well as the nuclear ITS region to assess intra- and interspecific variation of our samples. We chose to use the software TaxonGap as it produces a graphical representation of the heterogeneity and separability of different markers in comparison. We then chose the regions with the highest potential for being species specific (psbA-trnH and ITS) and used the R package SPIDER to conduct a distance based identification efficacy evaluation. Finally we applied a sliding window analysis of diagnostic nucleotides found in these regions and designed a specific ARMS marker based on these results to detect DNA from *D. moldavica*.

a) Preparation of Genomic DNA

DNA was extracted from young leaves with a modified CTAB protocol [Doyle 1987] using 2 % CTAB buffer (2 % w / v CTAB, 150 mM Tris, 30 mM EDTA, 1.4 M NaCl, 2 % v / v mercaptoethanol, 4 % w / v PVP). Quality and quantity of the DNA was evaluated using a spectrophotometer (ND1000, Thermo Scientific) and by electrophoresis in a 1 % agarose gel.

b) Amplification and Sequencing

All four barcoding markers were each amplified in a reaction volume of 30 µl containing ddH_2O, 1 X reaction buffer, 2 mM MgCl, 200 µM dNTPs, 0.2 µM of each primer, 150 ng template DNA and 0.1 Units of NEB Taq polymerase. For details on primer sequences please refer to Table 10. Successful amplification and size of product was verified by agarose gel electrophoresis (1.5 %, 25 minutes @ 100 V). Amplified fragments were purified using a silica-membrane technology based commercial kit (Macherey Nagel, Düren). In case several fragments were detected the correct-sized was extracted from agarose gel and purified. Sequencing reactions were prepared according to product specifications (LightRun96) and sequenced (GATC, Konstanz, Germany).

Trace files were examined to assess the quality of data using FinchTV 1.4.0 (Geospiza Inc.). Only data without significant ambiguity in the core region of the sequence was processed further. Each bidirectional data set was merged replacing "N"s and missing data in the terminal regions with complementary data using MEGA5 [Tamura 2011]. Sequence and specimen data was stored (private) at barcoding life data systems (BOLD) [Ratnasingham 2007]. Further modifications to the sequence data are elaborated in the respective sections.

c) Heterogeneity and Separability

TaxonGap uses marker similarity matrices of the respective taxa to calculate heterogeneity and separability.

For a given set of OTUs $\{O_1, O_2, ..., O_n\}$, the s-heterogeneity within taxon O_i is defined by

$$max\ x, y \in Oi, x \neq y\ ds(x, y)$$

Herein, $d_s(x, y)$ represents the distance between the (different) members x and y of the taxon O_i as measured from the biomarker s. These distances are presented in a separate distance matrix for each biomarker. Likewise, the s-separability of taxon O_i is defined by

$$min\ x \in Oi, y \notin Oi\ ds(x, y)$$

The taxon containing that member y for which a minimum distance is reached in the computation of the s-separability, is called the closest neighbor of taxon O. Note, however, that the closest

neighbor relationship is not necessarily symmetric: the fact that O_j is the closest neighbor of O_i does not imply that O_i is also the closest neighbor of O_j.

Matrices were prepared using the software MatGAT [Campanella 2003]. The program reads a collection of sequences in fasta format performs a series of pairwise alignments using a global alignment algorithm [Myers 1988] and calculates similarity and identity which can be exported to MS Excel in the form of a matrix. For this analysis all available unique *Dracocephalum* sequences were used to generate a similarity matrix. Further modification of the matrix was necessary to meet TaxonGap compliance: Using MS Excel taxon names were modified and self-similarity (100 %) was added to each taxon. Finally each matrix was exported as tab-delimited text file setting the encoding to UTF-8. In addition to the respective matrices of the marker TaxonGap requires either a tree in newick format or a simple taxon list containing the taxon names included in the analysis. We prepared a taxon list by creating a text file with one taxon name per line. Taxon list and matrices were then used with TaxonGap and output formats were adjusted accordingly before exporting the final graphical representation.

4. Species Level Diagnostics

a) Distance Based Identification Algorithms

Besides tree-based identification approaches several other algorithms are employed to test identification efficacy using sequence data directly. A pool of sequences derived from predetermined members of the taxonomic group of interest represents the barcode database. The algorithms work through the sequences picking one at a time trying to determine the species it belongs to. The most basic algorithms (nearNeighbour) finds the closest sequence (ties are decided by majority rule), compares taxon names and returns TRUE (taxon names match) or FALSE (identification failure). The algorithms bestCloseMatch (bCM) [Meier 2006] and threshID (tID) both use distance thresholds to find the closest neighbor. While bCM only uses the closest sequence(s) within the threshold tID is more inclusive by considering all sequences within the range. If no closest match is found the algorithms return “no id” (identification failure), if more than one species is closest (bCM) or within the threshold (tID) the algorithms return “ambiguous”. In case all matches (tID) or the closest match(es) are from one species the query returns TRUE (identification success).

We analyzed psbA-trnH and ITS marker sequence data as the most promising candidates with all three algorithms using the R package SPIDER (Brown 2012). Thresholds picked for the analysis

were 2 % which is used on BOLD to highlight high intra-specific distance (> 2 %) and low distance to another species (< 2 %) and an optimized threshold of 0.5 % (psbA-trnH) and 0.3 % (ITS) determined by evaluating the cumulative error in species assignment over a range from 0 to 2.5 % (psbA-trnH) and from 0 to 2 % (ITS). For the evaluation the SPIDER function threshOpt was used which determines the positive, negative, false positive and false negative rates of identification accuracy for a given threshold. Applying the function over the defined range the results were plotted to display FALSE positive and FALSE negative identifications. The cumulative error is defined as the sum of FALSE identification results and thus the threshold resulting in the lowest cumulative error represents an optimized threshold.

b) Diagnostic Nucleotides

For the diagnostic analysis the alignment containing unique sequences of *Dracocephalum* was used to generate a DNA sequence fingerprint [Lou 2007] visualizing the variability within respective markers using the online service at http://evol.mcmaster.ca/fingerprint/. Any position within a sequence alignment where all four possible nucleotide states occur is depicted by a black vertical bar. In contrast any position that only contains one state and thus is invariable is depicted by a white vertical bar. Anything in between is colored accordingly using greyscales. Additionally, the alignment was used for a sliding window analysis in R using the package SPIDER (Brown 2012). In particular a function was used to visualize regions with diagnostic nucleotides. The window size of 25 bp was used as it represents the common length of a primer. One plot of all *Dracocephalum* species was combined with a plot of a single species (i.e. *D. moldavica*).

c) Diagnostics of D. moldavica

Based on the described diagnostic nucleotide analysis we designed an ARMS primer targeting position 72 in the psbA-trnH inter-genic spacer region where a guanine is specific for the *Dracocephalum moldavica* group (all accessions of *D. moldavica* and unclassified accession 8142). The combination of psbA-trnH barcoding primers and the newly designed ARMS primer was used in a PCR reaction containing ddH_2O, single-strength reaction buffer, 2 mM MgCl, 200 µM dNTPs, 0.2 µM of each primer, 50 ng of template DNA and 0.1 units of Taq polymerase (NEB). In addition to represenatives of our Dracocephalum species we included two commercial products containing ‘Dragonhead’ (Table 7). The amplified fragments were separated in a 2 % agarose gel along with a 100 bp size marker (NEB).

V. Results

A. Basic Studies

For the maintenance of living specimens and seeds of reference plants an interactive database was established. In our basic studies we addressed limitations that might become apparent when dealing with processed plant material commonly used in commercial products (e.g. DNA fragmentation). We furthermore evaluated basic methods for the design and application of genetic food diagnostics. In the 'Lemon Myrtle' case we used PCR based Restriction Fragment Length Polymorphism (PCR-RFLP) to distinguish two species of different genera that are both called 'Lemon Myrtle' but differ in chemical composition and dose dependent toxicity. In a preliminary approach of the *Clematis* Complex, plants that are used in Traditional Chinese Medicine, we developed primers based on a PCR fragment derived from a DNA fingerprinting analysis (Random Amplified Polymorphic DNA) specific for an adulterant species from the Genus *Aristolochia*, known to contain components with high toxicity potential.

1. Interactive Plant Reference Database

In the beginning of this study information about plant accessions of the botanical garden were managed by one person using an MS Access database. Modifications only were possible through exchange of information based on personal contact or by sharing of excel sheets. By establishing an interactive database accessible through intra- / internet the maintenance of accession data is now possible by any authorized person simultaneously which greatly improves data consistency. Additionally, by integration of external information sources (The Plant List) plant name confusions are avoided by indicating the current status of the taxon name (accepted, synonym, unresolved, unknown). All known synonyms are available and can be included in search for scientific publications. Each modification of accession data is logged and informations regarding the plant can be entered manually in form of comments. For each accession information (e.g. descriptions, molecular data, images) derived from several external sources (Encyclopedia of Life, Biodiversity Heritage Library, Barcoding of Life Data Systems) is provided directly or in form of links. Since invasive plants are a threat to local biodiversity for each species it is indicated if any reports on weed characteristics exist in the online version of the Global Compendium of Weeds (www.hear.org/gcw/). The database can be accessed from within the KIT intranet:

http://172.22.97.24/science/Plants

2. PCR: Effects of DNA Fragmentation

While the presence of particular metabolites within DNA extracts represents the main challenge establishing PCR based diagnostics a solution in most cases can be found by adaptation of the extraction procedure or using tissue with low concentrations of the respective compounds. DNA fragmentation on the other hand poses a more serious problem since amplification of a certain genomic region is only possible if that region is available in full size. In this study we evaluated the effect of DNA fragmentation on the amplification of common barcoding markers.

The temperature measurements taken during the fragmentation process indicated an increase of roughly 0.5 ° C per minute up to 50 minutes followed by 0.3 ° C increase per minute for the remaining 50 minutes of the treatment. Results were visualized by separating 1.5 µg of fragmented DNA in agarose gel. The increase in fragmentation using the specified time points could best be observed using *Iris* samples (Figure 1 A, p. 38). It took between 6 and 30 seconds to destroy most of the genomic DNA in a fashion that no discrete band could be observed anymore. Between 5 and 10 minutes treatment the fragment size seems to reach a limit. This fragment population is reduced by more than half considering the 50 and 100 minutes treatment.

Using a final concentration of 2.5 ng / µl of the fragmented DNA as template in a PCR resulted in good amplification of each of the treatments (Figure 1 B). A slight decrease in concentration of the amplified fragment could be observed using the 100 minutes treatment. Template DNA of the 100 minutes treatment was then used in a dilution series to find a template concentration where amplification clearly is impaired (data not shown). All other samples were adjusted to the determined concentration (25 pg / µl final concentration) and used in another PCR (Figure 1 C). While PCR amplification does not seem to be affected using fragmented DNA of the 6 seconds up to the 5 minutes treatment, after the size of the main population of fragments drops below the rbcLa fragment size (~ 600 bp) amplification is strongly impaired (Figure 1 D).

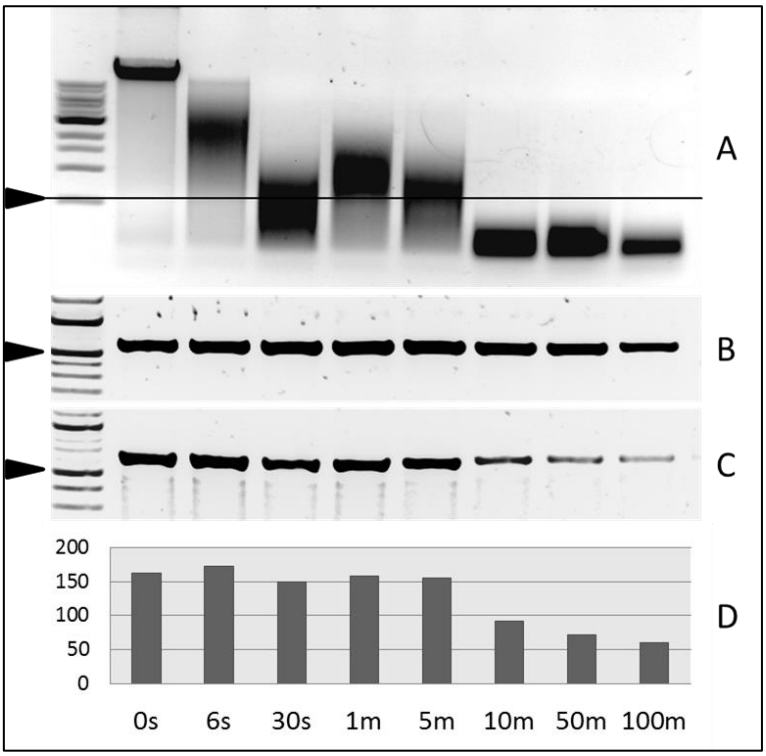


Figure 1: Effects of DNA fragmentation on rbcLa amplification using genomic DNA from *Iris domestica* (L.) Goldblatt & Mabberley. Black indicators on the left of A, B and C mark the fragment size of 500 bp. First lanes on the left of A, B and C are size markers 1 kb (A) and 100 bp (B and C). A: Genomic DNA after fragmentation using ultrasonic sound (US). Treatment duration is indicated below diagram D. B: PCR amplification of barcoding marker rbcLa using 50 ng / µl genomic DNA from the fragmentation series. C: PCR amplification using 500 pg / µl DNA template from the fragmentation series. Significant decrease in amplification can be observed using samples from highly fragmented genomic DNA (> 10 minutes US treatment). D: Image J measurements of the mean grey value of amplified fragment populations. While amplification of the rbcLa fragment keeps steady up to the 5 minutes treatment a significant drop can be observed starting with the 10 minutes treatment.

3. PCR-RFLP: The 'Lemon Myrtle' Case

The 'Lemon Myrtle' case addresses the issue of over simplified labeling. There are at least two very different species known under the vernacular name 'Lemon Myrtle' (*Backhousia citriodora* and *Leptospermum citratum*). Their taxonomic difference (two different genera) and chemical composition demands improvement in labeling of respective products. For that the clear determination of the species used for the product is necessary. By analyzing sequence data of the rbcLa marker derived from accessions of each of the two species we found a difference in restriction sites that allowed us to design a PCR-RFLP assay for the differentiation of the respective species (Figure 2 A, p. 39). This assay was first tested with our reference plants (Figure 2 B) and subsequently applied on commercial products (Figure 2 C). Further details are discussed on page 64 ff.

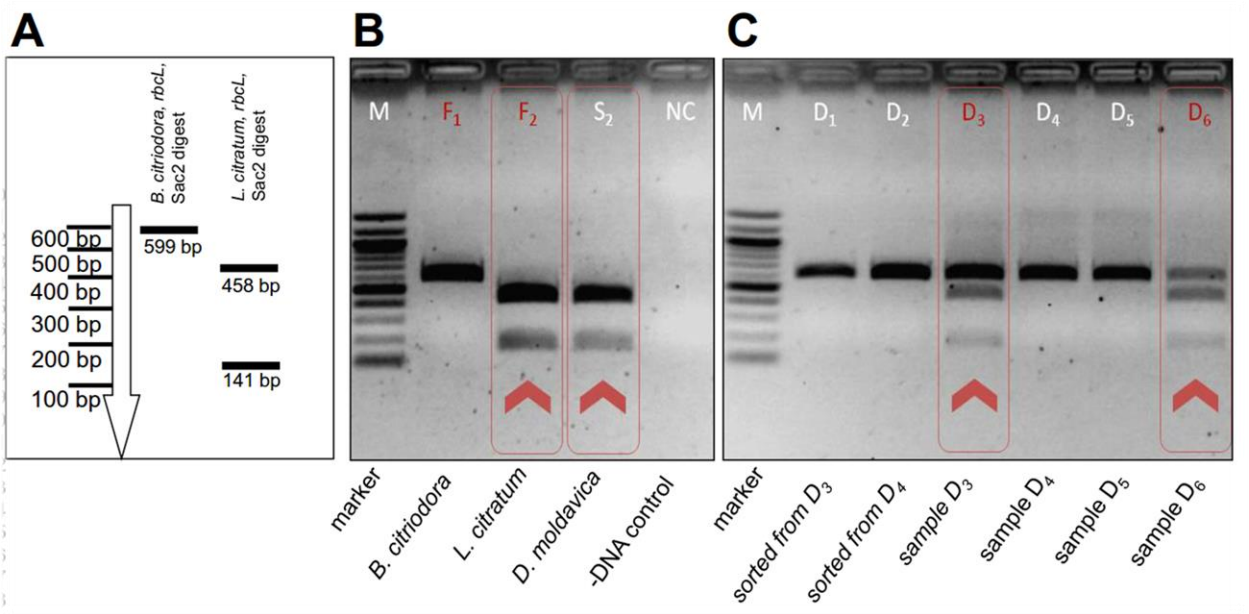


Figure 2: Molecular diagnostics of 'Lemon Myrtle' based on restriction digest of rbcLa amplification product. A: Predicted fragment sizes of digestion with Sac2 in *Backhousia citriodora* (none) versus *Leptospermum citratum*. B: Verification of the prediction using pure leaf samples of *Backhousia citriodora*, *Leptospermum citratum*, and the 'Lemon Myrtle' surrogate *Dracocephalum moldavica*. M: size marker (100 bp DNA ladder), -DNA control test for the specificity of the PCR without added DNA template. Arrows (red) indicate the two smaller fragments resulting from the restriction digest with Sac2. C: Validation of the assay using commercial tea blends containing 'Lemon Myrtle'. Samples D1 and D2 have been sorted as *B. citriodora* from the commercial products D3 and D4, respectively, using the microscopic discrimination given in Table 11. Commercial product D3 contained also *Dracocephalum moldavica*, commercial product D4 did not. Commercial product D5 contained *Backhousia citriodora*, D6 *Leptospermum citratum* as verified by microscopy.

4. SCAR Development: The *Clematis* Complex

Besides PCR-RFLP that is based on commonly used markers (e.g. rbcL, matK, ITS, etc.) DNA fingerprinting methods (e.g. RAPD or AFLP) can be used to develop new markers. To evaluate this approach we chose 11 RAPD primers with 50 % GC-content for a DNA fingerprinting analysis of plants involved in the production and adulteration of Traditional Chinese Medicine chuān mù tōng (Clematidis Armandii Caulis). While one sample of a surrogate species did not produce any pattern (s1) all other samples did. However, most of the patterns appeared very simple and sometimes weakly amplified (e.g. t2 and t5 in Figure 3 A, p. 40). Fragments amplified by 4 different RAPD primers unique for the adulterant species (a1) were isolated by gel extraction and used in another RAPD PCR as templates. The uniquely re-amplified fragment of primer 50e (Figure 3 B) was sequenced. Based on the resulting sequence information a pair of primers was designed to amplify a 629 bp fragment unique for the adulterant accession. The primers were tested with all available samples and showed specificity for the adulterant accession (Figure 3 C).

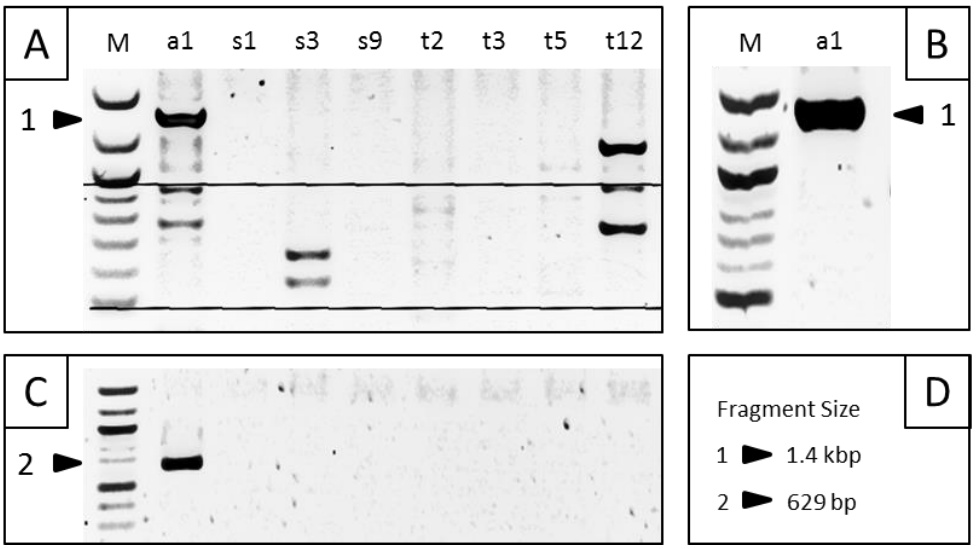


Figure 3: SCAR primer development for a selection of accessions involved in the *Clematis-Aristolochia* complex. A: RAPD pattern produced with primer 50e; Fragment 50e-a1 is the only fragment of the size 1.4 kbp and labeled by a black triangle (1). B: Re-amplification of fragment 50e-a1 for subsequent sequencing. C: SCAR PCR with primer pair S50e-9a1-2. D: Fragment size of the isolated RAPD fragment 50e-a1 (1) and the amplified SCAR fragment (2).

B. Case Study: *Dracocephalum* L.

Having compared two different ways to gain and use genetic information for the development of specific markers, questions on limitations of the specificity of the markers continuously came to the fore. With *Dracocephalum* we chose a member of the *Lamiaceae*, a family particularly rich in species used as spices, in traditional medicine and harbors many spontaneous hybrids that pose a great challenge not only to taxonomists but also for genetic diagnostics.

In the first part of this study we re-used the formerly established PCR-RFLP methodology, introduced another method which achieved the same result using only one PCR and authenticated commercial tea products containing 'Moldavian Dragonhead'. In the second part we detached our focus from the methodological perspectives and moved on to evaluate reference plants on which inevitably the claim for any primer specificity is based. We then evaluated four barcoding markers on their variability and potential use as diagnostic markers and applied the results to design a specific primer to distinguish between *D. moldavica* and other species of *Dracocephalum*.

1. Product Authentication using PCR-RFLP and ARMS

Dracocephalum moldavica, also called 'Moldavian Dragonhead', has a strong lemon scent and is used in several tea products. Morphological and chemical similarities between 'Moldavian Dragonhead' and 'Lemon Balm' (*Melissa officinalis)* as well as 'Catnip' (*Nepeta cataria*) - all Lamiaceae - are reasonable indicators to consider them potential adulterants. Alongside of morphological studies on leaf fragments of commercial products (data not shown) we analyzed

the plastid marker rbcLa for its diagnostic value in detecting these adulterants using the PCR-RFLP approach. Additionally, to confirm the presence of DNA from *D. moldavica*, we searched for differences that could be used to design a primer specific for *Dracocephalum* species. We found one restriction site present in *Melissa* and *Dracocephalum* while absent in *Nepeta* and another present only in *Melissa* and *Nepeta* (Figure 4, below). For the primer based (ARMS) authentication of *Dracocephalum* we found one site that was specific and could be used to amplify a diagnostic fragment from samples containing *Dracocephalum* DNA. Further details on the results are discussed on page 68 ff.

a) Restriction Analysis

Figure 4: Restriction analysis of rbcLa sequence data from selected *Dracocephalum*, *Melissa* and *Nepeta* accessions. Restriction maps of EcoRI (A) and BamHI (B) comparing accessions of *M. officinalis* (4639, Moff), *D. ruyschiana* (5156, Druy), *D. moldavica* (5156, Dmol) and *N. cataria* (7575, Ncat) are displayed. Recognition sites are in bold face. Perfect sites are labeled with a black triangle at the cut position below the sequence. Imperfect sites can be recognized by the underlined single nucleotide change with respect to a perfect site.

b) PCR-RFLP Assay

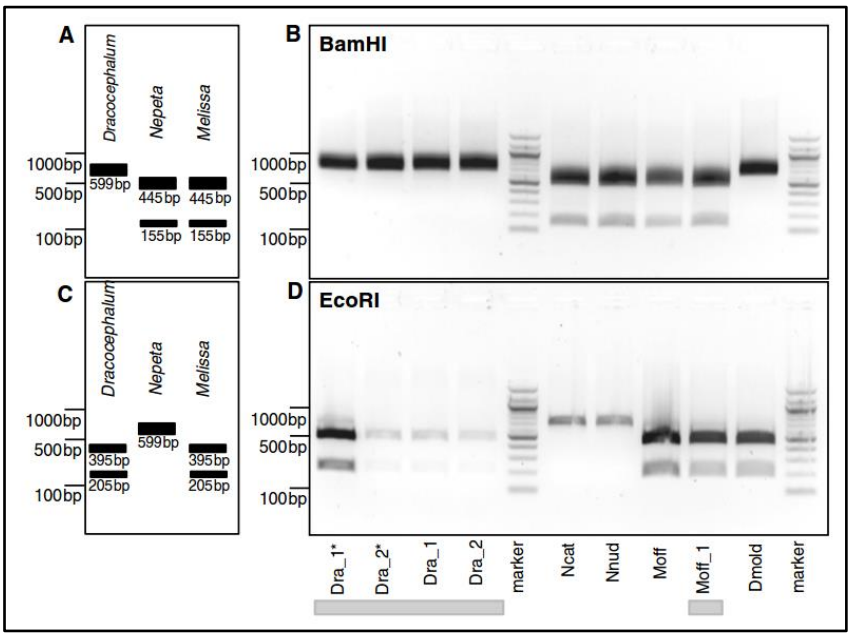


Figure 5: Discrimination of *Dracocephalum moldavica* L. from its surrogate species *M. officinalis* and *N. cataria* based on RFLP of the rbcLa marker in pure samples and commercial tea mixtures. B: RFLP produced by restriction with BamHI. D: RFLP produced by restriction with EcoRI. A, C: Banding patterns predicted from rbcLa sequence data. B, D: Representative electrophoretic patterns observed for pure samples of *N. cataria* L. (Ncat), *N. nuda* L. (Nnud), *M. officinalis* L. (Moff), and *D. moldavica* L. accession DRA2 (Dmold), along with commercial samples Dra_1, Dra_2, and Moff_1 along with samples Dra_1* and Dra_2* derived from the respective commercial sample by selecting putative dragonhead material based on microscopic features.

c) ARMS Assay

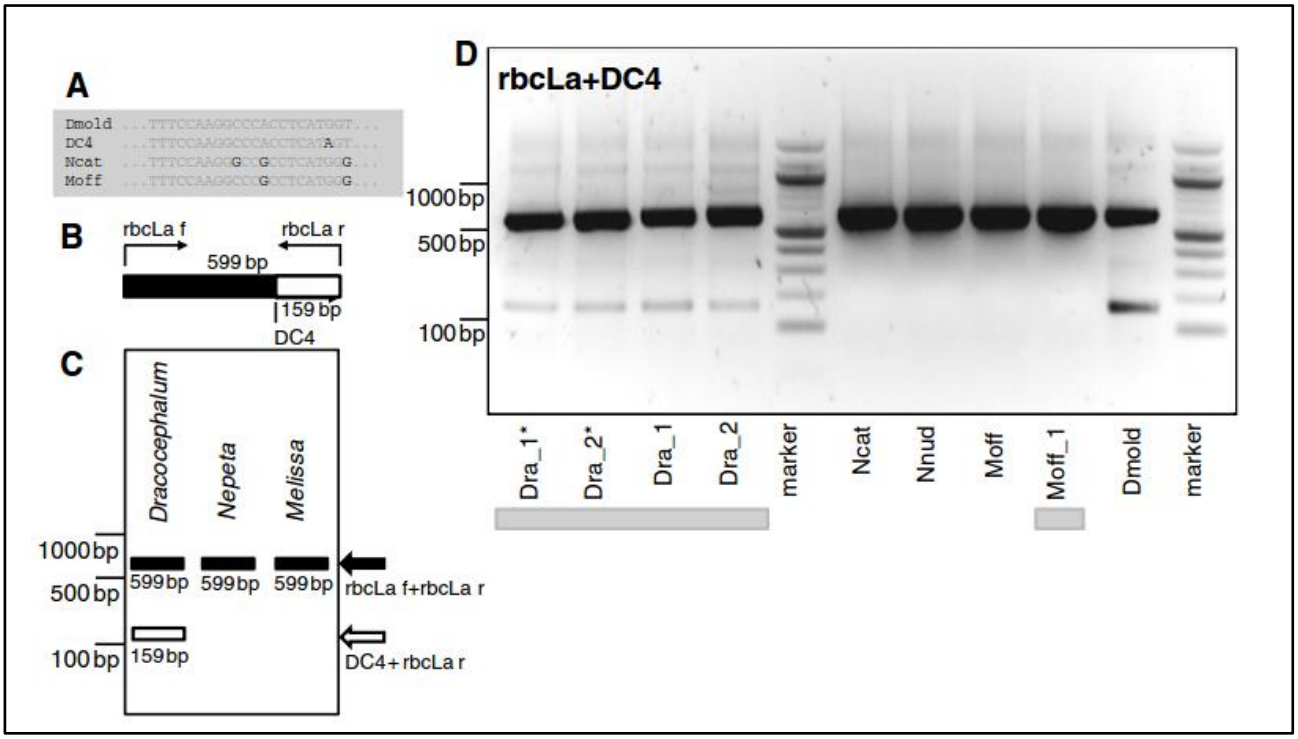


Figure 6: Discrimination of *D. moldavica* L. from its surrogate species *M. officinalis* and *N. cataria* based on ARMS of the rbcLa marker in pure samples and commercial tea mixtures. A: Design of ARMS primer DC4 in relation to the target sequence from *D. moldavica* L. (Dmold) versus *M. officinalis* (Moff) and *N. cataria* (Ncat). Bold letters indicate base changes relevant to annealing stability of the primers. B: Position of the ARMS primer in the rbcLa fragment; C: Banding patterns predicted for successful differential annealing during multiplex PCR using the primer set-up given in B. D: Representative electrophoretic patterns observed for multiplex PCR using primer DC4. For details on the samples see previous Figure.

2. Taxonomic Verification

All described exemplary studies utilized genetic information for diagnostic purposes. In two of those cases the taxonomic differentiation took place on the generic level by differentiating between *Backhousia* and *Leptospermum* in one case, and between *Dracocephalum* and *Melissa* and *Nepeta* in the second case. In this part we wanted to extend our knowledge on commonly used genetic markers evaluating there variability and determining their power of taxonomic resolution. For that aim we increased taxonomic sampling of the genus *Dracocephalum* from originally two different species (*D. moldavica* and *D. ruyschiana*) to 9 and also increased sampling on the species level as far as possible. All accessions were ordered from different botanical gardens and the Leibniz-Institut für Pflanzengenetik und Kulturpflanzen Forschung (IPK). Information on their wild origin however was very limited to completely absent (Table 5, p. 87). When applying methods that determine intra- and interspecific variation, calculate genetic distance or test identification efficacy based on genetic distance, the taxonomic determination must be correct. Any misidentification inevitably would introduce bias into the results. Hence the first step was to determine the identity of the plants. Since only some of the accessions reached maturity and developed sexual organs we used additional characters (i.e. seed morphology, Tubulin based Polymorphism (TBP) and Neighbor Joining (NJ) trees) to approximate their correct determination. To visualize the separation of *Dracocephalum* accessions from other genera we performed a Multidimensional Scaling (MDS) analysis including sequence data of neighboring genera downloaded from GenBank. We used sequence data of three plastid markers (rbcL, matK and psbA-trnH) and one nuclear marker (ITS) for the identity assessment (i.e. NJ and MDS) and subsequent marker evaluation.

a) Overview

Table 3: Results of the taxonomic verification of *Dracocephalum* accessions for subsequent studies. The columns contain accession ID, verification method, result (√ = taxon verified, X = taxon re-classified, / = taxon uncertain) and notes on re-classification (correct or closest taxon), used taxonomic key or reference sequence source. Molecular verification included sequence clustering based on genetic distance using the Neighbor Joining (NJ) method (methodology see page 31) and Tubulin based Polymorphism (TBP) analysis (methodology see page 31).

ID	Verification	Result	Note
Dracocephalum argunense (=Darg)			
7900	Classic	X	*Nepeta* spec. (Rothmaler)
8162	Molecular	√	
8252	Molecular, Seeds	X	*D. ruyschiana*
8262	Molecular, Seeds	√	
8347	Seeds	√	

ID	Verification	Result	Note
Dracocephalum austriacum(=Daus)			
7648	none	-	Failed to germinate
7686	none	-	Failed to germinate
Dracocephalum diversifolium (=Ddiv)			
7691	Classic (Images)	/	Flora of Russia
Dracocephalum foetidum (=Dfoe)			
7986	Molecular	X	*D.moldavica*
Dracocephalum grandiflorum (=Dgra)			
7684	Molecular, Seeds	X	*D.rupestre*
8012	Molecular, Seeds	X	*D.rupestre*
8167	Molecular, Seeds	X	*D.rupestre*
8253	Seeds	X	*D.rupestre*
8349	Molecular, Seeds	X	*D.rupestre*
Dracocephalum integrifolium (=Dint)			
7690	Seeds	/	References NA
Dracocephalum moldavica (=Dmol)			
5861	Classic	√	Julia Völker
5862	Classic	√	Julia Völker
5863	Classic	√	Julia Völker
7682	Classic	√	Flora of China
7687	Molecular	√	
7689	Molecular	√	
7899	Molecular	√	
8021	Classic	√	Flora of China
8138	Molecular	√	
8140	Molecular	√	
8175	Molecular	√	
8223	Molecular	√	
8251	Molecular	√	
8264	Molecular	√	
7701	Molecular	√	
Dracocephalum multicaule (=Dmul)			
7685	none	/	References NA
Dracocephalum nutans (=Dnut)			
8265	none	/	References NA
Dracocephalum parviflorum (=Dpar)			
7874	Molecular	√	GenBank accessions (p. 90)
Dracocephalum renati (=Dren)			
7088	Classic	√	Rothmaler
8013	Classic	√	Rothmaler
8263	Molecular	√	
Dracocephalum rupestre (=Drup)			
7089	Classic	√	Flora of China
8254	Morphology, Seeds, Molecular	X	*D. ruyschiana*
Dracocephalum ruyschiana (=Druy)			
5156	Classic	√	Julia Völker
7683	Molecular, Seeds	√	

ID	Verification	Result	Note
7688	Molecular, Seeds	√	
7921	Molecular, Seeds	√	
8011	Molecular, Seeds	√	
8020	Seeds	/	Molecular data NA
8137	Molecular, Seeds	√	
8141	Molecular, Seeds	√	
8166	-	/	Data NA
8174	Molecular, Seeds	√	
8194	Molecular, Seeds	√	
8209	Molecular, Seeds	√	
8230	Seeds	/	Molecular data NA
8261	Molecular, Seeds	√	
8348	Molecular, Seeds	√	
Dracocephalum scrobiculatum (=Dscr)			
8139	Classic	X	*Hyssopus* spec. (Rothmaler)
Dracocephalum spec. (=Dspec)			
8142	Molecular	/	Very close to *D.moldavica*

b) Morphological Determinations

According to morphological traits four accessions were misidentified. Accession 7900 *D. argunense* was determined to be of the genus *Nepeta* (Figure 22, p. 92), accession 8139 *D. scrobiculatum* was determined to be of the genus *Hyssopus* (Figure 23, p. 93) and accession 7986 *D. foetidum* was determined to be *D. moldavica*. Leaf morphology of accession 8254 suggested that it was not *D. rupestre* (leaves triangular-ovate, 1.4–5.5 × 1.2–4.5 cm) but rather a member of section Ruyschiana (linear to lanceolate-linear 3.4–6.2 cm, bracts ovate-elliptic). Accessions of *D. diversifolium* (7691), *D. foetidum* (7986), *D. moldavica* (all), *D. ruyschiana* (5156, 7688), *D. renati* (7088, 8013) and *D. rupestre* (7089) reached maturity and developed flowers. At least one accession of *D. foetidum*, *D. moldavica*, *D. ruyschiana*, *D. renati* and *D. rupestre* was morphologically determined using literature indicated in Table 3. Images of *D. diversifolium* were captured and later used for comparison with a drawing from the Flora of the USSR (1968) and with images from other sources (Figure 24 - Figure 26, p. 94 ff.). All accessions that reached maturity were included in a herbarium. None of the *D. grandiflorum* accessions (7684, 8012, 8167 and 8253) reached a generative state but displayed morphological similarity to *D. rupestre* (7089). While *D. rupestre* has numerous basal triangular-ovate leaves *D. grandiflorum* has basal oblong-elliptic, rarely ovate leaves. All *D. grandiflorum* accessions displayed the triangular-ovate leaf type. *D.argunense*, *D. integrifolium*, *D. multicaule*, *D. nutans, D. parviflorum* and unclassified accession 8142 did not reach maturity and could not be determined morphologically.

To further increase taxonomic certainty seed morphology and molecular analyses in particular of section *Ruyschiana* and the *D. grandiflorum* / *D.rupestre* complex were conducted.

c) Seed Morphology

Among the seed characters area (AS), circularity (CS) and length-width-ratio (LWR) appear to be the most useful. Considering CS of seeds (Figure 7, left, p. 46) we can divide our *Dracocephalum* accessions roughly in three groups. Group A contains *D. rupestre* (0.60), *D. grandiflorum* (0.59) and *D. renati* (0.61) with mean values around 0.60. Group B contains *D. moldavica* (0.66), *D. integrifolium* (0.68), *D. nutans* (0.69), *D. diversifolium* (0.69), *D. foetidum* (0.71) and the unclassified accession (0.71) with mean values between 0.65 and 0.72. And group C contains *D. ruyschiana* (0.75), *D. parviflorum* (0.78) and *D. argunense* (0.82) with mean values above 0.74. In case of seed area (Figure 7, right) the species with the highest value is *D. diversifolium* (5.69 mm^2) followed by *D. renati* (3.74 mm^2), *D.argunense* (3.44 mm^2), *D. grandiflorum* (3.20 mm^2), *D. rupestre* (3.10 mm^2) and *D. ruyschiana* (2.92 mm^2). D. moldavica and *D. foetidum* show very similar values (2.71 mm^2 and 2.67 mm^2). A rather continuous group consists of the unknown species (*Dracocephalum*) with 2.14 mm^2, *D. integrifolium* with 2.03 mm^2 and *D. parviflorum* with 1.96 mm^2. Among the smallest seed areas are displayed by *Nepeta* species ex *D. argunense* with 1.36 mm^2, *Hyssopus* species ex *D. scrobiculatum* with 1.25 mm^2 and *D. nutans* with 1.03 mm^2. The number of observations considered is lowest for *D. diversifolium* (9), *D. integrifolium* (25) and *D. parviflorum* (27). All other groups contained information of more than 50 observations.

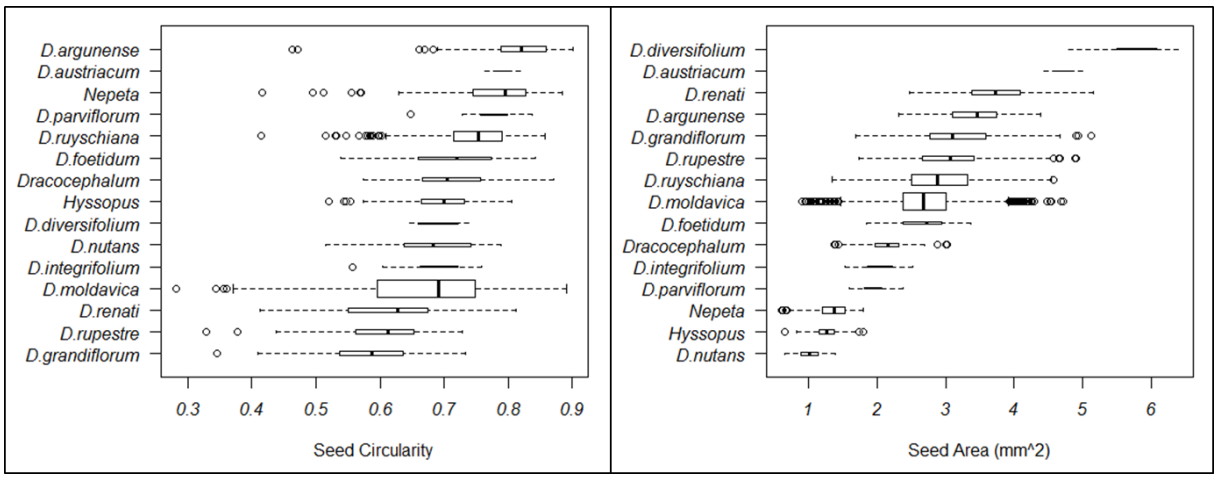


Figure 7: Comparison of seed area (right) and circularity (left) of putative *Dracocephalum* species. Boxplot heights were plotted proportional to the number of seeds analyzed. The order of accessions is descending by mean value.

Ruyschiana: Significant difference in seed LWR (Figure 8, p. 47) can be observed between two accessions of *D. argunense* (8262 and 8347), all *D. ruyschiana* and one putative *D. argunense* accession (8252). The arithmetic mean of LWR considering all seeds of all *D. ruyschiana* accessions is 1.906 (sd: 0.256, se: 0.007). The arithmetic mean of LWR of seeds from D. argunense (8262) is 1.386 (sd: 0.181, se: 0.013). Correlation between LWR and CD (Figure 9, p. 47) shows a more restricted distribution of respective values in *D. argunsense* than in *D. ruyschiana* accessions. The presence of malformed seeds and their influence on the dataset can be detected by comparison of the two regression lines. The lowest observation count (27) is represented by *D. argunense* accession 8347.

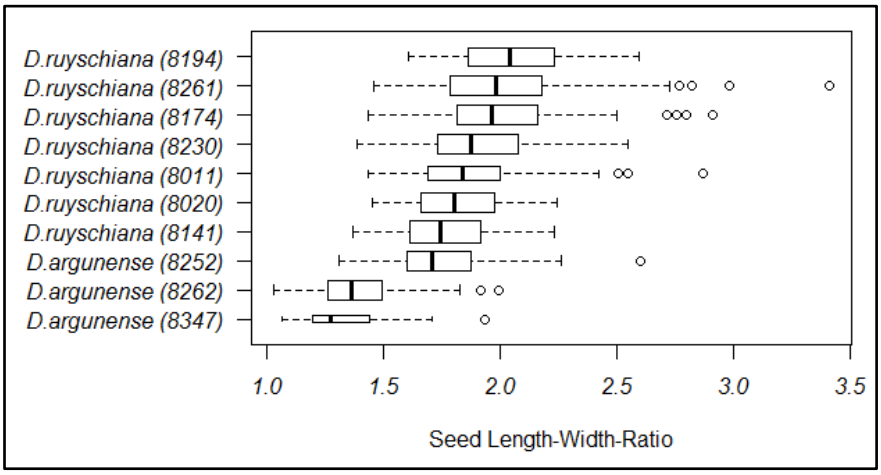


Figure 8: Comparison of seed length-width-ratio (LWR) of members of section *Ruyschiana* (*D. ruyschiana* and *D. argunense*). Boxplot heights were plotted proportional to the number of seeds analyzed. The order of accessions is descending by mean value.

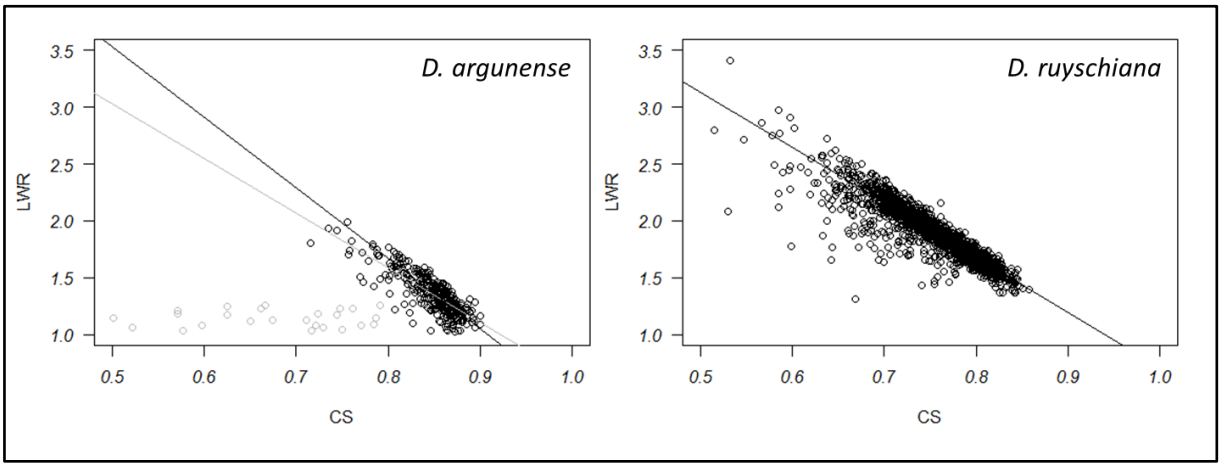


Figure 9: Correlation of seed length-width-ratio (LWR) and circularity (CS) of members of section *Ruyschiana* (*D. ruyschiana* and *D. argunense*). For overlaying a robust regression line in case of *D. ruyschiana* (right) all data was used and in the case of *D. argunense* (left) grey data points were excluded. The grey regression line includes all data points.

D.rupestre / D. grandiflorum: While there are no significant differences in CS (Figure 10, p. 48) between *D. rupestre* (7089) and all *D. grandiflorum* accessions the putative *D. rupestre* accession (8254) displays CS intermediate to both representatives of *D. ruyschiana*. Accession 8141 and 8174 represent the extremes in CS among *D. ruyschiana* accessions. While accession 8175 with a mean CS value of 0.70 does not fit into the defined circularity groups (see above) accession 8254 with a mean value of 0.74 comes very close. Additionally, CS of 8254 is significantly different from other *D. rupestre* accession. The lowest observation counts are represented by accessions 7684 (14) and 8254 (20).

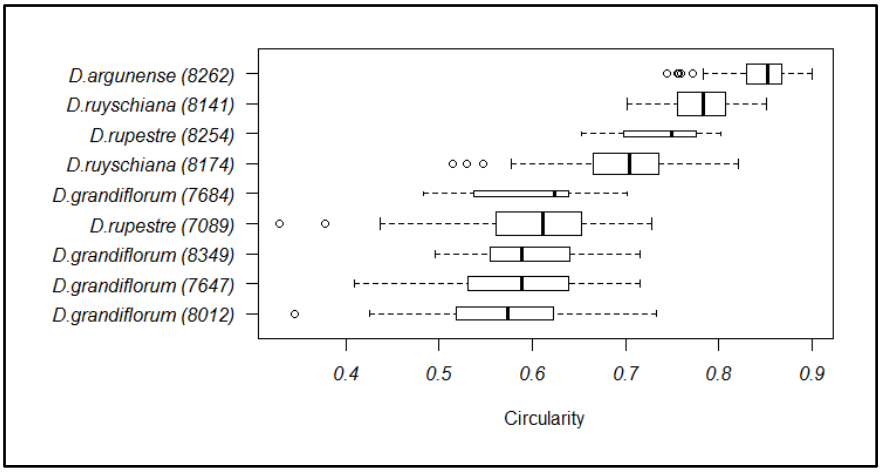


Figure 10: Comparison of seed circularity between *D. rupestre*, *D. grandiflorum*, *D. ruyschiana* and *D. argunense* accessions. Boxplot heights were plotted proportional to the number of seeds analyzed. The order of the accessions is descending by mean value.

d) Tubulin based Polymorphism

To see if we can correlate findings of morphological determinations and the seed character analysis with a molecular approach and to evaluate species specificity of the same, we analyzed TBP fingerprints for most of our *Dracocephalum* accessions.

Patterns observed in *D. moldavica* accessions were consistent showing no differences (data not shown). Comparing patterns between different species revealed unique patterns in all cases (data not shown).

Ruyschiana: While only one fragment is present in all samples (Figure 11, “1>”, 210 bp) another one (“2>”, 300 bp) is only shared by the *Ruyschiana* samples. Within *Ruyschiana* the absence of strongly amplified fragments at around 390 bp (“>3”) and 490 bp (“>4”) in two of the *D. argunense* accessions as well as the presence of a larger fragment (“>5”) indicates the difference between putative *D.argunense* accessions: 8162 and 8262 (blue) and 8252 (green). The pattern of 8252 can also be found in other *D.ruyschiana* accessions (8254 and 8141). While these patterns are

identical, all other accessions of *D.ruyschiana* are missing the fragment at around 490 bp (“>4”), which can be used to distinguish a subset of *D. ruyschiana* accessions.

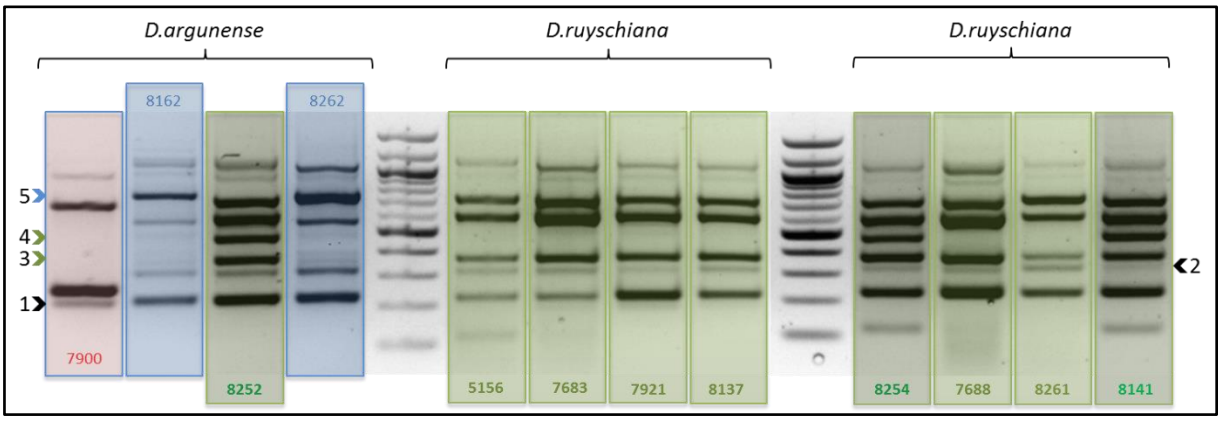


Figure 11: DNA fingerprinting pattern of putative *Ruyschiana* accessions (*D. ruyschiana* and *D. argunense*) based on tubulin intron polymorphism (TBP). The background coloring (red, blue and green) is based on the similarity of the fragment pattern. The uncolored lanes contain 100 bp size marker (NEB).

D. rupestre / D. grandiflorum: With the exception of 8254 all *D.rupestre* and *D.grandiflorum* accessions show a consistent pattern (Figure 12). The TBP fingerprint of 8254 is most similar with that of 8141 and shares more fragments with accessions 7688 and 8261 (*D. ruyschiana*) than with *D. rupestre* / *D. grandiflorum* accessions.

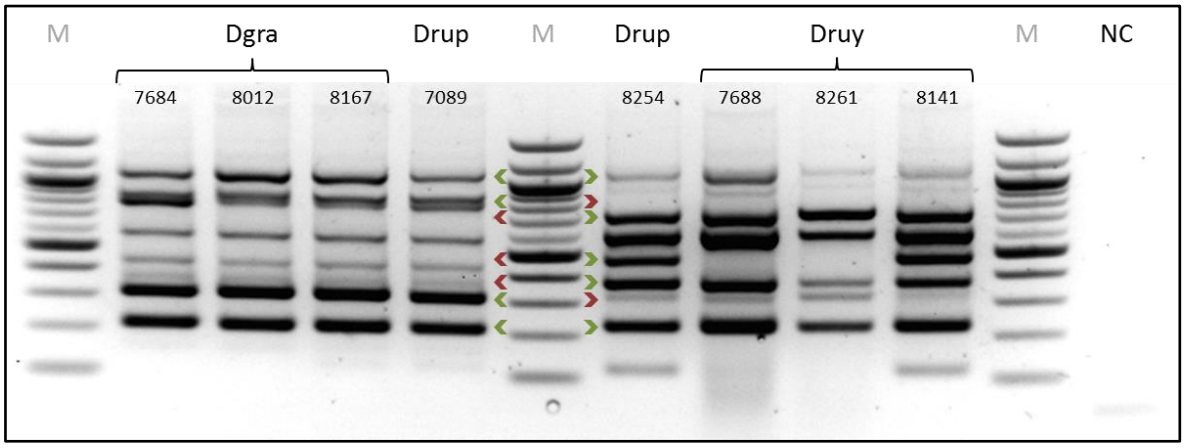


Figure 12: DNA fingerprinting pattern of putative *D. grandiflorum* (Dgra), *D. rupestre* (Drup) and *D. ruyschiana* (Druy) accessions based on tubulin intron polymorphism (TBP). Fragment inconsistencies between 8254 and *D. rupestre* / *D. grandiflorum* accessions are indicated with opposing green (present) and red arrows (absent or considerably weaker amplification) on the respective side of the middle 100 bp size marker (NEB).

e) Genetic Distance Analysis

In the final identity assessment we used sequence data from psbA-trnH and ITS markers to construct Neighbor Joining (NJ) trees based on Kimura 2-parameter (K2P) genetic distance (for rbcLa and matK-KIM NJ trees see Figure 27 and Figure 28 on page 96 ff.). In some cases external sequence data was available as reference for singleton species (Table 9, p. 90). To increase taxonomic certainty for singleton species where no external sequence data was available and to get an impression of the genetic distance of neighboring genera, we included sequences of respective accessions after a BLAST [Altschul 1990] search using several *Dracocephalum* sequences. Besides a tree representation we included a Multidimensional Scaling (MDS) plot that contains a broader range of neighbors while in the NJ tree only the closest are included along with our own sequence accessions from other *Lamiaceae* (*Lallemantia*, *Nepeta*, *Melissa* and *Hyssopus*).

ITS: Within the NJ tree build from K2P distances of ITS45 sequence data (Figure 13, p. 51) *D. parviflorum*, *D. ruyschiana* and *D. argunense* (external data) form one cluster supported by 68 % of the bootstrap trees. Internally *D. ruyschiana* is separated into two sub-clusters (8137, 8254 and 5156 vs. 8252) including the re-classified accession 8254 (ex *D. rupestre*). *D. parviflorum* adheres with the respective external accession (JQ669097). In the neighboring cluster *D.multicaule* is located at the base of a cluster containing accession 8139 (*Hyssopus* species ex *D. scrobiculatum*) and external *Hyssopus* accessions. The number of base substitutions per site (BSpS) from averaging over all sequence pairs between *Hyssopus* and *Dracocephalum* accessions is 0.0327 (0.0056). Accessions received as *D. grandiflorum* (7684, 8012 and 8167) share identical sequence data with the morphologically determined accession of *D. rupestre* (7089) and form a separate cluster. The external sequence of *D. grandiflorum* is located at the base of that cluster separated by 0.0047 BSpS. Their next closest neighbor is *D. bullatum* with 0.0064 BSpS compared to *D. rupestre*. The unclassified accession (8142) is located at the base of a cluster formed by all *D. moldavica* accessions. The number of BSpS compared to our *D. moldavica* accessions is 0.0046 (0.0024) and 0.0175 (0.0050) compared to the external accession. Comparing external and internal *D. moldavica* accessions yields 0.0127 (0.0043) BSpS. Excluding the *Hyssopus* group the number of BSpS from averaging over all sequence pairs between groups (*Dracocephalum* vs. non-*Dracocephalum*) is 0.0843 BSpS (0.0072) while averaging over all sequence pairs within each group returns 0.0248 BSpS (0.0035) for the *Dracocephalum* and 0.0982 BSpS (0.0074) for the non-*Dracocephalum* group respectively. Values in round brackets are standard error estimates obtained by a bootstrap procedure with 1000 replicates. According to the MDS result *Hyssopus*

and *Schizonepeta* are the two closest genera. While *Schizonepeta* is more distant to the *Dracocephalum* group *Hyssopus* appears to be nested within.

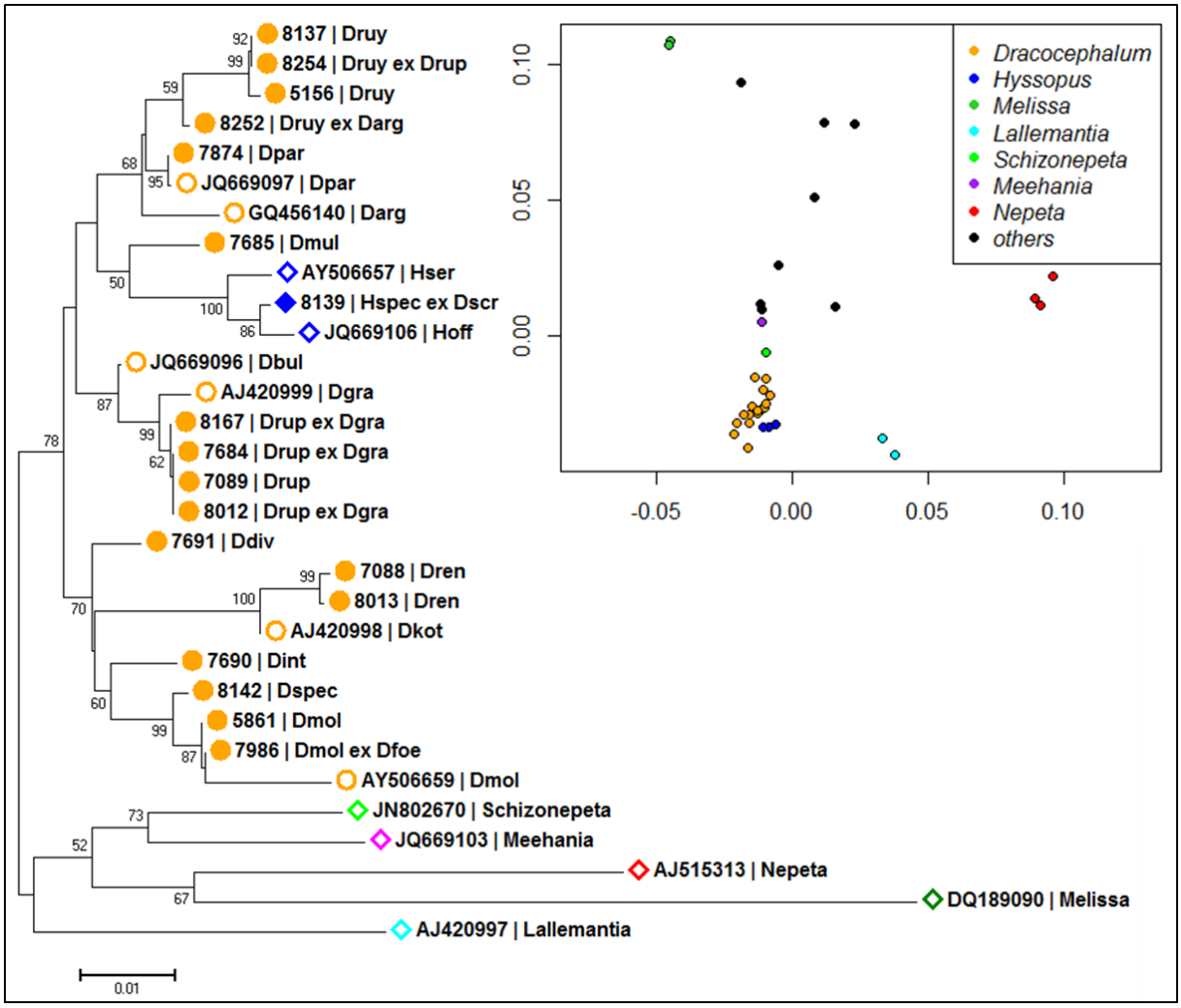


Figure 13: Neighbor-Joining tree and Multidimensional Scaling of selected *Dracocephalum* and neighboring genera accessions based on K2P distance of ITS45 sequence data. The optimal tree with the sum of branch length = 0. 3785 is shown. The root is placed on the branch containing *Dracocephalum* and *Hyssopus* (DH-Cluster). The analysis involved 31 nucleotide sequences and there were a total of 704 positions in the final dataset. All filled circles and rhombs indicate sequence data retrieved by the author. The respective forms with colored margin indicate sequence data downloaded from external sources (Table 9, p. 90). For taxon abbreviations please refer to Table 3.

psbA-trnH: The NJ tree based on K2P distances of psbA-trnH sequence data shows the unclassified accession (8142) in a cluster with all *D. moldavica* accessions. The average number of base substitutions per site (BSpS) between the group of *D. moldavica* accessions and the unclassified accession of 0.0143 (0.0057) is higher than the number of BSpS from averaging over all sequence pairs within the *D. moldavica* group (0.0005). *D. ruyschiana*, *D. argunense* and *D. parviflorum* are located on a separate branch. While *D. ruyschiana* accessions form two distinct clusters that are separated by 0.0204 (0.0070) BSpS, one of them shares a common ancestor with *D.argunense* and is separated by 0.0180 (0.0063) BSpS. The other appears more distant with 0.0330 (0.0088) BSpS. Similarly *D. parviflorum* is closer to the first (0.0225 BSpS) than to the second (0.0303 BSpS) *D.*

ruyschiana group and even more distant to *D. argunsense* (0.0348 BSpS). Accession 8252 (*D. ruyschiana* ex *D. argunense*) and 8254 (*D. ruyschiana* ex *D. rupestre*) are well separated from their original classifications. Accessions received as *D. grandiflorum* (7684, 8012, 8167 and 8349) share identical (8349 and 8012) or very similar (7684 and 8167) sequence data with the morphologically determined accession of *D. rupestre* (7089) and form a separate cluster. The external sequence of *D. rupestre* (FJ513110) is located at the base of that cluster separated by an average rate of BSpS of 0.0181 (0.0060). The number of BSpS from averaging over all sequence pairs between groups (*Dracocephalum* vs. non-*Dracocephalum*) is 0.1052 (0.0101) while averaging over all sequence pairs within each group returns 0.0433 BSpS (0.0058) for the *Dracocephalum* group and 0.0990 BSpS (0.0088) for the non-*Dracocephalum* respectively. Values in round brackets are standard error estimates obtained by a bootstrap procedure with 1000 replicates. According to the MDS result *Schizonepeta* and *Hyssopus* are the two closest genera placed near our *D. diversifolium* accession. BSpS from averaging over all sequence pairs between *Dracocephalum* accessions excluding *D. diversifolium* (Draco-Div) and *D. diversifolium* are 0.0376, between Draco-Div and *Schizonepeta* 0.0660 and between Draco-Div and *Hyssopus* 0.1010. This situation is more accurately represented by the NJ tree than by the MDS plot considering the orange dot close to *Hyssopus* (blue) and *Schizonepeta* (lightgreen) is *D. diversifolium*.

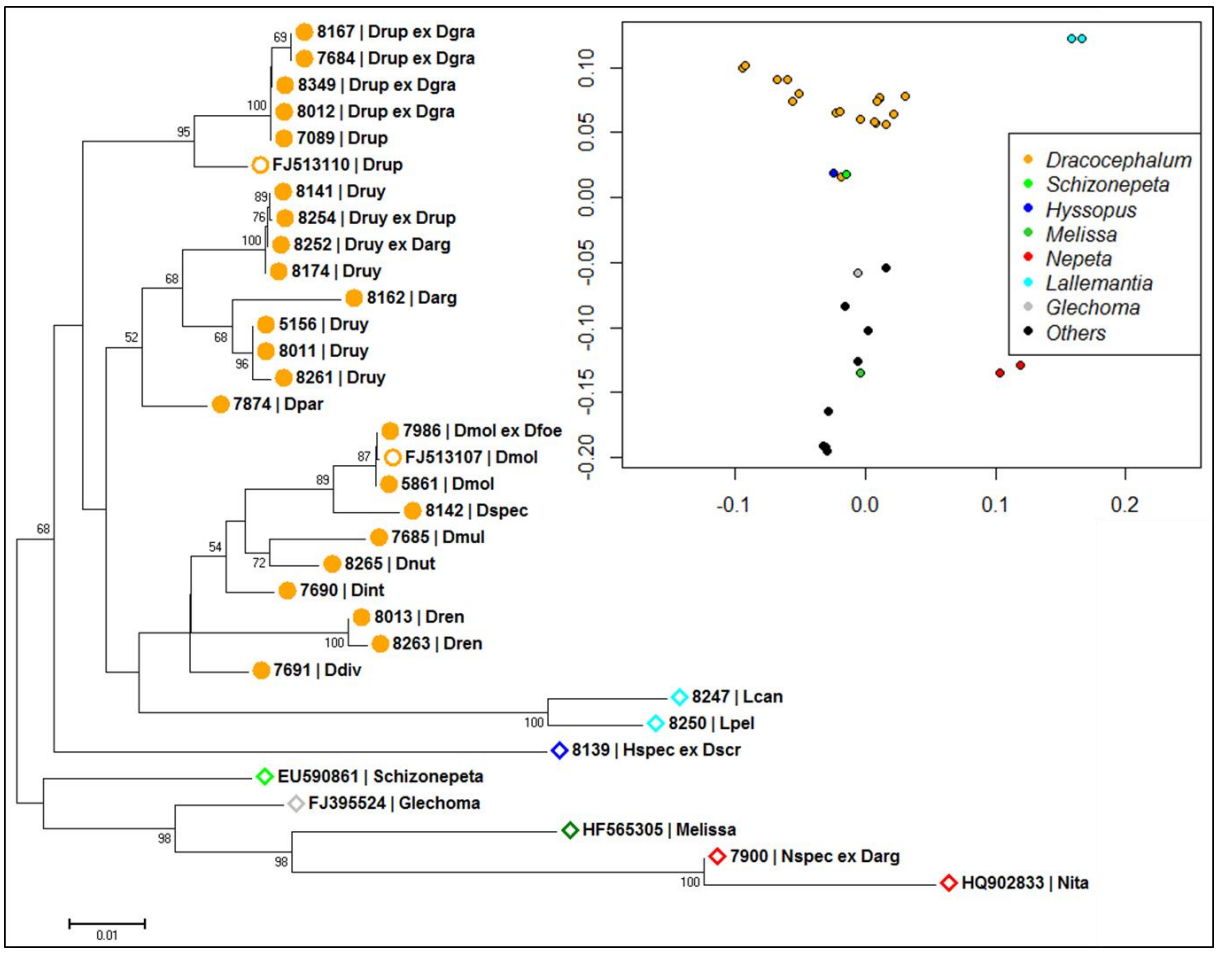


Figure 14: Neighbor-Joining Tree and Multidimensional Scaling of selected *Dracocephalum* and neighboring genera accessions based on K2P distance of psbA-trnH sequence data. The optimal tree with the sum of branch length = 0. 5494 is shown while the root is placed on the Cluster containing *Dracocpehalum*, *Hyssopus* and *Lallemantia* (DHL-Cluster). The analysis involved 33 nucleotide sequences and there were a total of 584 positions in the final dataset. All filled circles and rhombs indicate sequence data retrieved by the author. The respective forms with colored margin indicate sequence data downloaded from external sources (Table 9). For taxon abbreviations please refer to Table 3.

f) Summary

According to the reviewed data so far we decided to continue under the assumption that all accessions received as *D. grandiflorum* are instead *D. rupestre*. Seed data weakly indicates two groups supported by psbA-trnH sequence data. With respect to the section *Ruyschiana* group indications are considerably stronger and could be signs for the presence of a third species. However we decided to continue with the assumption that two accessions of *D. argunense* are available and all other accessions are from *D. ruyschiana* supported by shared TBP pattern types. All singleton species are closely aggregated with non-singleton species (MDS in Figure 13 and Figure 14, above) with the exception of *D. diversifolium* considering psbA-trnH data (Figure 14). Since morphological data of *D. diversifolium* is compelling and MDS and NJ clustering supports at least genus membership of the remaining singleton species we continued with all singleton species by their given taxon name.

Table 4: Results of retrieved marker sequence data after taxonomic verification. Number of sequences retrieved (n) and unique haplotypes (u) of marker regions rcbLa, matK-KIM, ITS45 and psbA-trnH are displayed. When more than one haplotype was found bold numbers are used.

Taxon	rbcLa		matK-KIM		ITS45		psbA-trnH	
	n	u	n	u	n	u	n	u
D. argunense	2	1	2	1	-	-	2	1
D. diversifolium	1	1	1	1	1	1	1	1
D. integrifolium	1	1	-	-	1	1	1	1
D. moldavica	15	1	13	1	13	1	15	1
D. multicaule	1	1	1	1	1	1	1	1
D. nutans	1	1	-	-	-	-	1	1
D. parviflorum	1	1	1	1	1	1	1	1
D. renati	3	1	3	1	3	**2**	3	**2**
D. rupestre	5	1	4	1	4	1	5	**2**
D. ruyschiana	11	1	10	**2**	10	**2**	14	**4**
D. species	1	1	1	1	1	1	1	1
Sequences	42	13	36	**10**	35	**16**	45	**18**
Species		12		9		13		11
D. argunense	-	-	-	-	-	1	-	-
D. bulatum	-	-	-	-	-	1	-	-
D. grandiflorum	-	1	-	-	-	1	-	-
D. kotschyi	-	-	-	-	-	1	-	-
D. moldavica	-	-	-	-	-	1	-	1
D. rupestre	-	1	-	-	-	-	-	1
Sequences from external sources (Details see Table 8 and Table 9)								

3. Marker Evaluation

A general aspect of genetic markers is their universality which is reflected by the ease of amplification from diverse taxa without the need of designing new primers and a consistent success rate of subsequent sequencing attempts. Since amplification success not only depends on the marker (i.e. used primers) but also on the quality of the DNA extract, we consider amplification success as the ability to amplify the marker region even if DNA had to be extracted more than once. For sequence recovery values we use the percentage of nucleotides recovered excluding the primer sequences. Using TaxonGap we display variation within the marker regions in terms of separability (interspecific variation) and heterogeneity (intraspecific variation).

a) Plastid Markers

Both rbcLa and psbA-trnH PCR amplification success was around 100 % while matK-KIM displayed some difficulties resulting in about 92 % amplification success. Regarding the recovery of good sequencing results rbcLa performed very well with 99.2 % recovery followed by psbA-trnH with

98.4 % and matK with 95.3 %. The GC-content within the marker regions is highest in rbcLa (42.9 %) followed by matK-KIM with 33.8 % and 29.5 % in the inter-genic region between psbA and trnH. Length variation only has been observed in the inter-genic region with an average of 416 bp, the maximum length in *D. ruyschiana* (425 bp) and the minimum length in *D. diversifolium* (363 bp). Note that primer sequences are not included in these values. A preliminary analysis indicated the presence of inverted repeats within the psbA-trnH region of *Dracocephalum* species (data not shown).

(1) Heterogeneity and Separability

Among the plastid markers (Figure 15) psbA-trnH shows the highest separability with an arithmetic mean of 4.95 % compared to 0.84 % for matK-KIM and 0.21 % for rbcLa respectively. Four species (*D. moldavica*, *D. multicaule*, *D. integrifolium* and *D. nutans*) show identical sequence data and cannot be separated using rbcLa sequence data. The minimum separability is 0.1 % using matK-KIM sequence data and 1.7 % using psbA-trnH data respectively. While there is no heterogeneity within rbcLa sequence data matK-KIM shows 0.4 % in *D. ruyschiana*. In psbA-trnH three additional species appear with heterogeneity within the analyzed accessions (*D. moldavica*, *D.renati* and *D. rupestre*). The values reach from 0.2 % (*D. renati*) up to 5.2 % in *D. moldavica*. Please note that for matK-KIM three species could not be included into the analysis.

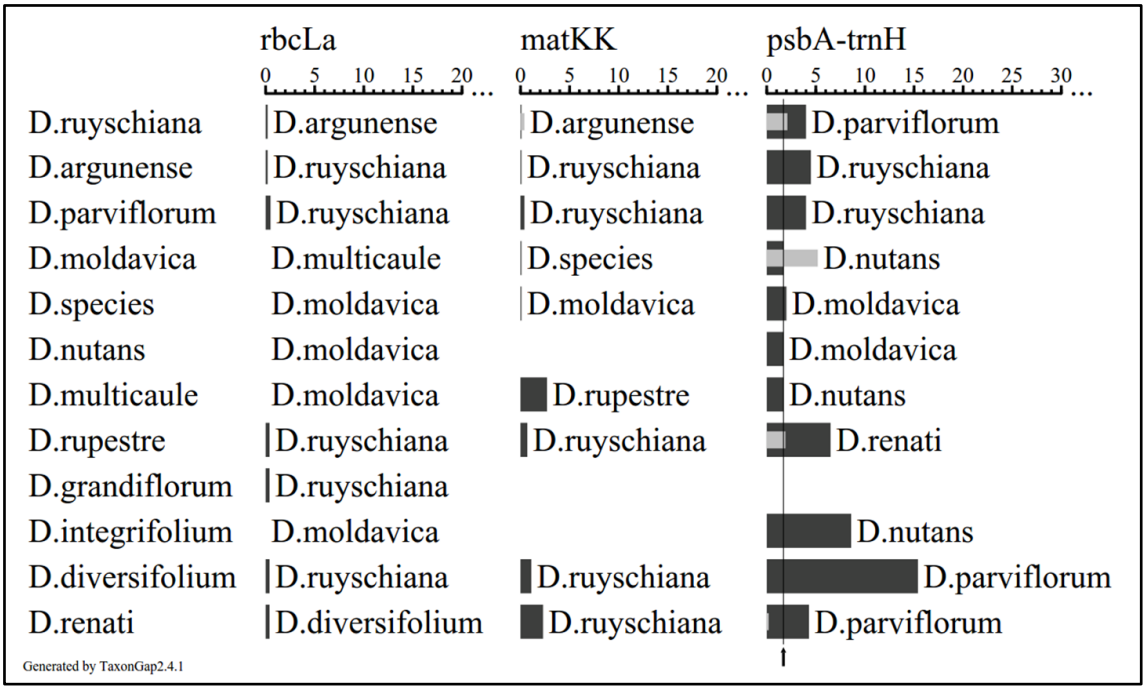


Figure 15: Evaluation of heterogeneity and separability of plastid barcoding markers. Heterogeneity (light grey bars) and separability (dark grey bars) of each *Dracocephalum* taxon are displayed. Taxon names in the marker columns indicate the closest species (in case more than one species are closest the first is displayed). The vertical line and the arrow in the psbA-trnH marker column indicate minimal separability.

b) Nuclear Markers

Amplification success using ITS45 primers was similar to matK-KIM with 89 %. The main problem here however was the occasional co-amplification of an additional fragment from samples of *D. ruyschiana*. In two instances a blast search indicated that the additional fragment originated from an epiphytic fungi and a nematode respectively. Regarding the recovery of good sequencing results again ITS45 performed similarly to matK-KIM with 93 %. The GC-content of 62 % within the marker region is the highest observed among all analyzed markers. Length variation was also observed within the ITS45 region. The minimum length of 671 bp was found in the external accession of *D. grandiflorum* and the maximum length of 695 bp is shared by *D. argunense* and *D. parviflorum* while *D. ruyschiana* with 694 bp is one base pair shorter. The average length of ITS45 excluding primer sequences is 682 bp. Considering ITS1 and ITS2 the length difference of ITS45 is mainly based on a 14 to 15 bp long Insertion / Deletion found in ITS2.

(1) Heterogeneity and Separability

The analysis of the divided nuclear marker (Figure 16) ITS45 is characterized by a mean separability of 0.19 % and a single species (*D. ruyschiana*) with heterogeneity within the region coding for the 5.8S subunit of ribosomal RNA. The internal transcribed spacer regions on the other hand show considerably higher values in separability (ITS1: 2.45 % and ITS2: 3.38 %). In both regions significant heterogeneity can be observed and *D. moldavica* shows enormous heterogeneity, which in both cases (ITS1 and ITS2) is mainly due to missing sequence data in one of the two accessions. Minimum separability values are situated between those of matK-KIM and psbA-trnH with 0.5 % for ITS1 and 0.8 % for ITS2. Although the mean separability is higher in ITS2 in some cases the individual value is higher in ITS1 (e.g. *D. argunense* with 13.5 % using ITS1 and 8.8 % using ITS2 sequence data). It might be worth mentioning that only 5 out of 13 taxa (*D. kotschyi*, *D. parviflorum*, *D. renati*, *D. ruyschiana* and *D*. species) display the same closest congeneric in both spacer regions.

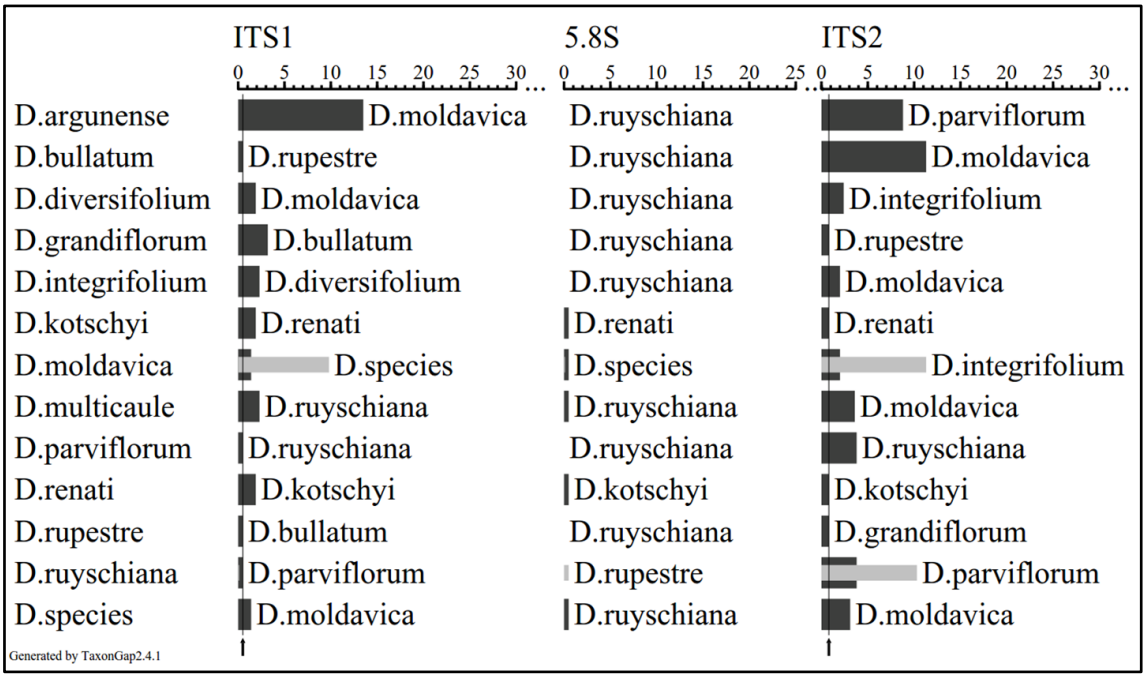


Figure 16: Evaluation of heterogeneity and separability of the internal transcribed spacer region. Heterogeneity (light grey bars) and separability (dark grey bars) of each *Dracocephalum* taxon for ITS1, 5.8S and ITS2 are displayed. Taxon names in the marker columns indicate the closest species (in case more than one species are closest the first is displayed). The vertical line and the arrow in the ITS1 and ITS2 marker columns indicate minimal separability.

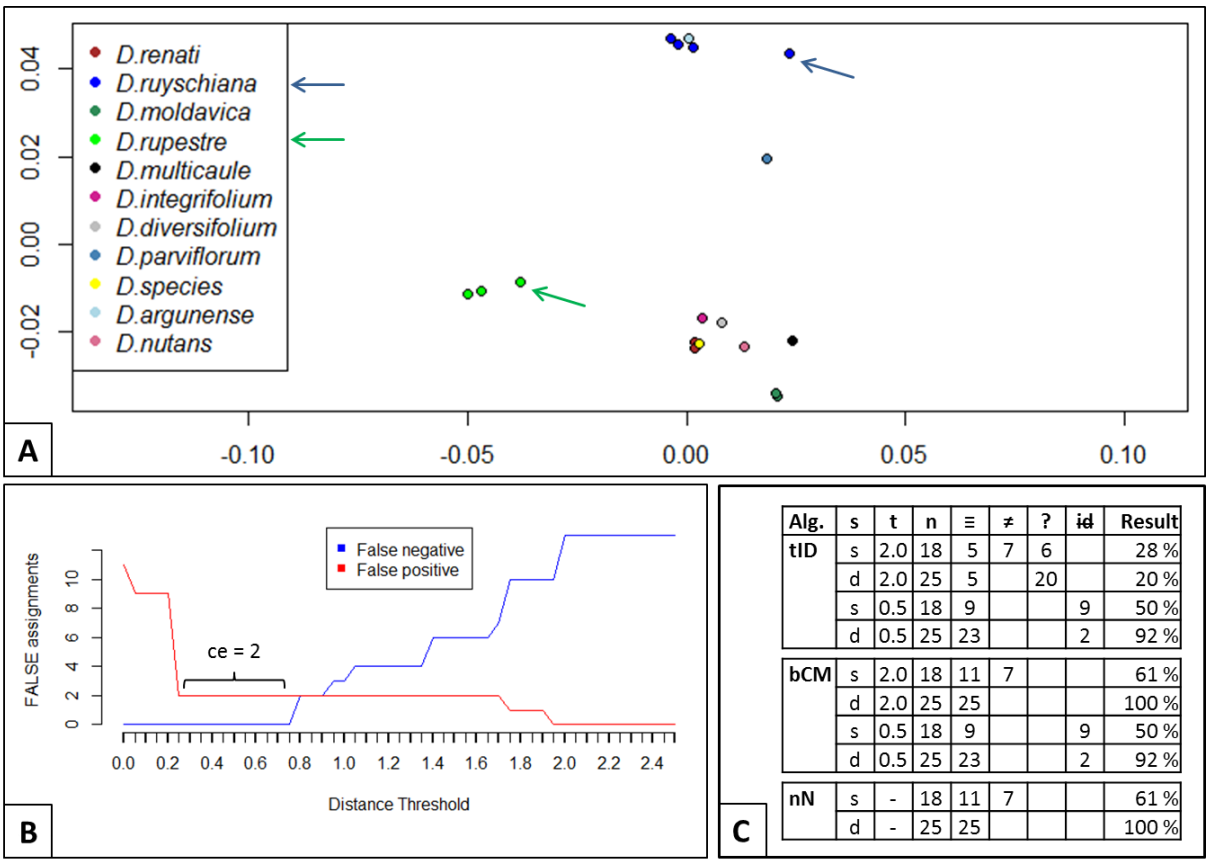
4. Species Level Diagnostics

Species identification efficacy was tested using the most suitable markers (psbA-trnH and ITS) and their diagnostic potential was determined by a sliding window analysis of diagnostic nucleotides. Finally considering the results we designed an ARMS primer based on psbA-trnH sequence data specific for *D. moldavica* and demonstrated its performance among samples from other *Dracocephalum* species and commercial products containing 'Moldavian Dragonhead'.

a) Distance Based Identification Algorithms

(1) psbA-trnH

Using psbA-trnH sequence data in a distance based identification efficacy analysis (Figure 17, p. 58) results in 92 % and 100 % identification success depending on the algorithm and threshold used. When using a 2 % threshold with bestCloseMatch (bCM) the outcome of 100 % identification success reflects that of using nearNeighbour (nN). With singletons present in the dataset the algorithms as expected fail to identify them and including copies resolves their failed identification. As soon as an optimized threshold (0.5 %) is being used the result of bCM reflects that of threshID (tID) where 2 sequences result in "no identification" due to genetic distances beyond the scope of the optimized threshold (*D.ruyschiana* accession 8141 with 2 % and *D.rupestre* accession FJ513110 with 1.8 % intra-specific distance, see Figure 17 A).



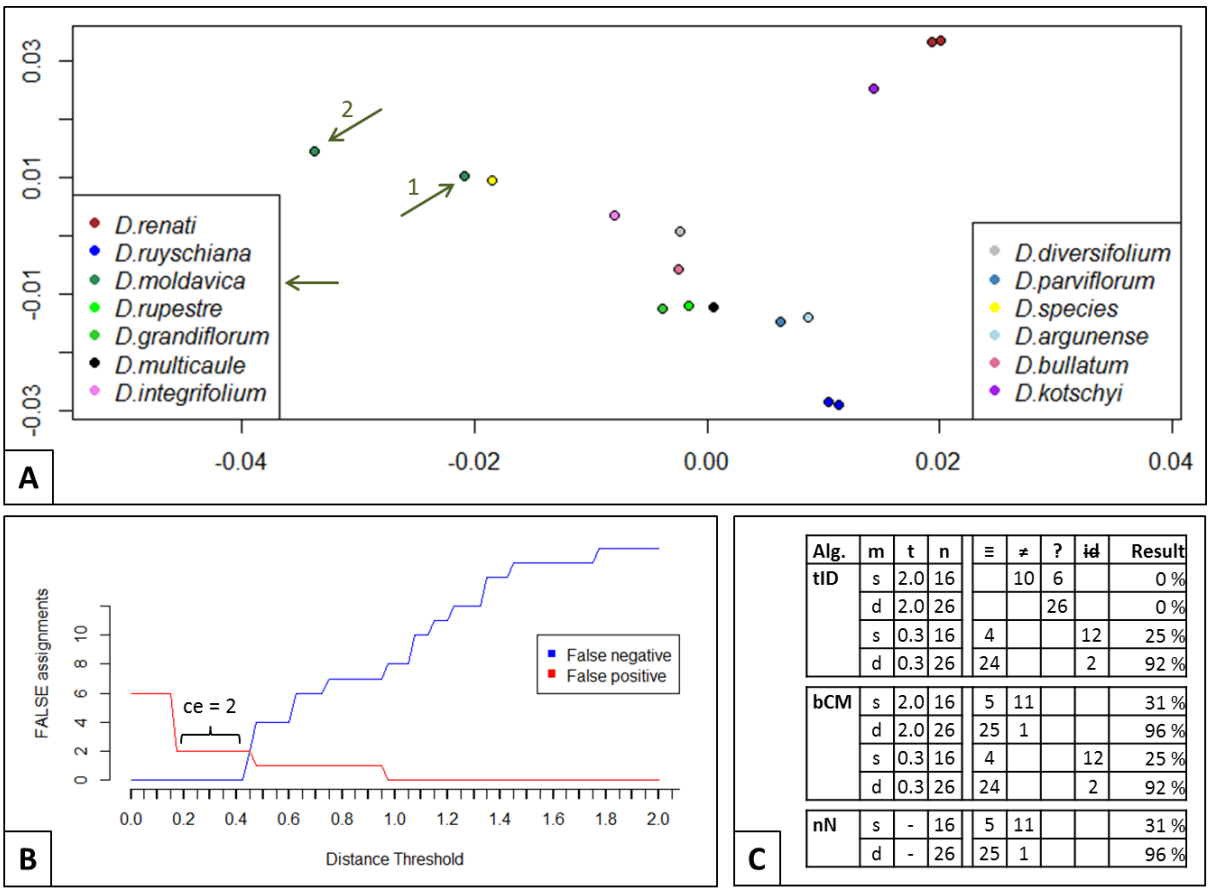
Alg.	s	t	n	≡	≠	?	id	Result
tID	s	2.0	18	5	7	6		28 %
	d	2.0	25	5		20		20 %
	s	0.5	18	9			9	50 %
	d	0.5	25	23			2	92 %
bCM	s	2.0	18	11	7			61 %
	d	2.0	25	25				100 %
	s	0.5	18	9			9	50 %
	d	0.5	25	23			2	92 %
nN	s	-	18	11	7			61 %
	d	-	25	25				100 %

C

Figure 17: Results of the accession identification efficacy analysis based on K2P-distance of psbA-trnH sequence data. A: Multidimensional Scaling of the used distance matrix. Arrows indicate species and respective individual sequences which could not be identified due to intraspecific distance outside the defined threshold (*D.ruyschiana* accession 8141 with 2 % and *D.rupestre* accession FJ513110 with around 1.8 % intraspecific distance). B: Determination of the optimal distance threshold. The plot shows false negative (blue) and false positive (red) species assignments between 0 and 2.5 % distance threshold. The lowest cumulative error (ce = 2) was found between 0.20 and 0.77 % distance. C: Table with results of threshID (tID), bestCloseMatch (bCM) and nearNeighbour (nN) algorithms. With each threshold two analyses were conducted. One that included singleton species (s = s) and one where the respective sequences were duplicated (s = d). In column "n" the number of sequences is indicated. Thresholds (t) were 2.0 and 0.5 % for tID and bCM analysis. Results are shown in the last 5 columns: identification success (≡), identification failure (≠), ambiguous result (?), no identification (~~id~~). The last column contains the accession identification efficacy value in percent.

(2) ITS

Using ITS45 sequence data in a distance based identification efficacy analysis (Figure 18) results in similar outcomes like with psbA-trnH data except for the identification success values which are slightly lower: 92 % using threshID or bestCloseMatch (bCM) with an optimized threshold of 0.3 % and 96 % using bCM with a 2 % threshold or nearNeighbour (nN). The species which could not be identified due to an intraspecific distance outside the defined threshold (0.95 %) is *D. moldavica* (local and external accession AY506659). In case of bCM and nN one of both sequences from *D. moldavica* is identified correctly while the other being closer to the unclassified species results in incorrect identification.



Alg.	m	t	n	≡	≠	?	id	Result
tID	s	2.0	16		10	6		0 %
	d	2.0	26			26		0 %
	s	0.3	16	4			12	25 %
	d	0.3	26	24			2	92 %
bCM	s	2.0	16	5	11			31 %
	d	2.0	26	25	1			96 %
	s	0.3	16	4			12	25 %
	d	0.3	26	24			2	92 %
nN	s	-	16	5	11			31 %
	d	-	26	25	1			96 %

Figure 18: Results of the accession identification efficacy analysis based on K2P-distance of ITS45 sequence data. A: Multidimensional Scaling of the used distance matrix. Arrows indicate sequences of the species and respective individual sequences which could not be identified due to intraspecific distance outside the defined threshold (*D. moldavica* local (1) and external (2) accession AY506659 with a distance of 0.95 %). B: Determination of the optimal distance threshold. The plot shows false negative (blue) and false positive (red) species assignments in the range of 0 and 2 % distance threshold. The lowest cumulative error (ce = 2) was found between 0.16 and 0.44 % distance. C: Table with results of threshID (tID), bestCloseMatch (bCM) and nearNeighbour (nN) algorithms. With each threshold two analyses were conducted. One that included singleton species (s = s) and one where the respective sequences were duplicated (s = d). In column "n" the number of sequences is indicated. Thresholds (t) were 2.0 and 0.3 % for tID and bCM analysis. Results are shown in the last 5 columns: Identification success (≡), Identification failure (≠), Ambiguous result (?), No identification (id). The last column contains the accession identification efficacy values in percent.

b) Diagnostic Nucleotides

(1) psbA-trnH

Applying a sliding window analysis to psbA-trnH sequence data searching for diagnostic nucleotides (Figure 19, p. 60) all species but *D. nutans*, *D. moldavica* (and unclassified accession 8142) return windows containing positions unique to the respective species. We included the closely related unclassified accession (8142) into the *D. moldavica* group and analyzed diagnostic nucleotides in more detail. The group offers two regions within psbA-trnH for possible primer design. Three positions in the 5-primed region (pos. 64, 65 and 72) and one in the 3-primed region (pos. 370) following the psbA primer are unique for the *D. moldavica* group.

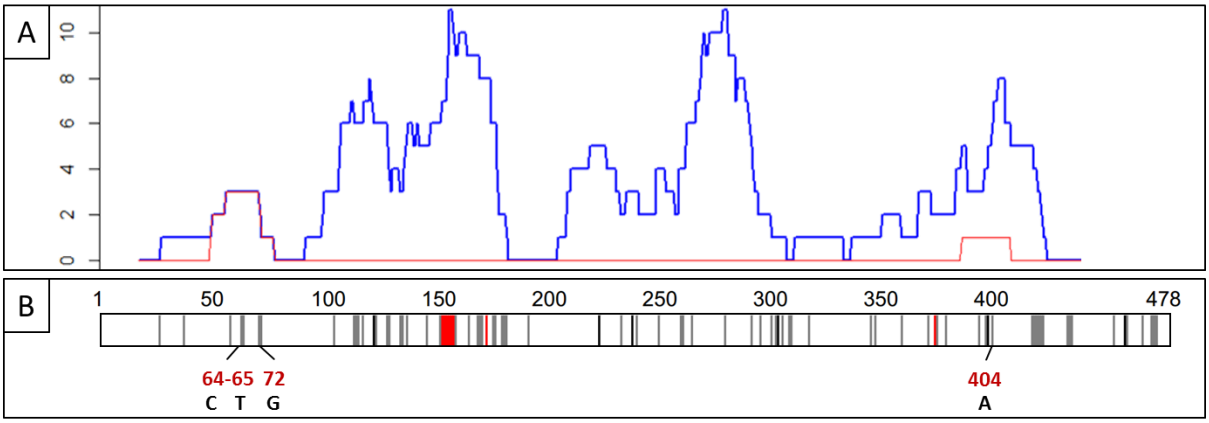


Figure 19: Diagnostic nucleotides in *Dracocephalum* derived from psbA-trnH inter-genic spacer sequence data. A: Sliding window (size = 25 bp, interval = 1) analysis of psbA-trnH sequence data displaying the number of diagnostic nucleotides within each window for all analyzed *Dracocephalum* species (blue) and for *D. moldavica* group only (red). B: DNA sequence fingerprint showing nucleotide variability at each position of the *Dracocephalum* psbA-trnH alignment. White areas are invariant while shading from light grey to black represents areas where 2, 3 or 4 variants are present. Red areas are alignment gaps (only one sequence with nucleotides in this area). Alignment positions unique for the *D. moldavica* group are indicated together with the respective nucleotide below the fingerprint.

(2) ITS

Applying a sliding window analysis to ITS45 sequence data searching for diagnostic nucleotides (Figure 20) all species but *D. parviflorum* returned windows containing unique positions. In the case of *D. ruyschiana* and *D. rupestre* only ITS2 and ITS1 regions respectively revealed diagnostic nucleotides. Although the unclassified accession (8142) could be authenticated by one diagnostic nucleotide in the ITS1 region we included it into the *D. moldavica* group and analyzed diagnostic nucleotides in more detail. The group offers four positions within ITS45 for possible primer design. Two positions in the ITS1 region (81 and 115), one in the 5.8S region (222) and one in the ITS2 region (625) following the ITS5 primer are unique for the *D. moldavica* group.

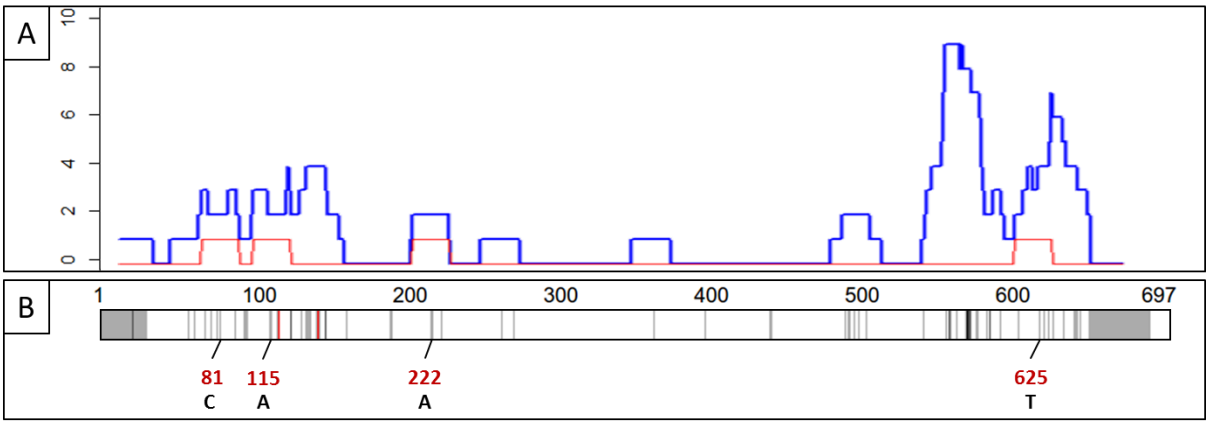


Figure 20: Diagnostic nucleotides in *Dracocephalum* derived from internal transcribed spacer sequence data. A: Sliding window (size = 25 bp, interval = 1) analysis of ITS45 sequence data displaying the number of diagnostic nucleotides within each window for all analyzed *Dracocephalum* species (blue) and for *D. moldavica* group only (red). B: DNA sequence fingerprint showing nucleotide variability at each position of the *Dracocephalum* ITS45 alignment. White areas are invariant while shading from light grey to black represents areas where 2, 3 or 4 variants are present. Red areas are alignment gaps (only one sequence with nucleotides in this area). Alignment positions unique for the *D. moldavica* group are indicated together with the respective nucleotide below the fingerprint.

c) Diagnostics of D. moldavica

Position 72 following the psbA primer was used to design an ARMS primer (psbA-72G) destabilized at the 3rd position from the 3-primed end by changing the guanine residue present in all known *Dracocephalum* accessions to a thymine. Using this primer in a multiplex PCR together with psbA-trnH primers an additional fragment of size 394 bp should be amplified in samples containing *D. moldavica* DNA and one with 417 bp in samples containing DNA from the unclassified accession (8142). The result of a multiplex PCR using the respective primers (Figure 21, p. 61) shows the verification of the described prediction.

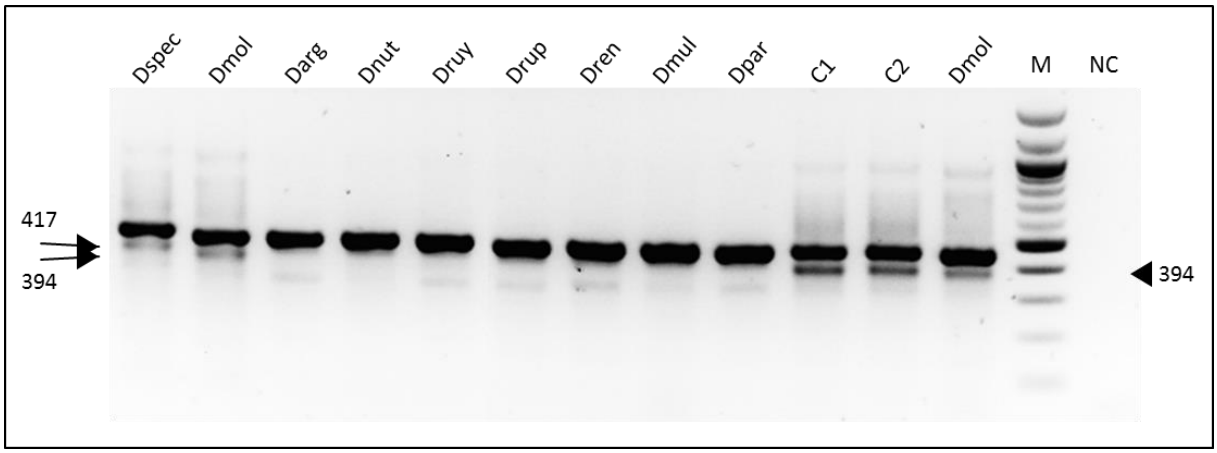


Figure 21: Multiplex PCR based on psbA-trnH using diagnostic ARMS primer psbA-72G to detect samples containing *D. moldavica*. A specific fragment for *D. moldavica* (Dmol = 394 bp) and *D. species* (Dspec = 417 bp) can be observed in all respective samples. C1 and C2 are derived from commercial teas containing *D. moldavica*. The size difference of the diagnostic as well as control fragment (psbA-trnH) between *D. moldavica* and *D. species* also can be observed. NC = no template control; M = 100 bp size marker (NEB)

VI. Discussion

A. Basic Studies

In our basic studies we established the methodological framework of PCR-RFLP and SCAR diagnostics, introduced a new database for managing reference collections either living or as seeds and addressed the question how processing of plant material in the form of DNA fragmentation effects results of PCR.

1. Interactive Plant Reference Database

The interconnectivity of plant accession data with general and specific information on the respective species represents the foundation of efficiency when working with a diverse pool of plant species. For example, there are over 1800 plants from 229 families known to be used in traditional medicine in north east India [Bhutani 2008] and about 7000 species used in Traditional Chinese Medicine [Hiller 2010]. Finding the most crucial cases related to genetic food diagnostics requires an enormous amount of information research that is aided by an integrative data management system. Furthermore, combining information of horticultural interest with scientific background information while both parties – the caretakers and the scientists – can input and share their experience improves cooperation and the value of work invested. Through collaborative ventures like 'The Plant List', a widely accessible working list of known plant species which has been developed and disseminated as a direct response to the Global Strategy for Plant Conservation (GSPC), it is today possible to adjust collections of hundreds of accessions to the most current taxonomic standard in little time. The integration of external data sources like 'The encyclopedia of life' [Wilson 2003] providing information on cultivation, geographical distribution, vernacular names, practical importance, ecological risks, toxicity, medicinal relevance and so forth, allows us to view a great portion of information instantly without searching for each piece individually. Through combination of internally and externally generated data comparative biological questions are easily discovered. As a data repository, information is preserved for the future and not lost with individuals leaving the local team. Extensions of the interactive plant reference database allow sample organization (tissue, herbar, DNA), barcoding organization (sequencing status of major barcoding markers) and chemical component organization using a chemical database and external data retrieved via ChemSpider [Pence 2010]. Future plans are the integration of seed list data from other botanical gardens with the intention of easily finding

replacement plants for accessions with unknown origin and the integration of endangered species information.

2. PCR: Effects of DNA Fragmentation

One of the first obstacles in establishing diagnostics based on PCR is the isolation of high quality DNA. Failure of PCR working with universal primers usually is related to quality of the DNA extract. Impurities (e.g. proteins, secondary metabolites) and DNA damages are the most frequent causes particularly common using extracts derived from plants rich in respective compounds and from processed material, respectively. In case of the latter, one might assume that if only highly fragmented DNA can be isolated, PCR failure is caused by the lack of DNA template molecules spanning the complete marker region.

Fragmentation of DNA induced by ultrasonic sound most likely leads to double strand breaks which do not influence amplification success as long as the break does not occur in the region of interest. The fragmentation series (Figure 1 A, p. 38) shows a gradual effect on fragment size. Fragment size distribution moves from “normal” sized genomic DNA to increasingly fragmented DNA while, even if the size of the main population passes the amplification size threshold (i.e. the size of the region of interest), a small undetected population of fragments with sufficient size is still present and permits a successful amplification (Figure 1 B and C). This notion is based on the assumption that the fragmentation process is gradual and the effect is strongest on the fragment population present in the highest concentration. In this case fragments present in low concentration evade detection by limited sensitivity of the used dye (SYBRSafe). High efficiency of PCR only requiring very few copies of the respective fragment further supports this hypothesis. Other explanations can be found in literature dealing with ancient DNA. For example a process called “jumping PCR” can result in the amplification of a fragment which size actually does not exist among the original template molecules. In this process partially extended primers are extended in subsequent cycles after annealing to a different fragment providing additional template information [Wilson 1989]. Another process that might contribute to a successful amplification from highly degraded DNA is termed “reconstructive polymerization”. During common PCR primers are abundantly available and polymerase is utilized by respective sites. However, in the absence of primers or due to lack of utilization caused by fragmentation or other DNA damages other sites formed by overlapping genomic fragments might get extended partially reconstructing genomic DNA [Golenberg 1996].

The two essential points from this experiment however are that even if we cannot see DNA in an agarose gel stained using SYBRSafe it does not necessarily mean that there is no template which can be used for amplification. Secondly, the fact that many DNA extractions derived from complicated plant material (e.g. dried or otherwise processed material) regularly resulted in highly fragmented DNA and subsequent PCR failure while amplification using respective fresh material succeeded, indicates that PCR failure most likely was a result of secondary metabolites inhibiting the polymerase or interacting with DNA and not related to DNA fragmentation. Ultimately however we also have to consider other DNA damages than mere double strand breaks that might prevent successful amplification [for review see Lindahl 1993].

3. PCR-RFLP: The 'Lemon Myrtle' Case

The taxonomy of the genus *Leptospermum* has been revised several times [Thompson 1989] and is still obscured by the coexistence of different synonyms, not to speak about the ambiguous common names used by commercial providers. We used a partial sequence of rbcL (rbcLa) proposed as one of the central molecular markers for plant genetic barcoding [CBOL Plant Working Group 2009] to define sequence divergences that might be useful for molecular diagnostics. We successfully obtained and verified rbcLa sequences for both *Leptospermum citratum* (GenBank: JN676919), and Backhousia citriodora (GenBank: JN676920). As to be expected, both sequences were highly similar. Despite the high conservation of the rbcLa marker between *Backhousia citriodora* and *Leptospermum citratum*, a molecular differentiation was possible by restriction analysis. We chose as target a Sac2 restriction site differing between the two species of 'Lemon Myrtle'. The restriction digestion should yield two fragments of 458 bp and 141 bp in *Leptospermum citratum*, whereas in *Backhousia citriodora*, the original 599 bp rbcLa amplificate would remain complete (Figure 2 A, p. 39). We tested this prediction using dried leaves from both species, amplified the rbcLa marker, and digested the amplified fragments using Sac2. The undigested rbcLa fragment was found to correspond to one band of around 600 bp in size for all species (data not shown). Upon Sac2 digestion, there was no change for the fragment of *Backhousia citriodora*, consistent with the prediction that, here, no Sac2 recognition site is present (Figure 2 B). In contrast, digestion of *Leptospermum citratum* yielded a larger fragment of around 450 bp, and a smaller fragment of around 150 bp, consistent with the prediction (Figure 2 A). A control, where the DNA template in the PCR reaction mix was omitted, was included as template negative control to check for potential contaminations. In addition to the two species of 'Lemon Myrtle' we tested *Dracocephalum moldavica*, which shares a similar citric flavor and is sometimes

used in combination with 'Lemon Myrtle' (e.g. commercial sample D_3, see Table 6, p. 89). Here, the PCR-RFLP pattern resembled that observed for *Leptospermum citratum*.

We finally tested whether this assay can be administered to commercial samples where typically dried leaf fragments of 'Lemon Myrtle' are blended with other herbal compounds (Table 6). Sample D_3 contained 'Lemon Myrtle' as second compound followed by 'Moldavian Dragonhead' and produced a pattern with three bands (Figure 2 C) of 600 bp (corresponding to the full-length rbcLa fragment), and 450 bp + 150 bp (corresponding to the fragments expected for *Leptospermum citratum*). The morphological analysis (data not shown) gave clear evidence for *Backhousia citriodora*, however. To test, whether the two smaller bands are actually originating from *Backhousia citriodora* or from the surrogate *Dracocephalum moldavica*, the leaf fragments recognized as 'Lemon Myrtle' were sorted from the mixture, extracted and analyzed separately. The resulting sample D_1 yielded only one band of 600 bp as to be expected for *Backhousia citriodora*, suggesting that the smaller bands originated from the surrogate *Dracocephalum moldavica*. Sample D_4 contained 'Lemon Myrtle' as one of the minor compounds and produced a band of 600 bp, which was consistent with the results of the morphological analysis clearly identifying *Backhousia citriodora*. Again, these fragments were sorted and analyzed separately yielding D_2 that produced a single band of 600 bp as expected. In sample D_5, 'Lemon Myrtle' (identified morphologically as *Backhousia citriodora*) was a major component. Again, a single band of 600 bp was observed. In the last sample, D_6, 'Lemon Myrtle' the morphological features identified *Leptospermum citratum* as major component, consistent with the presence of the 450 bp and 150 bp bands. The upper 600 bp band, corresponding to the undigested fragment, is most likely originating from the other components that do not harbor a corresponding Sac2 restriction site (data not shown).

As exemplary case study for many other cases, where plants used in traditional medical or culinary cultures are transferred as trend food supplements to the Western markets, we have analyzed 'Lemon Myrtle', actually two distantly related species of the Myrtaceae derived from the culture of the Australian Aborigines. By molecular diagnostics using restriction fragment length polymorphisms in the rbcLa marker we can discriminate between *Backhousia citriodora* and *Leptospermum citratum* in dried mixtures as typically encountered in commercial samples. Principally, this approach should be transferrable to more processed samples as well (such as 'Lemon Myrtle' flakes), where anatomical features are difficult to be assessed. However, in the presence of *Dracocephalum moldavica* (which in some samples is added due to its similar lemon-

type flavor), the restriction pattern hampers the discrimination between *Backhousia citriodora* and *Leptospermum citratum*. In those cases, it will be necessary to accompany the restriction analysis by microscopic investigation, or, in case that anatomical features have been lost due to strong processing, additional molecular markers. In the meantime microscopic as well as molecular markers are also available for the detection of *Dracocephalum*.

4. SCAR Development: The *Clematis* Complex

The naming of phytopharmaceuticals usually includes a taxonomic indicator accompanied by the part of the plant which is used (e.g. Matricariae flos = Flowers of *Matricaria chamomilla*). Original names of the two Traditional Chinese Medicines “mù tōng” (Akebiae Caulis) and “chuān mù tōng” (Clematidis Armandii Caulis) imply that they are related. While “mù tōng” is made from caulis of *Clematis armandii* or *C. montana* “chuān mù tōng” is derived from *Akebia quinata*, *A. trifoliate* or *A. trifoliata var. australis*. Furthermore substitutes for “mù tōng” are sometimes *C. armandii* or *C. montana* and for “chuān mù tōng” – among several other *Clematis* species – the standard species of “mù tōng”. More importantly for both medicinal products two species of *Aristolochia* are reported to be potential adulterants (*Aristolochia manshuriensis* and *Aristolochia moupinensis*).

The history of the medicinal herb “mù tōng” elaborates the confusion that might arise due to different naming strategies used in Traditional Chinese Medicine. Zhu [2002] reviewed historical records and revealed that the original source of “mù tōng” had been *Akebia* species until the 17th century. After that until the early 20th century *Clematis* species (“chuān mù tōng”) were used and only after the 1950s *Aristolochia manshuriensis* (“guan mù tōng”) became the main source. The name “mù tōng” seems to describe a habitat rather than a taxonomic group. Mù (木) translates to ‘wood’ or ‘tree’ while tōng (通) translates to ‘to pass’, ‘to get through’ or ‘to be linked to’. Although other sources (Wikipedia) suggest that “mù tōng” translates to ‘perforated wood’ it might be deduced that since all three species are climbers and inhabit forests “mù tōng” better translates to ‘linked to wood’.

Indicated by an incidence, where more than 100 women in Belgium lost their kidneys after a TCM based slimming regime [Vanherweghem 1993], the presence of components common in members of *Aristolochia* can have severe consequences for human health. Our aim was to establish a marker to detect the presence of *Aristolochia manshuriensis*. We used plants of the genus *Aristolochia*, *Akebia* and *Clematis* available at the time in a RAPD analysis. Our goal was to isolate RAPD fragments unique for our main reference plant (*Aristolochia manshuriensis*) and utilize the

contained sequence information to design diagnostic primers. Although the final assay (Figure 3 C, p. 40) using SCAR primers specific for our reference plant shows clear amplification using the respective sample and none using templates of the other species, there are several critical points to consider. During the RAPD analysis some of the samples consistently showed only weak (t_2 and t_5 in Figure 3 A) or no amplification (s_1 and t_3 in Figure 3 A). While weak amplification in some cases using different primers not necessarily indicates problematic template DNA, no amplification at all does. Before any final conclusions about the specificity of the primers can be drawn, the assay has to be repeated using new DNA extractions that have been verified to be working over a range of different template concentrations (i.e. characterized by stable fragment patterns). Additionally, since the plant which we received as *A. manshuriensis* later was determined to be *A. longgangensis* it might be a more reasonable approach first to develop primers that are able to detect DNA from the genus *Aristolochia*. For that the approximately 1 kb sized fragment present in both *Aristolochia* species (a_1 and t_{12}) could be used (upper black horizontal line in Figure 3 A). Finally, the design of SCAR primers using RAPD fingerprinting usually is based on terminal sequence information including the RAPD primer sequence. The presence of a RAPD fragment is a result of successful annealing and extension of the used primer. In this particular case however the final SCAR primers have been derived from internal regions of the respective RAPD fragment. This and the absence of amplification in *A. sempervirens* (t_{12} in Figure 3 C) might indicate that the specificity of RAPD fragments is not necessarily limited to the original RAPD primer sites. To investigate this hypothesis the SCAR primers should be tested with as many close related species as possible as well as other individuals of the same species.

The term ‘Sequence Characterized Amplified Region’ seems to be very roughly defined which makes it hard to generalize about the usefulness in genetic food diagnostics. Evaluating the literature on so called SCAR markers, it appears to be reasonable to differentiate between those that have been designed on well characterized amplified regions [Paran 1993] and those where characterization has been mostly superficial or completely avoided [Wang 2001, Dnyaneshwar 2006, Lee 2006, Huh 2006, Devaiah 2008, Theerakulpisut 2008]. In most of these cases only a limited number of individuals from the respective species have been used to retrieve a specific RAPD fragment. Additionally, the mere absence of a respective sequence in GenBank has been used as verification of specificity. The lack of sufficient characterization together with the absence of verified marker stability within the respective species using a sufficient sample highlights the enormous gap between claims of and scientific proof for species specificity in peer-reviewed journals.

B. Case Study: *Dracocephalum* L.

1. Product Authentication using PCR-RFLP and ARMS

Since morphological traits of *Dracocephalum moldavica* were too variable to serve for diagnostic purposes and the microscopic authentication based on the ratio between epidermal pavement cells and subtending palisade cells is somewhat cumbersome, we explored the rbcLa marker as base for genetic authentication. By genomic PCR we amplified rbcLa and obtained valid sequences for 3 accessions of *D. moldavica* L., 1 accession of *D. ruyschiana*, 1 accession of *M. officinalis* L., and 1 accession of *N. cataria* L. All these sequences were checked by a second run, and the taxonomic identity of all accessions had been verified before extraction of DNA. As to be expected, these sequences were highly similar, but several base exchanges were detectable, especially between the *Dracocephalum* accessions and those of *M. officinalis* and *N. cataria*. To evaluate these differences, we located the sequences with respect to other members of the *Nepetoidea* subfamily drawn from GenBank using *Stachys sylvatica* and *Ocimum basilicum* as outgroup based on the neighbour-joining algorithm. All three accessions of D. moldavica L. yielded identical sequences and clustered together with the D. ruyschiana L. in a separate *Dracocephalum* clade (data not shown). There were two sequences deposited in GenBank for both species. Whereas the sequence for *D. ruyschiana* L. was very similar, the three sequences isolated by us significantly differed from the sequence Z37389.1 deposited in GenBank and derived from a phylogenetic study on the *Nepetoidea* subfamily [Kaufmann 1994]. Whether these differences represent intraspecific variation or whether different species of the genus are involved is unclear, because no information on the identity or voucher references had been reported in that study. Irrespective of this minor detail, the reconstructed phylogeny shows a clear separation of *Dracocephalum* from its surrogate species *M. officinalis* and *N. cataria*. For both of these species, sequences were already available in the database and found to be identical with those isolated by us.

In the next step, the sequences were analyzed for differential restriction sites (Figure 4, p. 41). In fact, two of such sites could be identified. A base exchange of A for G at position 397 of rbcLa in *Nepeta cataria* L. eliminated a recognition site for EcoRI present in both *M. officinalis* and *Dracocephalum* (and also in all other *Nepetoidea* sequences analysed). Second, a base exchange of G for T at position 442 of the rbcLa in all *Dracocephalum* accessions analysed eliminated a recognition site for BamHI present in all other available *Nepetoidea* sequences including *M. officinalis* and *N. cataria*.

Based on these two diagnostic differences in restriction sites, a RFLP assay was designed. Digestion of the rbcLa fragment with BamHI was predicted to generate one band in *Dracocephalum* corresponding to the uncut fragment. In contrast, restriction should yield two smaller bands of 445 bp and 155 bp for *Nepeta* and *Melissa* (Figure 5 A, p. 42). This prediction was tested experimentally using pure and commercial samples of Dragonhead and its potential surrogates (Figure 5 B). As predicted from the sequence analysis, restriction of the rbcLa fragment yielded one band of around 600 bp for samples from *D. moldavica* L. or commercial samples declaring Dragonhead, whereas two bands of the predicted size (445 bp and 155 bp) were observed for the surrogate species *M. officinalis* L. and *N. cataria* L as well as *N. nuda*. To corroborate the validity of the approach, the commercial samples were sorted using the microscopic features given in Table 12 confirming the result obtained for pure samples of Dragonhead. Thus, RFLP based on restriction with BamHI allows detecting adulterations of Dragonhead by other *Labiatae* in commercial samples. However, in case of adulteration it would not be possible to tell, whether the surrogate is *M. officinalis* L. or a member of the genus *Nepeta*. A second drawback of this approach is that a failure of the restriction (for instance due to degradation of the BamHI enzyme) would also produce one uncut band indicative of Dragonhead which would leave adulterations gone unnoticed.

We therefore tested RFLP based on restriction of the rbcLa fragment with EcoRI predicted to yield one band of around 600 bp in members of the genus *Nepeta* (corresponding to the uncut rbcLa fragment), whereas both *D. moldavica* L. and *M. officinalis* L. should exhibit two bands of 395 bp and 205 bp (Figure 5 C). This prediction was verified and confirmed experimentally (Figure 5 D). Thus, RFLP based on restriction with EcoR I allows to unequivocally detect adulteration by *Nepeta* species. However, the two bands at 395 bp and 205 bp would not be unequivocal proof for the presence of Dragonhead, but would also be produced in case of adulteration with *M. officinalis* L. The advantage of that strategy over RFLP using BamHI is that the indication of Dragonhead in the sample is safeguarded against a failure of the restriction digest, since the characteristic double band is observed only for successful restriction.

Both RFLP-based strategies described above have their specific advantages and drawbacks: In one case (BamHI) Dragonhead can be clearly discriminated against adulteration with either *Melissa* or *Nepeta*, but failure of restriction digest would leave adulterations gone unnoticed. In the alternative strategy (EcoRI), the detection of Dragonhead is safeguarded against restriction failure, but the presence of the characteristic double band would also be obtained in case of adulteration

with *Melissa*. A general drawback of the RFLP approach is that it requires a two-step protocol: first, the rbcLa marker has to be amplified by PCR, and the obtained fragment has then to be digested overnight.

To overcome these drawbacks of RFLP-based strategies, we employed an alternative approach termed Amplification Refractory Mutation System (ARMS). This method is based on a multiplex PCR, whereby one intermediate primer will generate a second, smaller band in addition to the complete fragment. This intermediate primer is designed such that annealing is destabilized by introducing bases that are illegitimate with respect to the target sequence. In case of mutations in the target sequence this destabilization will prevent the intermediate primer from annealing such that the side band will not be observed. This strategy originally has been developed to rapidly screen populations for mutations in specific target sequences [Newton 1989]. However, if the sequence in the species of interest (in our case Dragonhead) is used as template for the ARMS design, any adulterant with even minor changes in the target motif would become detectable by its failure of producing the side band. A major advantage of the ARMS strategy over RFLP is that any failure of amplification itself would be immediately detected by the absence of the full-length band. The second major advantage is that no second restriction step is required – the result is obtained immediately after PCR.

We therefore designed a diagnostic ARMS primer for the rbcla sequence of dragonhead that should anneal 159 bp upstream of the rcbla reverse primer. To introduce destabilization of the 3'-end, a base exchange from G to A was introduced into the diagnostic primer (DC4) compared to the dragonhead target sequence. A multiplex PCR using this primer in combination with the two conventional rbcla primers (Figure 6 B, p. 42) should produce, in addition to the full-length fragment at 599 bp, a second band at 159 bp in a situation where the ARMS primer annealed to its target sequence (Figure 6 C). When this was tested experimentally (Figure 6 D), the multiplex PCR using primer DC4 produced the predicted diagnostic second band at 159 bp for both pure and commercial samples of *D. moldavica*, whereas for *M. officinalis* and two tested species of *Nepeta*, only a single band at 599 bp was observed. Thus the ARMS strategy delivered a single-step protocol that by the presence of the rbcla band simultaneously reports on the success of the PCR and the presence of dragonhead in the sample. It should be kept in mind that both ARMS and RFLP probe for specific species and thus are based on a hypothesis on the nature of potential adulterations and sequence information from potential adulterants. Moreover, the presence of more than one adulteration will create complex outcomes that are difficult to interpret. Thus,

fields of application for the ARMS strategy might be large sets of similar samples that have to be screened for specific adulterations.

Using common molecular markers to identify certain taxa is a frequently used method. However, similarly to the criticism already expressed for so called SCAR markers (p. 66) in many of these cases [Ma 2000, Lu 2005, Park 2006, Wang 2007, Chiou 2007] little characterization of the marker with respect to intra- and interspecific variation, including adequate taxonomic and distributional sampling, has been done and exceptions [Shiba 2006, Li 2007, Yang 2007, Kitaoka 2009] are less frequent.

2. Taxonomic Verification

a) Seed Morphology

Seed characteristics have been used for taxonomic identification [Grillo 2010] and can be correlated to the ecology of respective species [Westoby 1992]. Variation within species is mainly attributed to environmental effects during development rather than genetic differences [Vaughton 1998]. The aim of our comparative analysis was to screen for significant differences to identify other cases of mislabeling. Using circularity (CS) and the directly related length-width-ratio (LWR) seeds of *D. ruyschiana* and *D. argunense* accessions displayed significant differences with the exception of 8252 (*D. argunense*) which reflected characteristics more similar to *D. ruyschiana* accessions (Figure 8, p. 47). Characters of accession 7900 (*Nepeta* in Figure 7, p. 46) further supported the already established misidentification as *D. argunense*. Correlating LWR with CS of *D. argunense* accessions (Figure 9 left, p. 47) indicated the presence of seeds with LWR ratios and CS values outside the range of the main population. This might be related to abnormal seed development characterized by long and thin seeds. The same correlation made with *D. ruyschiana* seeds (Figure 9 right) shows an increased distributional range of data points along the regression line which might indicate either higher variation among the different accessions or increased developmental variation within *D. ruyschiana*. While circularity values of accessions 8174 and 8141 (Figure 10, p. 48) support the first theory, assessing the correlation of single accessions (data not shown) revealed that all accessions are in fact characterized by an increased range and that seeds of accession 8174 behave differently in that they are scattered more widely below the regression line with CS values smaller than 0.8. With the correlation of LWR and CS we can see a development from long thin seeds to more wide (more circular) seeds. Since we could not observe the same range in *D. argunense* we have to assume that either there is a temporal difference

which decreases the chance of collecting seeds that are not fully developed or selection after harvesting of the seeds took place. Although only two accessions of *D. argunense* are considered the fact that they are from different botanical gardens and show similar distributions indicates a significant difference in seed development. Comparing data sets of other species with sufficient observations an increased range of LWR and CS values can be observed that is more similar to *D. ruyschiana* accessions than to accessions of *D. argunense*. Correlation of recently acquired seeds from one accession of *D. austriacum* shows a slightly decreased distributional range but similar CS values compared to *D. ruyschiana* (data not shown). It will be interesting to see how genetic data from *D. austriacum* correlates with those of *D. ruyschiana* and *D. argunense*.

In the group of *D. grandiflorum* and *D. rupestre* accessions the determined misidentification of accession 8254 displaying significantly different leaf shape was further supported by seed character deviation especially pronounced in circularity (Figure 10). Seeds of *D. grandiflorum* / *D. rupestre* are considerably longer than wide and thus have low circularity values (0.6) while *D. ruyschiana* (0.7 - 0.8) and *D. argunense* (0.85) are increasingly more circular. Accession 8254 clearly fits into the range of *D. ruyschiana*. High similarity within the group of *D. grandiflorum* and *D. rupestre* (excluding 8254) in addition to the fact that they are from different origins suggests that either the two species are very close related or that actually only one species is represented. Some literature [Rothmaler 2008] indicates that most of the as *D. grandiflorum* cultivated accessions are actually *D. rupestre*. The two chosen representatives of *D. ruyschiana* on the other hand display the range of circularity variation found in this particular group. It is also worth mentioning that there is some variation within seed size (area) among putative *D. rupestre* / *D. grandiflorum* accessions which permits the definition of two groups partially supported by genetic data (Figure 14, p. 51).

Seed characters are at least in some cases good indicators to approach taxonomic verification. The software SmartGrain [Tanabata 2012] supplies 9 characters that allow a solid analysis. In our study we found several cases, where seed characters were able to identified miss identifications (7900, 8252 and 8254). Since our main source of reference plants is a seed exchange program, the first instance accessible to quality control is the seeds. Collection and characterization of seed data could be used to assess taxonomic classification prior to cultivation. Considering that one would need to wait months or even years before being able to morphologically determination the respective accession, only to realize that it is not the expected species, using characters that are immediately available does seem to be a reasonable approach. There are of course some

limitations to consider. Firstly, if seeds have to be removed from fruits or other appendixes prior to analysis, this would extend the time of analysis considerably. Secondly, seed color, if not homogenous, in some instances can prevent analysis using SmartGrain (*Oryza*, A. Grimm personal communication). Finally, to be able to reliably diagnose on the species level more information might be required. Along with the characters obtained using SmartGrain representative microscopic pictures of the seed coat (testa) can increase the efficacy of species level diagnostics [Fagúndez 2004, Oh 2008, Dadandi 2009]. The effort to establish a seed character database in the long run surely would benefit an exchange network like the international plant exchange network (IPEN) by increasing the rate of correct identity among the offered seeds. Additionally, the generated data could be used for further scientific studies, promoting cooperation among the involved institutions. E.g. by comparing seed data of accessions that have been cultivated in botanical gardens with data from respective wild relatives, the effect of long time cultivation could be studied contributing viable insights for ex-situ conservation projects.

b) Tubulin based Polymorphism

A relatively new DNA fingerprinting technique to infer genetic diversity is based on intron length variation in the first intron of plant β-tubulin genes – Tubulin based Polymorphism (TBP) [Bardini 2004]. The development of the degenerated TBP primers is based on characterized β-tubulin genes all containing a first intron flanked on both sides by fairly conserved coding regions. The value of application is derived from the fact that between different β-tubulin isotypes intron length differences have been observed while additional value is gained by the potential presence of different numbers of isotypes.

We used TBP for additional verification of the determined misidentifications and to evaluate the usefulness of this technique as tool for species identification among reference plants. All species included in this study showed differences in their TBP pattern related to intron size and most likely number of isotypes [data not shown]. In the group of *D. grandiflorum* and *D. rupestre* accessions no differences could be observed. Additionally, while all *D. moldavica* accessions showed identical patterns [data not shown] among *D. ruyschiana* two different patterns could be observed characterized by one additional fragment present in 3 of the analyzed accessions (No. 4 in Figure 11). Two fragments of *D. ruyschiana* accessions appear to be shorter versions of fragments found in *D. argunense* while another appears to be a larger version in *D. ruyschiana*. Since the number of β-tubulin genes in flowering plants can be as high as 20 [Oakley 2007] the presence of additional isotypes appears possible. However, further investigations are necessary to determine if the

fragment represents another β-tubulin isotype or is absent in other accessions because of differences in the priming site. Other differences expressed by variations in concentration of the respective fragments are more likely to be related to differences in the priming site or are the result of impaired quality of the used DNA template. Using different template concentrations showed effects ranging from slight differences in fragment concentration to almost complete failure of amplification [data not shown]. The most successful template concentration was 50 ng / µl. To further evaluate the amplified patterns sequence data of the respective introns could be retrieved and analyzed for regulatory functions while expression studies could be used to support the findings. It also would be interesting to determine if there is a morphological or physiological effect of β-tubulin intron variation reflected by differences in the respective plant accessions. Finally, the putative presence of additional isotypes within the same species would to some extend compromise the utility of TBP as species specific marker. However, as long as the overall TBP pattern is stable within a species and enough variation in intron size and number of isotypes is present between different species the effect of additional isotypes within a species should remain insignificant.

There are good indications that TBP is a useful tool to differentiate between species. In our case we used agarose gel electrophoresis and SYBRsafe staining. By increasing the resolution power and staining sensitivity applying polyacrylamide gel electrophoresis and silver staining this method can be far more informative as shown in the original TBP publication [Bardini 2004] and subsequent ones [Gavazzi 2012].

c) Genetic distance

Neighbor Joining (NJ) is an algorithm that takes a matrix of distances calculated between pairs of sequences and builds a tree by iteratively finding closest sequences, joining them at a new node and using this node as starting point for the search of the next closest neighbor. This is continued until the complete tree is resolved and all branch lengths are known. NJ trees are generally un-rooted and best visualized by rooting them first or using a radiation tree that shows a more accurate distance between the sequences. For the construction of NJ trees (Figure 13 and Figure 14, p. 51 ff.) we first conducted a classical Multidimensional Scaling analysis (MDS, syn. Principal Coordinates Analysis) using data of all available *Dracocephalum* accessions and closely related species from other genera identified by blast search. *Schizonepeta* and *Hyssopus* showed to be the closest genera of the genus *Dracocephalum* analyzing psbA-trnH and ITS data. Although in the ITS tree situated within the *Dracocephalum* cluster *Hyssopus* (ITS: > 3 %) and all other genra (ITS: > 8

%, psbA-trnH: > 10 %) are well separated. Already the construction of NJ tress indicated the superior resolution of psbA-trnH sequence data which is why we focused sequencing efforts on this marker. Clusters formed by the NJ approach confirmed all suspected misidentifications (Figure 14). Accession 8254 originally classified as *D. rupestre* and accession 8252 originally classified as *D. argunense* are located within one of two *D. ruyschiana* clusters sharing high sequence similarity with accession 8141 and 8174. The second cluster is formed by accessions 5156, 8011 and 8261 of which 5156 had been morphologically determined prior to this study. The fact that both clusters are separated by approximately 2 % sequence divergence and the remaining *D. argunense* accessions (represented by 8162) are closer to one of these clusters introduces room for speculation. *D. argunense* might be a paraphyletic group originating from an isolated group of *D. ruyschiana* populations. Results of seed character analysis (Figure 8, p. 47) and TBP pattern comparison (Figure 11, p. 49) strongly support the relation between the two *D. ruyschiana* clusters and the separation of the *D. argunense* group. It might be interesting to include *D. austriacum* sequence data as another member of this section distributed in overlapping regions with *D. ruyschiana*. Another very interesting aspect is the close relation of *D. parviflorum* – the only North American member of the genus – to one of the two *D. ruyschiana* clusters. This could indicate that there are two evolutionary distinct groups of *D. ruyschiana* one originating from the very east of Northern Asia where also *D. argunense* originated and the other group being the European / Western Asia representation at least geographically isolated from the other.

Both trees also indicate internal biogeographic groups. One containing *D. ruyschiana and D. argunense* distributed in the northern parts of Asia, Europe and the alpine regions of central and Eastern Europe as well as *D. parviflorum* distributed in North America. The second group is found in central Asia (*D. rupestre*) and the third is distributed from North Africa (*D. renati*) over western Asia (*D.kotschyi*, *D. multicaule*) into central Asia (*D. integrifolium*, *D. moldavica*).

3. Marker Evaluation

In this part of the study we analyzed three markers of the plastid (rbcL, matK and psbA-trnH) and one of the nuclear genome (ITS) derived from 10 species of *Dracocephalum*. Among the plastid markers amplification and sequencing success of approximately 100 % was best using rbcLa and psbA-trnH while matK-KIM with 92 % amplification success and 95 % sequencing coverage as expected proved to be challenging in some cases. With the exception of psbA-trnH (363 – 425 bp) no length variation was observed in the plastid markers. Amplification and sequencing success was moderate using ITS45. The cause can be deduced from the fact that respective primers originally

were designed for fungi [White 1990] which in some cases led to the amplification of additional fragments. Sequencing of respective fragments confirmed the presence of DNA of an epiphytic fungus (*Pseudozyma aphidis*) and of a nematode (*Phasmarhabditis hermaphrodita*), respectively. The nematode is frequently used as biological molluscicide but should not occur in DNA extracts from leaf material. By maximizing efforts to avoid contaminations the success rate of ITS45 can be increased. However, when the frequent occurrence of DNA from endo- or epiphytic fungi is to be expected, divergent paralogues of ITS require cloning of multiple copies or secondary structures lead to poor sequence data [Kress 2005], it might be advisable to use more specific primers or a more suitable marker, respectively.

a) Separability and Heterogeneity

TaxonGap uses similarity of a set of sequences to assess the separability between different species and heterogeneity within them. Reviewing the results from our TaxonGap analysis a minimum separability is indicated by a vertical black line in the plastid marker psbA-trnH (Figure 15, p. 55) and in two parts (ITS1 and ITS2) of the nuclear ITS marker (Figure 16, p. 57). Considering that in our data set only a small fraction of the around 70 known species of *Dracocephalum* are included strongly suggests that the primary barcoding markers (rbcL and matK) are very likely to fail in their role as specimen identification markers within *Dracocephalum*. Similar results can be found in other studies of *Lamiaceae* genera [De Mattia 2011, Guo 2011, Theodoridis 2012, Wang 2013, Federici 2013]. The data also suggests that with low divergence in the two coding regions the genus seems to contain groups with very recent speciation events. Exemplified by psbA-trnH, the difference in accumulation of mutations within inter-genic and spacer regions compared to coding regions becomes quite obvious. The main conclusion from this study is that for the purpose of species level diagnostics in *Dracocephalum* and most likely many other members of the *Lamiaceae* although coding regions are more desirable (stable length) both ITS spacers and the inter-genic psbA-trnH region are preferred candidates.

One particular point on the generation of the similarity matrix using MatGAT however has to be mentioned. In contrast to MEGA which uses pairwise or complete deletion of ambiguous and missing data (which also includes internal gaps) when comparing two sequences, MatGAT uses all available information. This has on the one hand the positive effect of more accurately describing the relationship between two sequences of introns and inter-genic spacers but on the other hand requires high quality sequence information that does not contain missing data. MatGAT does not recognize the “?” symbol as indicator for an unknown nucleotide thus handles it like a gap which

leads to biased similarity values as can be observed in the case of *D. moldavica* (heterogeneity, Figure 16).

4. Species Level Diagnostics

a) Identification Efficacy

We performed identification efficacy analyses using psbA-trnH and ITS45 data respectively. The respective distance matrix used for analysis was the same as for tree construction with the Neighbor Joining approach (Figure 13 and Figure 14, p. 51 ff.). Basic principle of the three used algorithms is to test for each sequence, whether the nearest (nearNeighbor = nN), the closest sequence within a defined threshold (bestCloseMatch = bCM) or all sequences within a defined threshold (threshID = tID), excluding the tested sequence, share the same species name. A respective threshold can be chosen arbitrarily (e.g. 2 % is used at the barcoding life data systems web site) or by minimizing the identification error rate over a range of thresholds. In the latter case for each threshold false positive and false negative results acquired represent the cumulative error. For our dataset the lowest cumulative error was 2 (false positives) for both psbA-trnH (threshold between 0.25 and 0.77 %, Figure 17 B, p. 58) and ITS (threshold between 0.16 and 0.44 %, Figure 18 B, p. 59) data. The optimized threshold chosen was 0.5 % and 0.3 % respectively. Using singletons (species represented by only one sequence) inevitably leads to identification errors because no representative of the species remains in the dataset. This and the loss of information by removing the respective sequences from the analysis are the reasons why we chose to analyze an additional dataset in which singleton sequences had been duplicated.

The identification efficacy using psbA-trnH sequence data (Figure 17 C) reached 100 % applying nN and bCM (2 % threshold) algorithms. Removing singletons from the analysis, here, had the same effect like duplication. Using a 2 % threshold with tID delivered the worst result with only 28 % efficacy in the normal dataset and 20 % using duplicated singletons. This reflects the increased inclusiveness of tID in considering all sequences within the threshold and the low divergence rate of the marker compared to COI with more than 10 % in animals [Erickson 2008]. A single sequence from another species within this threshold leads to an ambiguous identification result. As soon as we applied an optimized threshold both respective algorithms presented the same identification efficacy with 50 % using the normal dataset and 92 % using duplicated singletons. In this situation there was a relative difference between using duplicated singletons and removing them from the analysis. While the standard data set yielded 50 % efficacy, removing singletons would result in 82

%. By duplication the efficacy reaches 92 %, 10 % more than by removing the respective sequences from the analysis. One can now argue that both strategies lead to biased identification efficacy. Especially in case of an arbitrarily chosen threshold the removal of any sequence that falls within that threshold would lead to an overestimation of efficacy. Singleton duplication however simulates the presence of an identical sequence and retains available interspecific information. Furthermore the efficacy value is a result of testing any number of sequences disregarding if there are 10 haplotypes of one species and only 2 for several others. It is therefore preferable to calculate the efficacy in terms of species identification rather than specimen identification because we evaluate the marker and its ability to identify species. Applying this to the psbAtrnH data with duplicated singletons (10 species) leads to 80 % species identification efficacy in case of the optimized threshold where one sequence of *D. ruyschiana* (8141) and one of *D. rupestre* (FJ513110) could not be identified. In case of the 2 % threshold using bCM and nN the results remain true (100 %) since all species could be identified.

Results from using ITS45 data (Figure 18) only differed in efficacy values. Using a 2 % threshold tID fails completely which reflects the lower distance values of ITS45 sequences compared to psbA-trnH. By removing 5.8S rDNA data with its low separability values (Figure 16, p. 57) from the ITS45 sequences before calculating the distance matrix might lead to increased identification efficacy. In our case however removing the 5.8S rDNA data would lead to the loss of already rare non-singleton species caused by lack of differences in one of the two spacers combined with the elimination of differences found in the rDNA region. The sequence of accession AY506659 (*D. moldavica*) cannot be identified neither using nN nor bCM (2 % threshold). When using an optimized threshold (0.3 %) all two *D. moldavica* sequences present in the dataset cannot be identified. Considering the 12 species represented by ITS45 data and only one that could not be identified results in 92 % species identification efficacy independent of the used algorithm.

The tID algorithm clearly needs optimized thresholds to return useful information especially when working with plant taxa. With a sufficiently high threshold bCM returns basically the same results as nN which is reasonable because nN takes the closest match considering all sequences and as soon as the threshold includes all sequences bCM inevitably makes the same choice. As soon as optimized threshold values are used the result of bCM is basically the same as that of tID. This is because the threshold optimization is based on results from tID rather than bCM leading to a threshold that is best suited for an algorithm that includes all sequences within a given threshold while determining identification success.

All tested algorithms use genetic distances to determine identification success. Particularly in the early days of DNA Barcoding criticism about the approach in general and the use of genetic distance rather than character based data was common [DeSalle 2005]. One of the criticisms was that there is probably no universal distance cut-off for the delineation of species and that the process of finding cut-offs will therefore be very subjective. While the first part is true the second is unsubstantiated. In any type of approach finding delineation measures after addition of a new member to the group the measure might have to be adapted. Additionally, using threshold optimization is pretty straight forward and based on scientific data rather than subjective choice. However, there are also character based methods available (e.g. BRONX (Barcode Recognition Obtained with Nucleotide eXpose´s) [Little 2011], CAOS (Characteristic Attribute Organization System) [Sarkar 2008], DNA-BAR [DasGupta 2005] and BLOG system (Barcoding with LOGic formulas) [Bertolazzi 2009]) which might be useful for future studies.

b) Diagnostic Nucleotides

With the sliding window analysis of diagnostic nucleotides we applied a character based method to display regions within ITS (Figure 20, p. 60) and psbA-trnH (Figure 19, p. 60) that harbor diagnostic nucleotides in different numbers. With a window size reflecting the size of a primer (25 nucleotides) one can approximate the regions where most of the diagnostic primers will have to be designed. By overlaying the results of the genus analysis with the regions found in an individual species, specific diagnostic characters can be highlighted. Due to insufficient length differences of amplified fragments, regions with high density of diagnostic nucleotides (e.g. peak at pos. 575 in Figure 20) might be less suited for multiplex approaches that aim to diagnose several different species in one reaction. Such an approach would need to use separate reactions for each species using only one diagnostic primer at the time. We used one of the diagnostic characters found in the psbA-trnH region to design an ARMS primer and showed its potential use in discriminating between *D. moldavica* and other species of the genus (Figure 21, p. 61).

c) Diagnostics of D. moldavica

In our first approach (Figure 5 and Figure 6, p. 42) we designed a diagnostic primer based on the rbcLa marker. This allowed us to distinguish samples of *Dracocephalum* species from those of *Melissa* and *Nepeta*. After studying additional marker regions from both the plastid and nuclear genome we found that the inter-genic spacer psbA-trnH provides superior intra generic resolution followed by the internal transcribed spacer region (ITS). To test if we can proceed from inter to

intra generic diagnostics we designed an ARMS primer based on a specific nucleotide difference unique in the analyzed *D. moldavica* accessions and one very closely related (unclassified) accession. We evaluated the specificity of the approach including representatives of D. moldavica (2 reference accessions and 2 commercial products), 7 other *Dracocephalum* species and one unclassified accession. In the resulting electrophoresis (Figure 21, p. 61) specific fragments can be observed in all samples containing *D. moldavica* shortly below the control fragment. The size difference of the control and diagnostic fragment between the unclassified accession and other *D. moldavica* accessions also can be recognized. Although additional weakly amplified fragments can be observed in other species, they are of smaller size than the expected diagnostic fragment (394 bp / 417 bp) thus, do not compromise the authentication.

During the last decade, DNA Barcoding has been developed as new approach to address genetic identity in numerous organisms. DNA Barcoding is based on sequence diversity in short standardized gene regions of taxonomically pre-defined groups. It is used to identify species in food, wildlife trade, environmental monitoring projects, the validation of processed wood and lumber products and forensic investigations. For the animal kingdom, the mitochondrial cytochrome c oxidase I (COI) with more than 10 % divergence among species of different lineages [Erickson 2008] has been established as single locus DNA barcode. Due to limited variation, COI was dismissed as potential barcoding region for land plants [Fazekas 2008]. Although the plant working group of the consortium for the barcode of life [CBOL 2009] proposed the two plastid markers rbcL (RubisCo large subunit) and matK (maturase K) as primary barcodes, their resolution as single barcodes is limited in a wide range of taxa, such that laborious combination of several markers is required (e.g. 4 in Crocus [Seberg 2009]). A second limitation of DNA Barcoding arises from fundamental differences in the concept of “species” between animals and plants: In contrast to the animal kingdom, where species can be defined quite clearly as entities separated by propagation barriers (for review see [Mayr 1942]), a species concept that is merely based on gene flow is problematic in plants, since conventional sexual propagation is complemented by alternative mechanisms of propagation such as apomixis, vegetative propagation, and allopolyploidy leading to more reticulate models of phylogeny that limit the validity of molecular cladistics [Doyle 1995]. Especially in Angiosperms, numerous complexes of closely related, morphologically similar, phenotypically variable and readily interbreeding species that nevertheless may differ in their chemical composition are not an exception, but seem to represent the rule posing major challenges to species-based approaches of authentication.

The introduced methodologies using diagnostic primers (ARMS assisted diagnostics) or differential restriction patterns (PCR-RFLP) focus on single nucleotide differences and rely on universal primers. As shown in the comparison of separability and heterogeneity of common universal markers there are limitations in resolution. Patterns (i.e. a single nucleotide difference) that seem specific at one point in time might get rendered useless as soon as additional taxa are added to the data set. Although there are quite some reasons to follow these approaches, there are also reasons to consider different ones. For example it is true that it is much more laborious to develop new markers especially if the marker region is to be well characterized. The argument of the barcoding approach (millions of species have to be catalogued, loss of biodiversity etc.) cannot be transferred directly to the present question. Although there are many taxonomic groups involved in food and medicine and more will be added in the future, it appears to be possible to develop group specific markers that are characterized by robust delineation of species. New technology and genomic data being produced en mass presents new possibilities for different approaches.

C. Concluding remarks

1. *Dracocephalum* L.

Among the studied members of the genus *Dracocephalum* our initial focus was directed towards *D. moldavica*. As early as the 16th century, *D. moldavica* was brought to Europe and cultivated as ornamental, spice, medicinal and bee plant. In the late 20th century cultivation in Eastern Europe heralded its commercialization ultimately leading to nowadays tea products. In the first phase of our study with limited reference material available we established genetic diagnostics aimed to differentiate between potential adulterant genera *Melissa* and *Nepeta* and *Dracocephalum*. Subsequently after increasing sampling of *Dracocephalum* species we verified the capability of psbA-trnH as marker for species level diagnostics and developed an assay to differentiate between *D. moldavica* and other available species of the genus.

The unclassified *Dracocephalum* accession (*D*. species) provided by the botanical garden Halle, Germany, presents an interesting case. Although overall genetic similarity suggests classification as *D. moldavica* there are at least two regions within psbA-trnH showing clear signs of divergence from other *D. moldavica* accessions. In one of these regions *D*. species shares highly similar sequence motifs with *D. rupestre* while both *D. moldavica* accessions share the motif with *D. multicaule*. One particular observation in morphological development of *D*. species might also be of relevance. Both accessions from Halle were cultivation at the same time, but while the *D.*

moldavica accession developed its typical annual habitus, *D*. species remained in a state more typical for biennial or some perennial plants like *D. rupestre*, *D. multicaule* and *D. nutans*, not engaging in generative activities presumably until the next season. In one other region of psbA-trnH *D*. species shares a motif with the external *D. moldavica* sequence accession which is absent in all other accessions of *Dracocephalum*. Unfortunately we do not have access to the respective plant. Optimized seed data might also support a significant difference between *D. moldavica* and *D*. species and will be prepared and analyzed in the future.

Another interesting aspect of *Dracocephalum* besides *D. moldavica* is the *Ruyschiana* complex. Although many of our accessions did not flower last season and *D. austriacum* did not even germinate, considering the development so far, available molecular data and images available online, a group is revealed where a very close look is necessary to actually see distinctness and internal relations may represent one of the many paraphyletic cases that are typical in flowering plants. Hopefully we can have a closer look in the future revealing information of evolutionary history enabling us to explain the current critical status of populations in Europe.

2. Genetics and Species

For our purpose – specimen identification on the species level based on DNA – we either rely on species criteria that have already proven to apply for the particular case or we look for other criteria that can be used. In the latter case from almost 30 so called theoretical species concepts published only one can be considered appropriate for species diagnostics while others are merely subordinate to that concept describing derived special cases essential for the study of species in practice [Mayden 1997, Haveman 2013].

Since we - as any taxonomist would be - are confronted with an epistemological question our primary interest lies with properties that we can use to identify an individual member of a species. Our species criterion is genetic even if we use other criteria to identify reference material. What we need in either case is sufficient divergent change for a group of organisms to become distinct – the way that we can recognize them. Sexual reproduction as operational property by which species stay distinct and the interruption of which might lead to divergence and eventually speciation is only one factor among others we have to consider in the search of characters that can be used to determine identity. In Land plants increased genomic plasticity reveals itself by less impermeable reproductive barriers and frequent polyploidization (20 – 40 % of flowering plant species [Stebbins 1950, Grant 1971]; 70 % of angiosperms [Masterson 1994]). Considering our

genetic data derived from common markers *Dracocephalum* appears to contain closely aggregated groups of species rather than distinctly defined single species which might indicate increased speciation frequencies within these intra generic groups.

The DNA Barcoding approach which successfully uses genetic distance of cytochrome c oxidase I (COI) in animals, in plants still is limping behind caused by low mitochondrial substitution rates [Mower 2007, Fazekas 2008] and the lack of alternative universal markers with comparable divergence rates. The importance of the question whether to use a genetic distance or character based identification approach seems to be mainly determined by the divergence rate of the used marker. We therefore argue that if interspecific genetic differences are abundant and intraspecific conservation predominates, delineation of plant species using genetic distance will succeed. From the diagnostic point of view however approaches using PCR based techniques are character based thus rely on taxon specific nucleotides. Character based approaches also represent solutions for groups were interspecific divergence is highly heterogeneous rendering distance threshold optimization ineffective. If neither distance nor characters of a particular marker suffice to delineate species of interest, we need to reconsider the evolutionary history of the group. This leads us back to the only appropriate species concept mentioned above.

> *A species is a single lineage of ancestral descendant populations of organisms which maintains its identity from other such lineages and which has its own evolutionary tendencies and historical fate.*
>
> The evolutionary species concept by Simpson 1961 modified by Wiley 1978

The evolutionary species concept which if considered carefully ultimately will lead us to suitable genetic markers. Marker universality is not obligatory for genetic food diagnostics which means even if we are confronted with cryptic taxa (e.g. containing apomicts or hybrids) as long as there are identifiable unique characters there is a way to distinguish species of such groups. Finally, there is one more question that seems most relevant for species diagnostics when considering the term identity in combination with the fact that species have an evolutionary tendency: Does identity change and if so how fast does it change?

Basically the identity of a species only changes when characters change and those only change when the genetic foundation changes. The primary evolutionary tendency of each species is to reproduce and secure its own survival. Changes on the genetic level appear frequently but only changes that are either mostly irrelevant to the primary tendency or increase success of the same

are retained. Considering the many cases where species have been described hundreds of years ago and up till today have retained their key characters, it seems unlikely that on a non-geological time scale considerable changes will mutate those species identities. However, there are factors that can increase the rate of change within a species (e.g. bottleneck effect) and considering each nucleotide as a character, change can be very fast. This pronounces the importance of choosing an adequate genomic region for genetic species diagnostics.

For all other cases, i.e. cryptic species for which no stable species criteria can be found, the question of change in identity does not apply as long as it has not been defined. For example it has been revealed that polyploid plant species are typified by multiple origins [Soltis 1999]. This means that each polyploidization event and subsequent genome reorganization leads to a single lineage with its own genetic identity. With multiple origins we would consider this a cryptic species consisting of different micro species. Approaching diagnostics here will only be possible on a level where identity is genetically coherent. By considering each lineage as a separate micro species and providing that genome structure is sufficiently stable allows us to diagnose even cryptic species.

To summarize: In genetic food diagnostics we only can diagnose taxa – most basically species – that are based on a genetically coherent definition. The evolutionary species concept most adequately and generally describes the key feature of life – its capability to reproduce. By reproduction genetic material is transferred into the next generation maintaining identity. Change is introduced by mutation, horizontal gene transfer or other genomic events that may lead to loss or change of identity through directional selection. Hybridization as part of the history of all higher life forms in plants might be a retained key feature to overcome their limitations in movement. By frequently producing new genetically isolated entities there are more evolutionary tendencies that can be developed.

As long as there is coherence – the maintenance of identity by each entity – we can diagnose them. However, we have to consider that not all lineages are meant to be and that there are groups that are in the process of finding identity. In theory it is possible to diagnose any genetically coherent entity.

3. Outlook

For the methodological part of genetic food diagnostics there are still many open questions. Firstly, will we be able to apply simultaneous authentication of different species contained in complex mixtures? By using a marker of limited length (e.g. psbA-trnH ~ 400 bp in *Dracocephalum*)

and with diagnostic nucleotides concentrated in certain regions there will be restrictions of how many species we can authenticate simultaneously depending on size difference of diagnostic fragments. Additionally, the requirement for multiple primers to have homogeneous melting temperatures, minimal primer-primer interactions and the need for reaction optimizations [for review see Henegariu 1997] might complicate such an endeavor considerably. Thus, it might be necessary to switch to next generation sequencing (NGS) based diagnostics exemplified by environmental barcoding studies with the shared objective to deal with complex samples composed of a mix of many species [e.g. Sogin 2006, for review see Valentini 2009b]. However, most herbal tea mixtures contain less than 10 species and by investing in high resolution agarose while focusing on different small (< 500 bp) diagnostic fragments it should be possible to unambiguously separate reactions containing even more than 10 fragments.

Secondly, will we be able to authenticate species even when DNA degradation becomes a limiting factor? With psbA-trnH, ITS1 and ITS2 (215 – 260 bp) we already have relatively short markers that could be useful amplifying DNA from highly degraded samples. Since it is to be expected that these markers will not be able to differentiate between species in all instances, even shorter markers that contain enough information have to be developed first. Exemplified by another NGS approach applied in diet analysis [Valentini 2009a] using common markers will greatly reduce the ability to distinguish between related species. The use of reconstructing PCR as mentioned earlier (p. 63) to recover fragments of sufficient size might be an option to overcome size limitations. Ultimately, the only feasible way to reliably authenticate highly processed products where only low concentrations of very short DNA fragments can be recovered will be to combine information of several sufficiently short markers. To maintain our current approach using a universal marker as positive control, in this context will present another challenge we have to face in the future.

Lastly, will we be able to improve the situation regarding the availability of sufficient exotic reference plants? Our current sources for respective plants are botanical gardens that participate in seed or plant exchange programs. An evaluation of the offered resources [manuscript in preparation] revealed that the quality with regard to background information on accessions is very heterogeneous. Many garden collections lack essential informations on their accessions and plants have been cultivated for an unknown time. Artificial selection and hybridization are factors that are likely to compromise the representative value of the respective accession [Ye 2006, Aplin 2007]. Botanical gardens that offer wild collected seeds are limited and those that have organized expeditions to exotic regions and returned with documented seeds remain marginal. It seems that

the only way to acquire enough reference material in a relatively short time and with limited expenses is to engage in cooperation with partners from the respective regions.

VII. Appendix

A. *Dracocephalum* Plant Accessions

Table 5: List of plant accessions used in the *Dracocephalum* case study. The table shows the internal ID, provider of seed material (BG/GB = botanical garden, BGU/JBU/HBU = botanical garden associated with a university, BAZGRC = Braunschweig Genetic Resources Centre), an indicator for IPK and IPEN accessions, an indicator of CBD compliance for accessions that were acquired outside IPEN and if available the origin of the plant.

ID	Provider	IPK	IPEN	CBD	Origin
Dracocephalum argunense					
7900	BGU Hohenheim DE		√		
8162	BG RAS Amur RU			√	
8252	BGU Cluj-Napoca RO		√		
8262	HBU Tartu EE			-	
Dracocephalum austriacum					
7648	BG St. Gallen CH		√		Wild: Ardez GR, Burghügel
7686	BGU Bern CH	√			
Dracocephalum diversifolium					
7691	BAZGRC: 29087	√			
Dracocephalum foetidum					
7986	MHG Washington US			-	
Dracocephalum grandiflorum					
7684	BGU Basel CH	√			
8012	BGU Brno (FacSci) CZ			√	
8167	BGU Innsbruck AT		√		
8253	BGU Cluj-Napoca RO		√		
8349	BGU Zagreb HR		√		ex BV Milano
Dracocephalum integrifolium					
7690	BAZGRC: 29088	√			
Dracocephalum moldavica					
5861	Fa. Rühlemanns DE			-	
5862	BGU Zürich CH	√			
5863	Dr. Th. Gladis	√			
7682	BGU Frankfurt DE	√			
7687	BAZGRC: 55841	√			
7689	Zierpflanzens. Erfurt: 8742	√			
7899	BGU Hohenheim DE		√		
8021	GB Alpino Rezia Bormio IT			√	
8138	BG Turku FI			-	
8140	BGU Halle DE		√		
8175	BGU Kaunas LT			-	
8223	BG Nancy FR		√		
8251	BGU Cluj-Napoca RO		√		

ID	Provider	IPK	IPEN	CBD	Origin
8264	HBU Tartu EE			-	
7701	Templiner Kräutergarten DE			-	
Dracocephalum multicaule					
7685	Inst. Arzneipfl. Poznan PL	√			
Dracocephalum nutans					
8265	HBU Tartu EE			-	
Dracocephalum parviflorum					
7874	RBG-KEW EN			√	Wild: central Alberta, Canada
Dracocephalum renati					
7088	BG St. Gallen CH		√		
8013	BGU Brno (FacSci) CZ			√	
8263	HBU Tartu EE			-	
Dracocephalum rupestre					
7089	BG St. Gallen CH		√		
8254	BGU Cluj-Napoca RO		√		
Dracocephalum ruyschiana					
5156	BGU Karlsruhe DE		√		
7683	BGU Frankfurt DE	√			
7688	BAZGRC: 29000	√			
7921	BG Muenchen DE		√		Ex-situ – Italy, South Tyrol
8011	BGU Brno (FacSci) CZ			√	
8020	GB Alpino Rezia Bormio IT			√	
8137	BG Turku FI			-	
8141	BGU Halle DE		√		
8166	BGU Innsbruck AT		√		
8174	BGU Kaunas LT			-	
8194	BG St. Gallen CH		√		
8209	JBU Grenoble FR			-	Wild: Lautaret, 2100 m
8230	BG Wroclaw PL		?	?	
8261	BGU Petrozavodsk RU			√	Wild: Petrozavodsk
8348	BGU Zagreb HR		√		
Dracocephalum scrobiculatum					
8139	BGU Halle DE		√		
Dracocephalum spec.					
8142	BGU Halle DE		√		

B. Commercial Products

Table 6: Declared composition of commercial samples used in the 'Lemon Myrtle' study and rbcL fragments obtained after digest with Sac2.

Sample	Declared composition	rbcL-Sac2 fragments [bp]
D_3	'Lemon Grass', 'Lemon Myrtle', 'Moldavian Dragonhead', 'Elderberry Flowers'	~600, ~450, ~150
D_4	'Ginger', 'Lemon Grass', 'Lemon Peel', 'Lemon Myrtle', 'Liquorhize'	~600
D_5	'Rooibos', 'Lemon Myrtle', 'Lemon Grass', 'Lemon Verbena', 'Cranberry', 'Orange Peel', 'Peppermint', 'Lemon Peel'	~600
D_5	'Orange Leaves', 'Ginger', 'Sweet Bramble Leaves', 'Lemon Myrtle', 'Natural Orange Oil'	~600, ~450, ~150

Table 7: Commercial 'Dragonhead' tea products used for species level diagnostics

Name	Vendor	Type	Content
Drachenkopf	Berglandkräuter aus Hessen Monika Grebe-Schuchhardt & Ute Kern GbR	Loose	'Dragonhead'
Bio Kräutertee Drachenkopf	Hollerbuschhof, Hartmut Herrmann	Loose	'Dragonhead'

C. External Sequence Data

Table 8: List of external rbcLa and matK sequence data used in this study. For rbcLa and matK the BOLD process ID is indicated. All sequences can be retrieved via BOLD. If no year is mentioned the sequence is only available at BOLD. All other sequences have been published at GenBank in the indicated year and if used in a publication the respective journal and publication year is mentioned in the note column.

Taxon	BOLD	M	Year	Note
D.grandiflorum	GBVG2945-11	rbcL	1994	Biosci. Rep. 49, 635-645 (1994)
D.parviflorum	BBYUK1219-12	rbcL	-	
D.parviflorum	BBYUK1220-12	rbcL	-	
D.rupestre	GBVS3770-13	rbcL	2010	Unpublished
D.ruyschiana	GBVG2948-11	rbcL	1994	Biosci. Rep. 49, 635-645 (1994)
D.ruyschiana	RINGV009-12	rbcL	-	
Hyssopus officinalis	GBVG2989-11	rbcL	1994	Biosci. Rep. 49, 635-645 (1994)
Melissa officinalis	GBVG3073-11	matK	2010	Ecol. Lett. 14 (4), 389-396 (2011)
Melissa officinalis	GBVG3075-11	rbcL	1994	Biosci. Rep. 49, 635-645 (1994)
Melissa officinalis	GBVX216-13	rbcL	2013	Unpublished
Nepeta cataria	KSR161-07	matK	2010	Unpublished
Nepeta cataria	POWNA1443-12	matK	2011	Unpublished
Nepeta cataria	GBVG3182-11	rbcL	1994	Biosci. Rep. 49, 635-645 (1994)

Taxon	BOLD	M	Year	Note
Nepeta italica	GBVS3888-13	matK	2011	Mol Ecol Resour 12 (4), 620-633 (2012)
Nepeta italica	GBVS3954-13	rbcL	2011	Mol Ecol Resour 12 (4), 620-633 (2012)

Table 9: List of external ITS and psbA-trnH sequence data used in this study. Since these markers are not integrated into the BOLD (plant) system, the GenBank locus ID is indicated. Sequences have been published at GenBank in the indicated year and if used in a publication the respective journal and publication year is mentioned in the note column.

Taxon	GenBank	marker	Year	Note
Agastache barberi	AJ421001	ITS	2001	Thesis (2001)
Agastache rugosa	EU590857	psbA-trnH	2008	Unpublished
Bupleurum falcatum	AY551290	ITS	2004	Direct Submission
Cedronella canariensis	JQ669079	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)
Clinopodium chinense	FJ513089	psbA-trnH	2008	Unpublished
Dorystaechas hastata	DQ667252	ITS	2006	Ann. Bot. 100 (2), 375-391 (2007)
D. argunense	GQ456140	ITS	2009	Unpublished
D. bullatum	JQ669096	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)
D. grandiflorum	AJ420999	ITS	2001	Thesis (2001)
D. kotschyi	AJ420998	ITS	2001	Thesis (2001)
D. moldavica	AY506659	ITS	2003	Syst. Bot. 29 (3), 702-715 (2004)
D. moldavica	FJ513107	psbA-trnH	2008	Unpublished
D. parviflorum	JQ669097	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)
D. rupestre	FJ513110	psbA-trnH	2008	Unpublished
Glechoma hederacea	DQ006014	ITS	2005	Proc. Natl. Acad. Sci. U.S.A. 102 (23), 8369-8374 (2005)
Glechoma hederacea	FJ395524	psbA-trnH	2008	Unpublished
Hyssopus officinalis	JQ669106	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)
Hyssopus seravschanicus	AY506657	ITS	2003	Syst. Bot. 29 (3), 702-715 (2004)
Isodon shikokianus var. occidentalis	AB523532	psbA-trnH	2009	Unpublished
Lallemantia canescens	JQ669108	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)
Lallemantia peltata	AJ420997	ITS	2001	Thesis (2001)
Lycopus uniflorus	DQ667302	ITS	2006	Ann. Bot. 100 (2), 375-391 (2007)
Meehania cf. henryi	JQ669103	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)

Taxon	GenBank	marker	Year	Note
Melissa axillaris	JQ669114	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)
Melissa officinalis	DQ189090	ITS	2005	Pharmazie 61 (11), 912-915 (2006)
Melissa officinalis	HF565305	psbA-trnH	2012	Int. J. Legal Med. 127 (6), 1109-1123 (2013)
Mentha canadensis	JN406990	psbA-trnH	2011	Food Control 25 (2), 758-766 (2012)
Nepeta cataria	AJ515313	ITS	2002	Taxon 52 (1), 21-32 (2003)
Nepeta cataria	DQ667301	ITS	2006	Ann. Bot. 100 (2), 375-391 (2007)
Nepeta cataria	JQ669126	ITS	2012	Am. J. Bot. 99 (5), 933-953 (2012)
Nepeta italica	HQ902833	psbA-trnH	2011	Mol Ecol Resour 12 (4), 620-633 (2012)
Nepeta laxiflora	AJ420995	ITS	2001	Thesis (2001)
Origanum vulgare	HQ596773	psbA-trnH	2010	Unpublished
Perovskia scrophulariifolia	DQ667330	ITS	2006	Ann. Bot. 100 (2), 375-391 (2007)
Rosmarinus officinalis	FJ513141	psbA-trnH	2008	Unpublished
Salvia deserta	DQ132865	ITS	2005	Unpublished
Salvia sclarea	KC473222	psbA-trnH	2013	Gene 528 (2), 206-215 (2013)
Schizonepeta tenuifolia	JN802670	ITS	2011	Unpublished
Schizonepeta tenuifolia	EU590861	psbA-trnH	2008	Unpublished
Thymbra capitata	HE819478	psbA-trnH	2012	Unpublished
Thymus praecox subsp. polytrichus	HE819474	psbA-trnH	2012	Unpublished

D. Primers and PCR programs

Table 10: List of primers used in this study.

Primer		5‘-Sequence-3‘	Marker	Ann. (° C)
TBPF1	fw	GARGCYGARAAYTGYGAYTG	TBP	55
TBPR1	rv	RTCHGGRTAYTCYTCHCKRAT	TBP	55
rbcLa_f	fw	ATGTCACCACAAACAGAGACTAAAGC	rbcL	58
rbcLa_rev	rv	GTAAAATCAAGTCCACCRCG	rbcL	58
3F_KIM f	rv	CGTACAGTACTTTTGTGTTTACGAG	matK	52
1R_KIM r	fw	ACCCAGTCCATCTGGAAATCTTGGTTC	matK	52
ITS4	rv	TCCTCCGCTTATTGATATGC	ITS	50
ITS5	fw	GGAAGTAAAAGTCGTAACAAGG	ITS	50
trnH	rv	CGCGCATGGTGGATTCACAATCC	psbA-trnH	55
psbA	fw	GTTATGCATGAACGTAATGCTC	psbA-trnH	55
DC4	fw	TTTCCAAGGCCCACCTCATAGT	ARMS	58

Primer		5‘-Sequence-3‘	Marker	Ann. (° C)
pbsA72G	fw	TAAGACTTGTTCTTAGCTTGTATGG	ARMS	55
10RP-50a		TGGTCACTGA	RAPD	35
10RP-50b		AGGTCACTGA	RAPD	35
10RP-50c		TCGTCACTGA	RAPD	35
10RP-50d		TGCTCACTGA	RAPD	35
10RP-50e		TGGACACTGA	RAPD	35
10RP-50f		TGGTGACTGA	RAPD	35
10RP-50g		TGGTCTCTGA	RAPD	35
10RP-50h		TGGTCAGTGA	RAPD	35
10RP-50i		TGGTCACAGA	RAPD	35
10RP-50j		TGGTCACTCA	RAPD	35
10RP-50k		TGGTCACTGT	RAPD	35
S50e-9a1-fw2	fw	ACCACCTACAACGAATTGGTAGCCA	SCAR	58
S50e-9a1-rv2	rv	CACCGCTGACCAGAGTTGGAGT	SCAR	58

E. Classification

1. Morphology

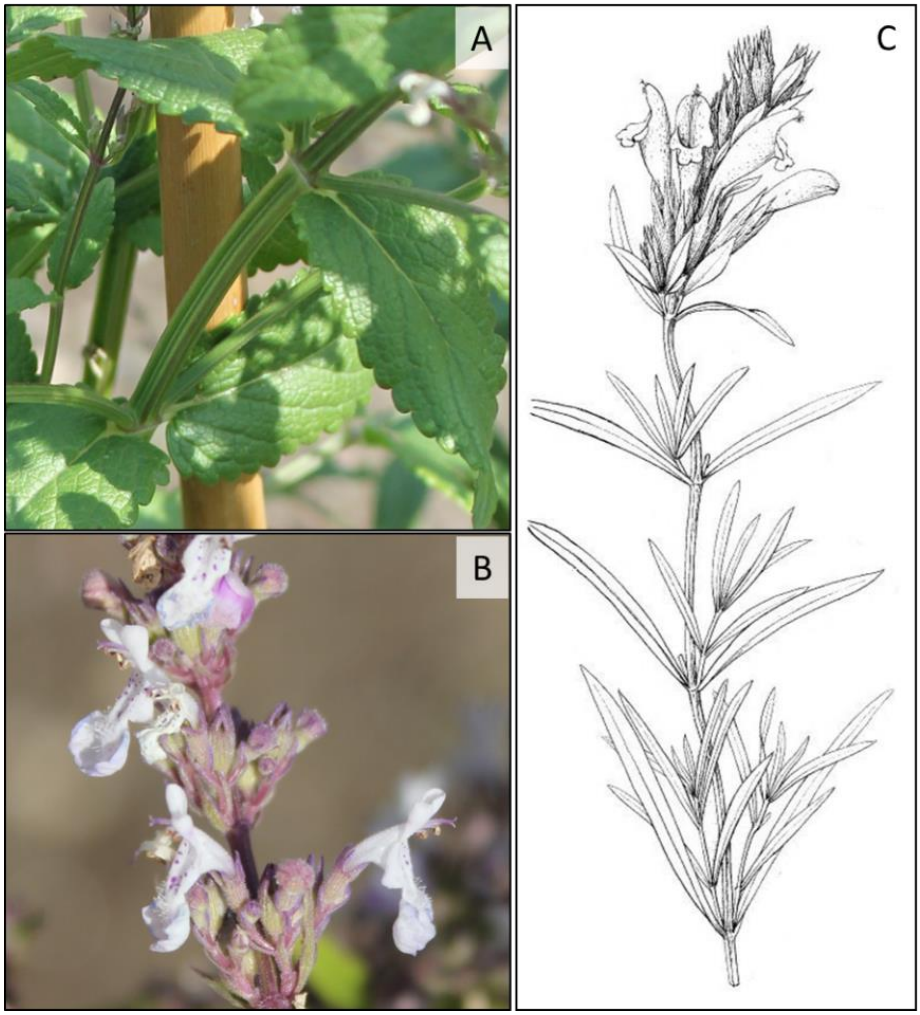


Figure 22: Key features of accession 7900 (*Nepeta* species ex *D. argunense*). A: Common characteristics in Lamiaceae are alternating opposite leaves and a squared stem. In this case we see ovate crenate leaves while the true *D. argunense* has entire lanceolate-linear leaves typical for the subgenus/section Ruyschiana. B: The corolla displayed by accession 7900 is purplish-

white and less than 10 mm long. The tube is abruptly dilated into a short throat. The upper lip continues almost straight and is more or less deeply divided into two ovate lobes. In contrast the corolla of *D. argunense* is usually 30 – 40 mm long, azure-blue and has a large throat. (Based on the Flora of the USSR (*D.argunense*) and the Flora of China (*Nepeta nuda*)). C: Drawing of *D. argunense* (Flora of China)

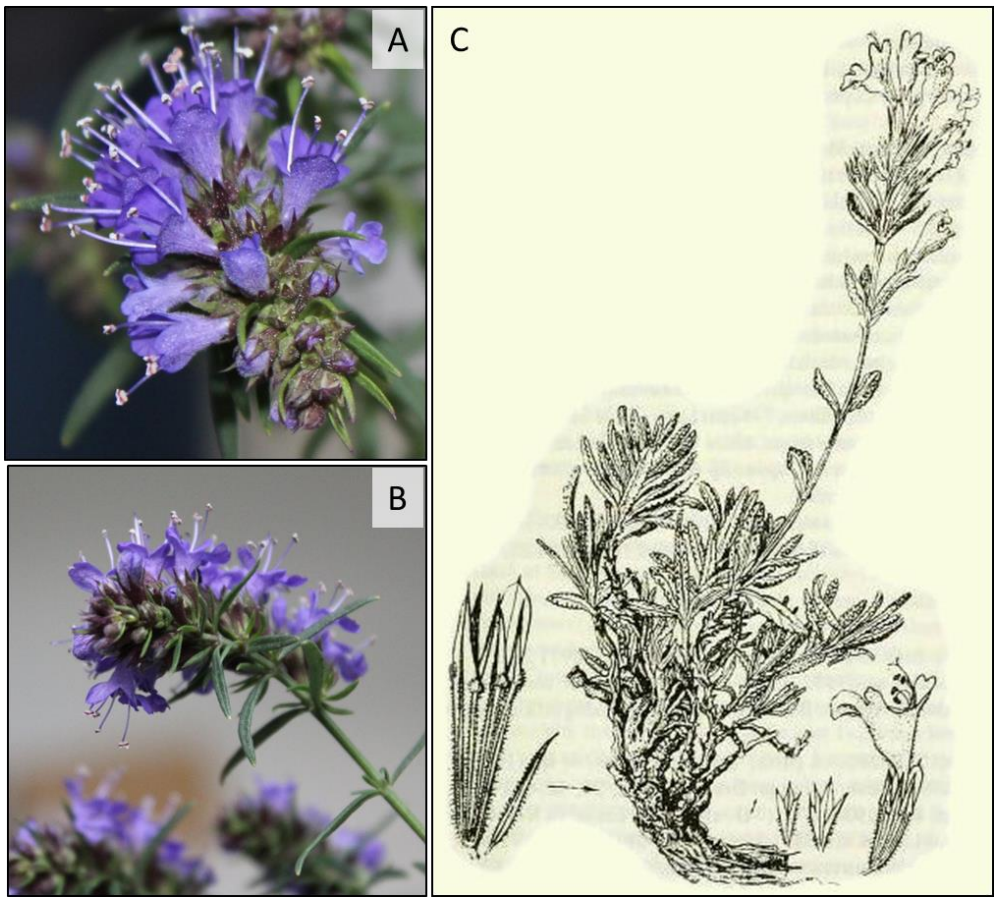


Figure 23: Key features of accession 8139 (*Hyssopus* species ex *D. scrobiculatum*). A: Inflorescence with emerging stamens of which the longer pair is almost twice as long as the throat of the corolla. Dragonhead species rarely display exserted stamens and as can be derived from C it is not a particular characteristic of *D. scrobiculatum*. B: Inflorescence captured from a lower angle displaying the second verticillaster a feature of *H. officinalis* and not mentioned within the *Dracocephalum* literature. C: Drawing of *D. scrobiculatum* (Flora of the USSR)

Flora of the U.S.S.R., Springfield, Va. :Israel Program for Scientific Translations, 1968-
http://www.biodiversitylibrary.org/bibliography/43751
v. 20:
http://www.biodiversitylibrary.org/item/95895
Page(s): Page 314

Dracocephalum diversifolium Rupr. in Mem. Acad. Sc. Petersb. VII ser., XIV, 4 (1869) 66; Lipsk. in Tr. Bot. sada, XXVI, 586. - Exs.: HFAM, No. 206.

Perennial; rhizome long, to 1 .5 cm thick, ascending, strongly branching above; stems few, 20-40 cm long, covered with very short hairs, branching, with opposite or alternate branches, rarely simple; leaves ovate, crenate-dentate or ovate-lanceolate, entire, obtuse or acute above, tapering to base, abruptly short-petioled, 2-4 cm long, 1-1.5 (2) cm wide, the abbreviated shoots in their axils with smaller, usually entire leaves ca. 1 cm long, 0.4 cm wide; flowers in axils of upper elUptic or lanceolate leaves in remote or approximate many-flowered verticillasters; calyx 15 mm long, covered with very short hairs, 2-lipped, with aristate teeth; upper lip with 3 ovate teeth 1/3-1/2 as long as the lanceolate teeth of lower lip; bracts lanceolate or ovate, with 4 lanceolate teeth at either side, the teeth with slender awn 4-5 mm long, the terminal tooth broadly ovate, with shorter awn; corolla (dry) straw-colored, sometimes bluish, densely covered with soft white hairs, ca. 30 mm long; nutlets pale brown, trigonous, 4 mm long, 2.2 mm wide. July-August.

(Plate XXVII, Figure 1.)
Stony and gravelly slopes, taluses, sometimes in mountain scrub at altitudes of 1500-4000 m. - Centr. Asia; T.Sh., Pam.-Al. Endemic. Described. from Tien Shan. Type in Leningrad.

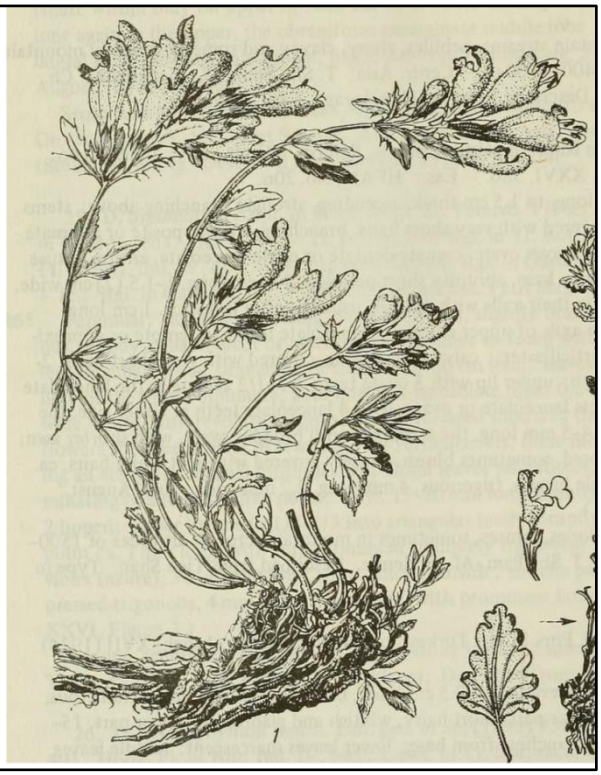


Figure 24: Drawing of *D. diversifolium* and description: Flora of the U.S.S.R.

Figure 25: Images of local *D. diversifolium* accession 7691.

Figure 26: Images of *D. diversifolium* retrieved from the internet.

2. Genetic Distance

a) rbcLa - Neighbour Joining

The optimal Neighbor Joining tree constructed using rbcLa sequence data of Dracocephalum accessions has the sum of branch length 0.0537. The tree is drawn to scale, with branch lengths in the same units as those of the evolutionary distances used to infer the phylogenetic tree. The analysis involved 35 nucleotide sequences. There were a total of 553 positions in the final dataset. All black-filled forms indicate new sequence data retrieved by the author. The respective white-filled forms indicate external sequence data. All *Dracocephalum* share a major cluster (*Dracocephalum-Lallemantia*-Cluster = DLC) with *Lallemantia* and one external *Hyssopus* sequence (GBVG2989-11). In contrast our accession 8139 which originally was classified as *D. scrobiculatum* and later determined to be a Hyssopus species forms a single cluster outside the DLC. Within the DLC *D. ruyschian and D. argunense*, *D. parviflorum*, *D. renati*, *D. diversifolium* and *D. rupestre* (including the external Hyssopus sequence) are separated into clusters. Due to identical rbcLa sequence data *D. moldavica*, *D. multicaule*, *D. integrifolium*, *D. nutans* and the unclassified accession (8142) are clustering together and cannot be separated by this marker. The sequence of accession 8252 (*D. ruyschiana* ex *D. argunense*) is identical with *D. ruyschiana* accessions and

different from *D. argunense* accessions. All accessions received as *D. grandiflorum* (7684, 8012, 8349 and 8167) share identical sequence data with the morphologically determined *D. rupestre* accession (7089). Additionally, the external *D. grandiflorum* sequence (GBVG2945-11) is placed at a different position within the DLC. Our *Melissa* and *Nepeta* form clusters together with respective external accessions well separated from the DLC, from each other and accession 8139 (*Hyssopus* spec).

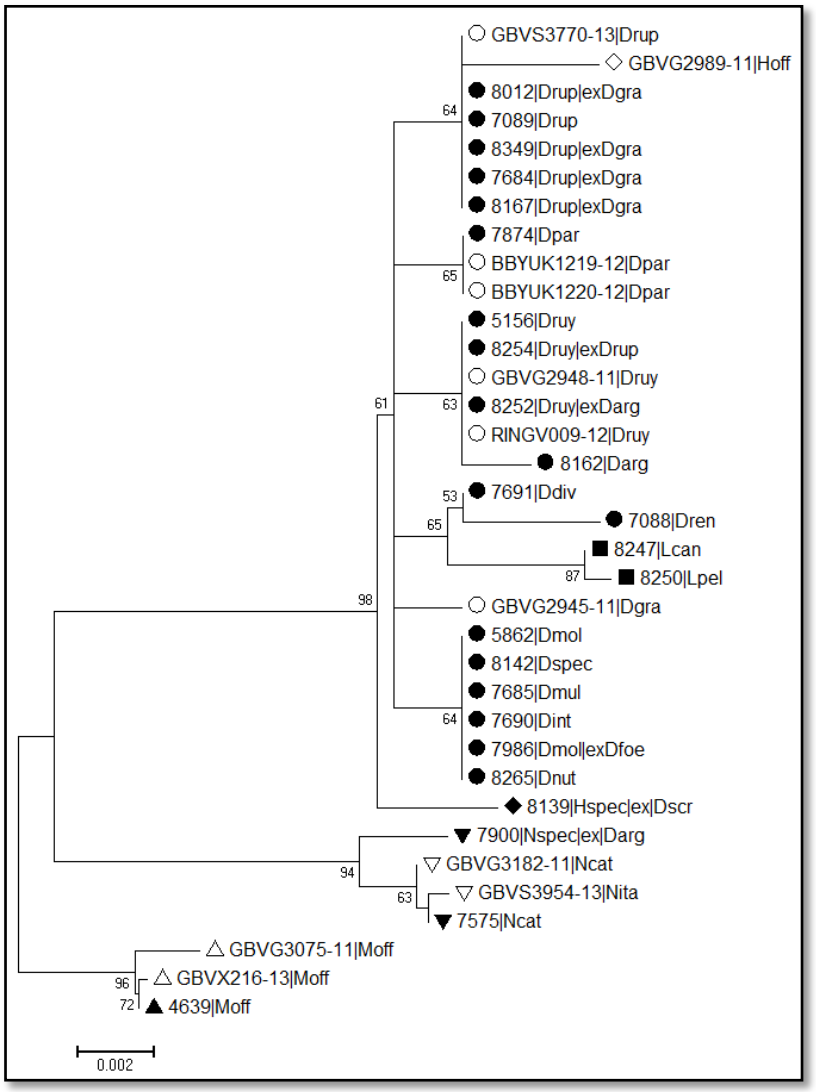


Figure 27: Neighbor-Joining Tree of selected *Dracocephalum* accessions based on rbcLa sequence data. The root is placed on the Melissa branch. For taxon abbreviations please refer to Table 3

b) matK-KIM – Neighbour Joining

Neighbor-Joining Tree of selected *Dracocephalum* accessions based on matK-KIM sequence data. The optimal tree with the sum of branch length = 0.1805 is shown while the root is placed on the *Melissa* branch. The tree is drawn to scale, with branch lengths in the same units as those of the

evolutionary distances used to infer the phylogenetic tree. The analysis involved 24 nucleotide sequences. There were a total of 732 positions in the final dataset. All black-filled forms indicate sequence data retrieved by the author. The respective white-filled forms indicate sequence data retrieved from external sources. The matK-KIM sequence data shows *D.parviflorum*, *D. ruyschiana* and *D. argunense* together in one cluster while two accessions of *D. ruyschiana* (7683 and 8261) form a distinct nested cluster. *D. renati* and *D. diversifolium* form separated clusters. The unclassified accession (8142) forms a cluster together with the *D. moldavica* accessions. Accessions received as *D. grandiflorum* (7684, 8012 and 8167) share identical sequence data with the morphologically determined *D. rupestre* accession (7089). Additionally, a single read sequence of *D. multicaule* is located within the same cluster. Accession 7900 which we received as *D. argunense* forms a cluster together with *Nepeta* sequences from external sources. Like in the rbcLa tree our accession 8139 which was received as *D. scrobiculatum* is located at the base of the *Dracocephalum-Lallemantia* cluster.

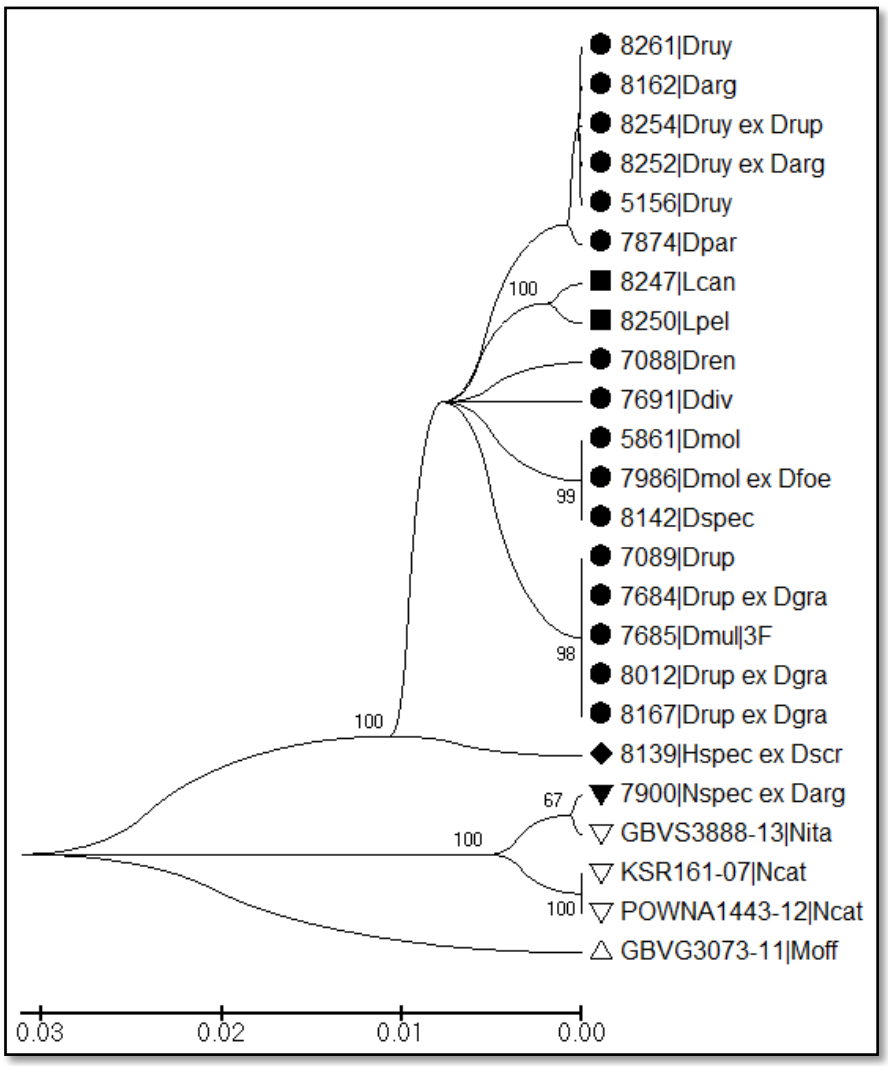


Figure 28: Neighbor-Joining Tree of selected *Dracocephalum* accessions based on matK-KIM sequence data. The root is placed on the Melissa (Moff) branch. For taxon abbreviations please refer to Table 3.

F. 'Lemon Myrtle' Diagnostics

Table 11: Characteristic features observed in fully developed leaves of Backhousia citriodora F.Muell, and Leptospermum citratum Challinor, Cheel & A.R.Penfold (Abbreviations: E-ad = adaxial epidermis, E-ab = abaxial epidermis)

	Backhousia citriodora	*Leptospermum citratum*
Leaf type	bifacial	equifacial
E-ad	puzzle-shaped pavement cells, polygonoid cells above vascular bundles	polygonal pavement cells, distinct cuticular folds
E-ab	like in E-ad	like E-ad, with cuticular folds
guard cells E-ad	none	tetracytic, anomocytic
guard cells E-ab	anomocytic	tetracytic, anomocytic
trichomes E-ad	unicellular, unbranched	unicellular, unbranched, weakly

		coiled trichomes that are aligned
trichomes E-ab	unicellular, unbranched, like in E-ad, in addition, more thick-walled and longer unicellular, unbranched, trichomes	less than in E-ad
mesophyll	numerous schizogenic oil cavities idioblasts, calcium oxalate druses cell files filled with single calcium oxalate crystals lining vascular bundles	numerous schizogenic oil cavities idioblasts, calcium oxalate druses cell files filled with single calcium oxalate crystals lining vascular bundles
palisade cells / pavement cell	mostly 6-8	mostly 2-5

G. *Dracocephalum* Diagnostics

Table 12: Microscopic features of leaves of *Dracocephalum moldavica* versus *Melissa officinalis*

	Dracocephalum moldavica	*Melissa officinalis*
Type of leaf	Bifacial	Bifacial
E-ad	Puzzle-shaped, weakly indented cell walls	Puzzle-shaped weakly indented cell walls
E-ab	Puzzle-shaped strongly indented cell walls	Puzzle-shaped strongly indented cell walls
Stomata E-ad	Diacytic	Absent
Stomata E-ab	Diacytic	Diacytic
Trichomes		
E-ad	Short, unbranched trichomes (1–3 cells, mostly 1–2 cells)	Short, unbranched trichomes (1 cell), and a few long unbranched trichomes (2–8 cells)
Trichomes E-ab	Short, unbranched trichomes (1–3 cells, mostly 2–3 cells)	Like E-ad
G-hairs E-ad	Short hairs with 2 G-cells, and long hairs with 1 G-cell	Short hairs with 2 G-cells, and few long hairs with 1 G-cell
G-hairs E-ab	Like E-ad	Like E-ad
G-scales E-ad	Absent	Absent
G-scales E-ab	Scales with 14–24 G-cells	Scales with 8 G-cells
Palisades/E-cell	Mostly 5–7	Mostly 10–20

Abbreviations Epidermal cells adaxial E-ad, Epidermal cells abaxial E-ab, Glandular G

VIII. Abbreviations

AFLP	Amplified Fragment Length Polymorphism
ARMS	Amplified Refractory Mutation System
AS	Seed Area
bCM	Algorithm: best Close Match
BIN	Barcode Index Number
BOLD	Barcoding Of Life Datasystems
COI	Gene for cytochrome c oxidase I
CS	Seed Circularity
CTAB	Cetyl Trimethyl Ammonium Bromide
DAF	DNA Amplification Fingerprinting
dNTP	Deoxynucleotide
GCW	Global Compendium of Weeds
IPEN	International Plant Exchange Network
IPK	Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung
ITS	Internal Transcribed Spacer
LWR	Seed Length-Width-Ratio
matK	Gene for maturase K
MgCl	Magnesium Chloride
NaCl	Sodium Chloride
NEB	New England Biolabs
NFR	Novel Food Regulation
NJ	Neighbor-Joining
nN	Algorithm: near Neighbour
PBD	Pattern based Diagnostics
PCR	Polymerase Chain Reaction
psbA	Gene for D1 protein of Photosystem II
RAPD	Random Amplified Polymorphic DNA
rbcL	Gene for ribulose-bisphosphate carboxylase, large subunit
RFLP	Restriction Fragment Length Polymorphism
SBD	Sequence based Diagnostics
SCAR	Sequence Characterized Amplified Region
TBP	Tubulin based Polymorphismus
TCM	Traditional Chinese Medicine
tID	Algorithm: threshID
TPL	The Plant List
trnH	Gene for tRNAHis (GUG)

IX. References

H. H. Abd El-Baky & G. S. El-Baroty (2008). 'Chemical and biological evaluation of the essential oil of Egyptian Moldavian balm'. *Advances in food sciences* **30**(3):170–175.

L. Adam & B. Hoppe (2012). *Arznei- und Gewürzpflanzen A - K*. Eigenverl. d. Vereins für Arznei- und Gewürzpflanzenbau SALUPLANTA e.V., Bernburg.

L. Alan Prather, et al. (2002). 'Monophyly and Phylogeny of Monarda (Lamiaceae): Evidence from the Internal Transcribed Spacer (ITS) Region of Nuclear Ribosomal DNA'. *Systematic Botany* **27**(1):127–137.

R. G. Albaladejo & A. Aparicio (2007). 'Population Genetic Structure and Hybridization Patterns in the Mediterranean Endemics *Phlomis lychnitis* and *P. crinita* (*Lamiaceae*)'. *Annals of Botany* **100**(4):735–746.

Alonso & Negueruela (1984). 'Essential oil analysis of thymus villosus subsp. lusitanicus'. *Phytochemistry* **23**(3):581–582.

S. F. Altschul, et al. (1990). 'Basic local alignment search tool.'. *Journal of molecular biology* **215**(3):403–410.

D. Aplin, et al. (2007). 'Indices seminum: are they really worth the effort? '. In *Building a sustainable future: the role of botanic gardens. Proceedings of the 3rd Global Botanic Gardens Congress, Wuhan, China, 16-20 April, 2007.*, pp. 1–16. Botanic Gardens Conservation International.

M. Bardini, et al. (2004). 'Tubulin-based polymorphism (TBP): a new tool, based on functionally relevant sequences, to assess genetic diversity in plant species'. *Genome* **47**(2):281–291. PMID: 15060580.

D. A. Benson, et al. (2008). 'GenBank'. *Nucleic Acids Research* **36**(suppl 1):D25–D30.

J. Bergsten, et al. (2012). 'The Effect of Geographical Scale of Sampling on DNA Barcoding'. *Systematic Biology* **61**(5):851–869.

P. Bertolazzi, et al. (2009). 'Learning to classify species with barcodes'. *BMC Bioinformatics* **10**(Suppl 14):S7+.

K. K. Bhutani (2008). *Herbal Wealth of North-East India: Database and Appraisal*. Department of Natural Products, National Institute of Pharmaceutical Education and Research.

C. R. Bijoy (2007). 'Access and Benefit Sharing from the Indigenous Peoples' Perspective: The Tbgri-Kani Model'. *Law Env't & Dev. J.* **3**:1.

D. Botstein, et al. (1980). 'Construction of a genetic linkage map in man using restriction fragment length polymorphisms.'. *American journal of human genetics* **32**(3):314–331.

S. D. J. Brown, et al. (2012). 'Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding'. *Molecular Ecology Resources* **12**(3):562–565.

A. L. Budantsev (1987). 'System of the genus *Dracocephalum* (*Lamiaceae*)'. *Botanicheskii zhurnal* .

J. Bunsawat, et al. (2004). 'Phylogenetics of *Mentha* (*Lamiaceae*): Evidence from Chloroplast DNA Sequences'. *Systematic Botany* **29**(4):959–964.

G. Caetano-Anolles, et al. (1991). 'DNA Amplification Fingerprinting Using Very Short Arbitrary Oligonucleotide Primers'. *Nature Biotechnology* **9**(6):553–557.

J. Campanella, et al. (2003). 'MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences'. *BMC Bioinformatics* **4**(1):29+.

D. R. Campbell (2004). 'Natural selection in Ipomopsis hybrid zones: implications for ecological speciation'. *New Phytologist* **161**(1):83–90.

A. P. Carnat, et al. (1998). 'The aromatic and polyphenolic composition of lemon balm (*Melissa officinalis* L. *subsp. officinalis*) tea'. *Pharmaceutica Acta Helvetiae* **72**(5):301–305.

S. Chen, et al. (2010). 'Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species'. *PLoS ONE* **5**(1).

S.-J. Chiou, et al. (2007a). 'Authentication of medicinal herbs using PCR-amplified ITS2 with specific primers'. *Planta medica* **73**(13):1421–1426.

Y. Cho, et al. (2004). 'Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plants'. *Proceedings of the National Academy of Sciences of the United States of America* **101**(51):17741–17746.

E. C. M. Coxworth (1965). 'Oil and protein content, and oil composition of the seeds of some plants of the Canadian prairies'. *Journal of the American Oil Chemists' Society* **42**(10):891–894.

A. Cronquist (1978). 'Once again, what is a species? .'. In *Symposium on Biosystematics in Agriculture. Beltsville, Md.(USA). 8-11 May 1977.*

A. Cronquist (1988). 'The evolution and classification of flowering plants'. *New York Botanical Garden, Bronx* .

M. Dadandi, et al. (2009). 'Seed macro and micro morphology of the selected *Nigella* (Ranunculaceae) taxa from Turkey and their systematic significance'. *Biologia* **64**(2):261–270.

B. DasGupta, et al. (2005). 'DNA-BAR: distinguisher selection for DNA barcoding'. *Bioinformatics* **21**(16):3424–3426.

K. Dastmalchi & Others (2008). '*Dracocephalum moldavica* L. and *Melissa officinalis* L.: Chemistry and bioactivities relevant in Alzheimer's disease therapy' .

K. Dastmalchi, et al. (2007). 'Chemical composition and antioxidative activity of Moldavian balm (*Dracocephalum moldavica* L.) extracts'. *LWT - Food Science and Technology* **40**(9):1655–1663.

F. De Mattia, et al. (2011). 'A comparative study of different DNA barcoding markers for the identification of some members of *Lamiacaea*'. *Food Research International* **44**(3):693–702.

R. DeSalle, et al. (2005). 'The unholy trinity: taxonomy, species delimitation and DNA barcoding'. *Philosophical Transactions of the Royal Society B: Biological Sciences* **360**(1462):1905–1916.

K. M. Devaiah & P. Venkatasubramanian (2008). 'Development of SCAR marker for authentication of *Pueraria tuberosa* (Roxb. ex. Willd.) DC'. *Current science* **94**(10):1306–1309.

W. Dnyaneshwar, et al. (2006). 'Development and Application of RAPD-SCAR Marker for Identification of *Phyllanthus emblica* L.'. *Biological and Pharmaceutical Bulletin* **29**(11):2313–2316.

T. Dobzhansky (1935). 'A critique of the species concept in biology'. *Philosophy of Science* pp. 344–355.

T. Dostálek, et al. (2010). 'Genetic diversity and its effect on fitness in an endangered plant species, *Dracocephalum austriacum* L.'. *Conservation Genetics* **11**(3):773–783.

J. J. Doyle (1987). 'A rapid DNA isolation procedure for small quantities of fresh leaf tissue'. *Phytochem bull* **19**:11–15.

J. J. Doyle (1995). 'The Irrelevance of Allele Tree Topologies for Species Delimitation, and a Non-Topological Alternative'. *Systematic Botany* **20**(4):574+.

R. C. Edgar (2004). 'MUSCLE: multiple sequence alignment with high accuracy and high throughput'. *Nucleic Acids Research* **32**(5):1792–1797.

D. L. Erickson, et al. (2008). 'DNA Barcoding in land plants: Developing standards to quantify and maximize success.'. *Taxon* **57**(4):1304–1316.

D. S. Fabricant & N. R. Farnsworth (2001). 'The value of plants used in traditional medicine for drug discovery.'. *Environmental health perspectives* **109 Suppl 1**:69–75.

J. Fagúndez & J. Izco (2004). 'Taxonomic value of seed characters in the *Erica tetralix* L. group (*Ericaceae*)'. *Plant Biosystems - An International Journal Dealing with all Aspects of Plant Biology* **138**(3):207–213.

N. Faham, et al. (2008). 'Calycopterin, an immunoinhibitory compound from the extract of *Dracocephalum kotschyi*'. *Phytother. Res.* **22**(9):1154–1158.

References

A. J. Fazekas, et al. (2008). 'Multiple Multilocus DNA Barcodes from the Plastid Genome Discriminate Plant Species Equally Well'. *PLoS ONE* **3**(7).

A. J. Fazekas, et al. (2009). 'Are plant species inherently harder to discriminate than animal species using DNA barcoding markers? '. *Molecular Ecology Resources* **9**(s1):130–139.

S. Federici, et al. (2013). 'DNA barcoding to analyse taxonomically complex groups in plants: the case of *Thymus* (*Lamiaceae*)'. *Bot J Linn Soc* **171**(4):687–699.

J. Felsenstein (1985). 'Confidence Limits on Phylogenies: An Approach Using the Bootstrap'. *Evolution* **39**(4):783–791.

F. Fritsche & O. Kaltz (2000). 'Is the *Prunella* (*Lamiaceae*) hybrid zone structured by an environmental gradient? Evidence from a reciprocaltransplant experiment'. *American Journal of Botany* **87**(7):995–1003.

T. Gao, et al. (2010). 'Identification of medicinal plants in the family *Fabaceae* using a potential DNA barcode ITS2'. *Journal of Ethnopharmacology* **130**(1):116–121.

F. Gavazzi, et al. (2012). 'Technical improvement of the TBP (tubulin-based polymorphism) method for plant species detection, based on capillary electrophoresis'. *ELECTROPHORESIS* **33**(18):2840–2851.

V. Gobert, et al. (2002). 'Hybridization in the section *Mentha* (*Lamiaceae*) inferred from AFLP markers'. *American Journal of Botany* **89**(12):2017–2023.

E. M. Golenberg, et al. (1996). 'Effect of Highly Fragmented DNA on PCR'. *Nucleic Acids Research* **24**(24):5026–5033.

V. Grant (1971). *Plant speciation*, vol. 183. Columbia University Press New York.

O. Grillo, et al. (2010). 'Statistical seed classifiers of 10 plant families representative of the Mediterranean vascular flora'. *Seed Science and Technology* pp. 455–476.

C. P. W. Group & D. H. Janzen (2009). 'A DNA Barcode for Land Plants'. *Proceedings of the National Academy of Sciences of the United States of America* **106**(31).

C. P. B. Group, et al. (2011). 'Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants'. *Proceedings of the National Academy of Sciences* **108**:19641–19646.

X. Guo, et al. (2011). 'DNA Barcodes for Discriminating the Medicinal Plant *Scutellaria baicalensis* (*Lamiaceae*) and Its Adulterants'. *Biological and Pharmaceutical Bulletin* **34**(8):1198–1203.

X. Han, et al. (2008). 'Protective effects of naringenin-7-O-glucoside on doxorubicin-induced apoptosis in H9C2 cells'. *European Journal of Pharmacology* **581**(1-2):47–53.

X. Han, et al. (2012). 'Protective effect of naringenin-7-O-glucoside against oxidative stress induced by doxorubicin in H9c2 cardiomyocytes.'. *Bioscience trends* **6**(1).

P. Hanelt & R. Büttner (2001). *Mansfeld's encyclopedia of agricultural and horticultural crops:(except ornamentals)*, vol. 1. Springer.

R. Haveman (2013). 'Freakish patterns – species and species concepts in apomicts'. *Nordic Journal of Botany* **31**(3):257–269.

P. D. N. Hebert, et al. (2003). 'Biological identifications through DNA barcodes'. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **270**(1512):313–321.

O. Henegariu, et al. (1997). 'Multiplex PCR: critical parameters and step-by-step protocol'. *Biotechniques* **23**(3):504–511.

M. Hermann (2009). 'The impact of the European Novel Food Regulation on trade and food innovation based on traditional plant foods from developing countries'. *Food Policy* **34**(6):499–507.

K. Hiller & M. Melzig (2010). *Lexikon der Arzneipflanzen und Drogen: in zwei Bänden*. Spektrum, Akad. Verlag.

P. M. Hollingsworth, et al. (2011). 'Choosing and Using a Plant DNA Barcode'. *PLoS ONE* **6**(5).

References

Y. Holm, et al. (1988). 'Variation of the main terpenes in dragonhead (*Dracocephalum moldavica* L.) during growth'. *Flavour Fragr. J.* **3**(3):113–115.

J. Holub & F. Procházka (2000). 'Red List of vascular plants of the Czech Republic 2000'. *Preslia* **72**(2-4):187–230.

M. K. Huh & K. H. Bang (2006). 'Identification of *Atractylodes japonica* and *A. macrocephala* by RAPD analysis and SCAR Markers'. *Silvae Genetica* **55**(3):101–104.

institut Akademiia nauk SSSR, et al. (? ? ? ?). *Flora of the U.S.S.R.*, vol. v. 20. [Springfield, Va. :Israel Program for Scientific Translations ;. http://www.biodiversitylibrary.org/bibliography/43751.

A. Jenks, et al. (2011). 'Evolution and origins of the Mazatec hallucinogenic sage, *Salvia divinorum* (Lamiaceae): a molecular phylogenetic approach'. *Journal of Plant Research* **124**(5):593–600.

M. R. Joharchi & M. S. Amiri (2012). 'Taxonomic evaluation of misidentification of crude herbal drugs marketed in Iran.'. *Avicenna Journal of Phytomedicine* **2**(2).

M. Kaufmann & M. Wink (1994). 'Molecular systematics of the Nepetoideae (family *Labiatae*): phylogenetic implications from rbcL gene sequences'. *Zeitschrift fur Naturforschung C-Journal of Biosciences* **49**(9):635–645.

P. Khazaeli & M. Mehrabani (2008). 'Screening of Sun Protective Activity of the Ethyl Acetate Extracts of Some Medicinal Plants'. *Iranian Journal of Pharmaceutical Research* **7**(1):5–9.

S. H. Kim & T. Y. Shin (2006). 'Effect of *Dracocephalum argunense* on mast-cell-mediated hypersensitivity'. *International archives of allergy and immunology* **139**(2):87–95.

M. Kimura (1980). 'A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences'. *Journal of Molecular Evolution* **16**(2):111–120.

F. Kitaoka, et al. (2009). 'Molecular characterization of *Akebia* plants and the derived traditional herbal medicine.'. *Biological & pharmaceutical bulletin* **32**(4):665–670.

F. E. Koehn & G. T. Carter (2005). 'The evolving role of natural products in drug discovery'. *Nature Reviews Drug Discovery* **4**(3):206–220.

K. Kondo, et al. (1996). 'Phylogenetic relationship of medicinally important *Cnidium officinale* and Japanese *Apiaceae* based on rbcL sequences'. *Journal of Plant Research* **109**(1):21–27.

L. B. Koski & G. B. Golding (2001). 'The closest BLAST hit is often not the nearest neighbor.'. *Journal of molecular evolution* **52**(6):540–542.

W. J. Kress & D. L. Erickson (2007). 'A Two-Locus Global DNA Barcode for Land Plants: The Coding rbcL Gene Complements the Non-Coding trnH-psbA Spacer Region'. *PLoS ONE* **2**(6):e508+.

W. J. Kress, et al. (2005). 'Use of DNA barcodes to identify flowering plants'. *Proceedings of the National Academy of Sciences of the United States of America* **102**(23):8369–8374.

P. Lazarević, et al. (2009). 'On the distribution of *Dracocephalum ruyschiana* (Lamiaceae) in the Balkan Peninsula'. *Phytologia balcanica* **15**(2):175–179.

M. Y. Lee, et al. (2006). 'Development of SCAR Marker for Discrimination of *Artemisia princeps* and *A. argyi* from Other *Artemisia* Herbs'. *Biological and Pharmaceutical Bulletin* **29**(4):629–633.

S. B. Lee, et al. (2007). 'The antimicrobial activity of essential oil from *Dracocephalum foetidum* against pathogenic microorganisms'. *JOURNAL OF MICROBIOLOGY-SEOUL-* **45**(1):53.

X. Li, et al. (2007). 'Molecular authentication of *Alisma orientale* by PCR-RFLP and ARMS'. *Planta medica* **73**(1):67.

T. Lindahl (1993). 'Instability and decay of the primary structure of DNA'. *Nature* **362**(6422):709–715.

D. P. Little (2011). 'DNA Barcode Sequence Identification Incorporating Taxonomic Hierarchy and within Taxon Variability'. *PLoS ONE* **6**(8):e20552+.

References

M. Lou & G. B. Golding (2007). 'fingerprint: visual depiction of variation in multiple sequence alignments'. *Molecular Ecology Notes* **7**(6):908–914.

H. Lou, et al. (2012). 'Eriodictyol protects against H2O2-induced neuron-like PC12 cell death through activation of Nrf2/ARE signaling pathway'. *Neurochemistry International* **61**(2):251–257.

K. Lu, et al. (2005). 'Identification of saposhinkoviae radix in concentrated Chinese medicine preparations by nested PCR and DNA sequencing methods'. *Journal of Food and Drug Analysis* **13**(3):219.

X. Q. Ma, et al. (2000). 'Species identification of Radix Astragali (Huangqi) by DNA sequence of its 5S-rRNA spacer domain'. *Phytochemistry* **54**(4):363–368.

M. Martínez-Vázquez, et al. (2012). 'Neuropharmacological study of *Dracocephalum moldavica* L. (Lamiaceae) in mice: Sedative effect and chemical analysis of an aqueous extract'. *Journal of Ethnopharmacology* **141**(3):908–917.

J. Masterson (1994). 'Stomatal Size in Fossil Plants: Evidence for Polyploidy in Majority of Angiosperms'. *Science* **264**(5157):421–424.

R. L. Mayden (1997). 'A hierarchy of species concepts: the denouement in the saga of the species problem'. *Systematics Association Special Volume* **54**:381–424.

E. Mayr (1940). 'Speciation phenomena in birds'. *American Naturalist* pp. 249–278.

E. Mayr (1942). *Systematics and the Origin of Species: From the Viewpoint of a Zoologist*. No. 13. Harvard University Press.

R. Meier, et al. (2006). 'DNA Barcoding and Taxonomy in Diptera: A Tale of High Intraspecific Variability and Low Identification Success'. *Systematic Biology* **55**(5):715–728.

L. Meng, et al. (2009). 'Polyphenol extraction from eight Perilla frutescens cultivars'. *Comptes Rendus Chimie* **12**(5):602–611.

C. P. Meyer & G. Paulay (2005). 'DNA Barcoding: Error Rates Based on Comprehensive Sampling'. *PLoS Biol* **3**(12):e422+.

L. N. Misra & A. Ahmad (1992). 'An Oxygenated Tetrahydrobergamotene from the Essential Oil of *Dracocephalum nutans**.'. *Planta medica* **58**(5):478–479.

L. N. Misra, et al. (1988). 'Volatile constituents of Dracocephalum nutans'. *Planta medica* **54**(02):165–166.

N. Moslem, et al. (2009). 'Effects of total extract of Dracocephalum *moldavica* on ischemia / reperfusion induced arrhythmias and infarct size in the isolated rat heart'. *Iranian Journal of basic medicinal sciences* **11**(4):229–235.

J. Mower, et al. (2007). 'Extensive variation in synonymous substitution rates in mitochondrial genes of seed plants'. *BMC Evolutionary Biology* **7**(1):135+.

K. Mullis, et al. (1986). 'Specific Enzymatic Amplification of DNA In Vitro: The Polymerase Chain Reaction'. *Cold Spring Harbor Symposia on Quantitative Biology* **51**:263–273.

J. Měsíček & J. Soják (1973). 'Karyological and taxonomic observations on *Dracocephalum foetidum* Bunge and *Koenigia islandica* L.' **8**(1):105–112.

E. W. Myers & W. Miller (1988). 'Optimal alignments in linear space'. *Computer applications in the biosciences : CABIOS* **4**(1):11–17.

M. Nei & M. Nozawa (2011). 'Roles of mutation and selection in speciation: from Hugo de Vries to the modern genomic era.'. *Genome biology and evolution* **3**:812–829.

C. R. Newton, et al. (1989). 'Analysis of any point mutation in DNA. The amplification refractory mutation system (ARMS)'. *Nucleic Acids Research* **17**(7):2503–2516.

A. S. Nikitina, et al. (2008). 'Studies of the essential oil of *Dracocephalum moldavica* cultivated in the Stavropol region' **42**(4):203–207.

References

D. R. Nuñez & C. O. de Castro (1990). 'Hybridization between *Sideritis serrata* Lag. and *Sideritis bourgaeana* Boiss. (Lamiaceae) in Their Hybrid Zone in Spain'. *Annals of Botany* **66**(2):147–154.

S. R. Numonov, et al. (2013). 'Chemical composition of *Dracocephalum heterophyllum*'. *Chemistry of Natural Compounds* **49**(3):511–513.

R. V. Oakley, et al. (2007). 'Differential Expansion and Expression of α- and β-Tubulin Gene Families in *Populus'*. *Plant Physiology* **145**(3):961–973.

G. B. Oganesyan (2009). 'Minor flavonols from *Dracocephalum multicaule*'. *Chemistry of Natural Compounds* **45**(2):242–243.

I.-C. Oh, et al. (2008). 'Comparative seed morphology and character evolution in the genus *Lysimachia* (*Myrsinaceae*) and related taxa' **271**(3-4):177–197.

R. G. Olmstead, et al. (1992). 'Monophyly of the *Asteridae* and Identification of Their Major Lineages Inferred From DNA Sequences of rbcL'. *Annals of the Missouri Botanical Garden* **79**(2).

W. H. Organization & Others (2002). 'WHO traditional medicine strategy 2002-2005.' .

I. Paran & R. W. Michelmore (1993). 'Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce'. *TAG Theoretical and Applied Genetics* **85**(8):985–993.

M.-J. Park, et al. (2006). 'Molecular identification of Korean ginseng by amplification refractory mutation system-PCR'. *Food Research International* **39**(5):568–574.

H. E. Pence & A. Williams (2010). 'ChemSpider: An Online Chemical Information Resource'. *J. Chem. Educ.* **87**(11):1123–1124.

V. Povilaityee & P. R. Venskutonis (2000). 'Antioxidative activity of purple peril (*Perilla frutescens* L.), moldavian dragonhead (*Dracocephalum moldavica* L.), and roman chamomile (*Anthemis nobilis* L.) extracts in rapeseed oil' **77**(9):951–956.

V. Povilaityté, et al. (2001). 'Antioxidant properties of Moldavian Dragonhead (*Dracocephalum moldavica* L.)'. *Journal of Food Lipids* **8**(1):45–64.

S. Ratnasingham & P. D. N. Hebert (2007). 'bold: The Barcode of Life Data System (http://www.barcodinglife.org)'. *Molecular Ecology Notes* **7**(3):355–364.

D.-M. Ren, et al. (2008). 'Simultaneous determination of nine major active compounds in *Dracocephalum rupestre* by HPLC'. *Journal of Pharmaceutical and Biomedical Analysis* **48**(5):1441–1445.

L. H. Rieseberg & L. Brouillet (1994). 'Are many plant species paraphyletic? '. *Taxon* pp. 21–32.

W. Rothmaler & E. J. Jäger (2008). *Bd. 5: Krautige Zier-und Nutzpflanzen*. Heidelberg [etc.]: Spektrum.

N. Saitou & M. Nei (1987). 'The neighbor-joining method: a new method for reconstructing phylogenetic trees.'. *Molecular Biology and Evolution* **4**(4):406–425.

I. N. Sarkar, et al. (2008). 'caos software for use in character-based DNA barcoding'. *Molecular Ecology Resources* **8**(6):1256–1259.

A.-C. Scheen & V. A. Albert (2009). 'Molecular Phylogenetics of the Leucas Group (Lamioideae; *Lamiaceae*)'. *Systematic Botany* **34**(1):173–181.

O. Seberg & G. Petersen (2009). 'How Many Loci Does it Take to DNA Barcode a Crocus? '. *PLoS ONE* **4**(2):e4598+.

E. Selenge, et al. (2013). 'Flavone Tetraglycosides and Benzyl Alcohol Glycosides from the Mongolian Medicinal Plant *Dracocephalum ruyschiana*'. *J. Nat. Prod.* **76**(2):186–193.

L. Sen, et al. (2011). 'Molecular evolution of rbcL in three gymnosperm families: identifying adaptive and coevolutionary patterns'. *Biology Direct* **6**(1):29.

J. Shaw, et al. (2005). 'The tortoise and the hare II: relative utility of 21 noncoding chloroplast DNA sequences for phylogenetic analysis'. *American Journal of Botany* **92**(1):142–166.

M. Shiba, et al. (2006). 'Identification of medicinal *Atractylodes* based on ITS sequences of nrDNA'. *Biological and Pharmaceutical Bulletin* **29**(2):315.

T. Shuge, et al. (2009). 'Essential Oil Composition of the *Dracocephalum moldavica* L from Xinjiang in China'. *Pharmacognosy Research* **1**(4):172.

G. G. Simpson (1961). *Principles of animal taxonomy*, vol. 20. Columbia University Press.

H. O. Smith & K. W. Welcox (1970). 'A Restriction enzyme from *Hemophilus influenzae*'. *Journal of Molecular Biology* **51**(2):379–391.

M. L. Sogin, et al. (2006). 'Microbial diversity in the deep sea and the underexplored "rare biosphere"'. *Proceedings of the National Academy of Sciences* **103**(32):12115–12120.

D. E. Soltis & P. S. Soltis (1999). 'Polyploidy: recurrent formation and genome evolution'. *Trends in Ecology & Evolution* **14**(9):348–352.

C. L. Stebbins & Others (1950). 'Variation and evolution in plants.'. *Variation and evolution in plants.* .

M. Y. Stoeckle, et al. (2011). 'Commercial Teas Highlight Plant DNA Barcode Identification Successes and Obstacles'. *Sci. Rep.* **1**.

A. A. Taherpour, et al. (2011). 'Chemical composition analysis of the essential oil of *Melissa officinalis* L. from Kurdistan, Iran by HS/SPME method and calculation of the biophysicochemical coefficients of the components'. *Natural Product Research* **26**(2):152–160.

K. Tamura, et al. (2011). 'MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods.'. *Molecular biology and evolution* **28**(10):2731–2739.

T. Tanabata, et al. (2012). 'SmartGrain: High-Throughput Phenotyping Software for Measuring Seed Shape through Image Analysis'. *Plant Physiology* **160**(4):1871–1880.

P. Theerakulpisut, et al. (2008). 'Development of species-specific SCAR markers for identification of three medicinal species of *Phyllanthus*'. *Journal of systematics and evolution* **46**(4):614–621.

S. Theodoridis, et al. (2012). 'DNA barcoding in native plants of the Labiatae (*Lamiaceae*) family from Chios Island (Greece) and the adjacent Çeşme-Karaburun Peninsula (Turkey)'. *Molecular Ecology Resources* **12**(4):620–633.

J. Thompson (1989). 'A revision of the genus *Leptospermum* (*Myrtaceae*)'. *Telopea* **3**:301–449.

A. Valentini, et al. (2009a). 'New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach'. *Molecular Ecology Resources* **9**(1):51–60.

A. Valentini, et al. (2009b). 'DNA barcoding for ecologists'. *Trends in Ecology & Evolution* **24**(2):110–117.

J.-L. Vanherweghem, et al. (1993). 'Rapidly progressive interstitial renal fibrosis in young women: association with slimming regimen including Chinese herbs'. *The Lancet* **341**(8842):387–391.

A. Varma, et al. (2007). 'Plant genomic DNA isolation: An art or a science'. *Biotechnology Journal* **2**(3):386–392.

G. Vaughton & M. Ramsey (1998). 'Sources and consequences of seed mass variation in *Banksia marginata* (Proteaceae)'. *Journal of Ecology* **86**(4):563–573.

R. Verpoorte (1998). 'Exploration of nature's chemodiversity: the role of secondary metabolites as leads in drug development'. *Drug Discovery Today* **3**(5):232–238.

P. Vos, et al. (1995). 'AFLP: a new technique for DNA fingerprinting'. *Nucleic Acids Research* **23**(21):4407–4414.

L. J. Wallace, et al. (2012). 'DNA barcodes for everyday life: Routine authentication of Natural Health Products'. *Food Research International* **49**(1):446–452.

N. G. Walsh, et al. (1996). *Flora of Victoria. Volume 3: dicotyledons Winteraceae to Myrtaceae.* Inkata Press.

References

J. Wang, et al. (2001). 'Application of sequence characterized amplified region (SCAR) analysis to authenticate *Panax* species and their adulterants'. *Planta medica* **67**(08):781–783.

C. Z. Wang, et al. (2007). 'Simultaneous identification of Bulbus Fritillariae cirrhosae using PCR-RFLP analysis'. *Phytomedicine* **14**(9):628–632.

L. Wang, et al. (2012). 'Phenolic alkaloids from the aerial parts of *Dracocephalum heterophyllum*'. *Phytochemistry* **82**:166–171.

M. Wang, et al. (2013). 'Potential use of DNA barcoding for the identification of *Salvia* based on cpDNA and nrDNA sequences'. *Gene* **528**(2):206–215.

M. Westoby, et al. (1992). 'Comparative evolutionary ecology of seed size'. *Trends in Ecology & Evolution* **7**(11):368–372.

B. WhiteTJ, et al. (1990). 'Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics'. *PCR Protocols: A guide to methods and applications.(Innis MA, Gelfand DH, Sninsky JJ, White TJ, eds.). Academic Press, New York* pp. 315–322.

J. A. Wiens (1989). 'Spatial scaling in ecology'. *Functional ecology* **3**(4):385–397.

E. O. Wiley (1978). 'The Evolutionary Species Concept Reconsidered'. *Systematic Biology* **27**(1):17–26.

J. G. K. Williams, et al. (1990). 'DNA polymorphisms amplified by arbitrary primers are useful as genetic markers'. *Nucleic Acids Research* **18**(22):6531–6535.

A. C. Wilson (1989). 'Ancient DNA and the polymerase chain reaction' .

E. O. Wilson (2003). 'The encyclopedia of life'. *Trends in Ecology & Evolution* **18**(2):77–80.

P. Xiao, et al. (1984). 'Ethnopharmacologic study of chinese rhubarb'. *Journal of Ethnopharmacology* **10**(3):275–293.

Z.-Y. Yang, et al. (2007). 'ITS sequence analysis used for molecular identification of the *Bupleurum* species from northwestern China'. *Phytomedicine* **14**(6):416–423.

M. Yavari, et al. (2011). 'Composition and Antibacterial Activity of the Essential Oil of a Green Type and a Purple Type of *Ocimum basilicum* L. from Iran'. *Journal of Essential Oil Research* **23**(1):1–4.

Q.-G. Ye, et al. (2006). 'Potential Risk of Hybridization in Ex Situ Collections of Two Endangered Species of Sinojackia Hu (*Styracaceae*)'. *Journal of Integrative Plant Biology* **48**(7):867–872.

W. Zheng-Yi & P. H. Raven (1994). 'Flora of China. Vol. 17'.

Y.-P. Zhu (2002). 'Toxicity of the Chinese Herb Mu Tong (*Aristolochia manshuriensis*)' **21**(4):171–177.

X.-F. Zhu, et al. (2009). 'Molecular and morphological evidence for natural hybridization between *Primula secundiflora* Franchet and *P. Poissonii* Franchet (*Primulaceae*)'. *Acta Biol Cracov Bot* **51**:29–36.