

Article

# Applied Barcoding: The Practicalities of DNA Testing for Herbals

Caroline Howard <sup>1,2,\*</sup>, Claire Lockie-Williams <sup>2</sup> and Adrian Slater <sup>1,\*</sup> 

<sup>1</sup> Biomolecular Technology Group, Leicester School of Allied Health Science, Faculty of Health and Life Sciences, De Montfort University, Leicester LE1 9BH, UK

<sup>2</sup> BP-NIBSC Herbal Laboratory, National Institute for Biological Standards and Controls, Potters Bar EN6 3QG, UK; Claire.Lockie-Williams@nibsc.org

\* Correspondence: choward@dmu.ac.uk (C.H.); ads@dmu.ac.uk (A.S.)

Academic Editor: Erica Lumini

Received: 7 June 2020; Accepted: 28 August 2020; Published: 4 September 2020



**Abstract:** DNA barcoding is a widely accepted technique for the identification of plant materials, and its application to the authentication of commercial medicinal plants has attracted significant attention. The incorporation of DNA-based technologies into the quality testing protocols of international pharmacopoeias represents a step-change in status, requiring the establishment of standardized, reliable and reproducible methods. The process by which this can be achieved for any herbal medicine is described, using *Hypericum perforatum* L. (St John's Wort) and potential adulterant *Hypericum* species as a case study. A range of practical issues are considered including quality control of DNA sequences from public repositories and the construction of individual curated databases, choice of DNA barcode region(s) and the identification of informative polymorphic nucleotide sequences. A decision tree informs the structure of the manuscript and provides a template to guide the development of future DNA barcode tests for herbals.

**Keywords:** applied barcoding; *rbcL*; *matK*; *psbA-trnH*; nrITS; intron; *Hypericum*

## 1. Introduction

The use of DNA-based methods to identify herbal materials has been widely accepted as a complementary method to phytochemical and physical testing methods [1–6]. The ability of DNA-based methods to detect and identify contaminating materials has been its most beneficial feature, and it is this quality that sets these methods apart from others. The information provided by a well-performing, controlled and standardized genetic method enables producers, manufacturers and regulators to increase the quality of products, and prevent toxic materials from entering the supply chain [7].

The literature provides many examples of new methodologies, high sensitivity techniques and methods aimed at processed materials and degraded samples [5,8,9]. However, it is widely accepted that the most beneficial and appropriate application of DNA-based methods to the herbal industry will be upstream of production [7]. Ideally, these methods are applied to select seeds to be grown under conditions in line with cGACP guidelines. Although cultivated materials are preferred by the industry, a large proportion of medicinal plant material is still wild-harvested and this is arguably the most important area to which DNA barcoding methods should be applied [7]. Before any processing has occurred, DNA-based identification of the raw materials can provide a certainty of identity that is unparalleled, together with the detection of contaminating plant material.

As DNA barcoding projects have advanced, the steady accumulation of barcode sequences in public databases has proved to be a valuable tool for the design of DNA barcode authentication methods [10]. However, this wealth of information requires to be treated with caution to avoid poor

quality and incorrectly labeled sequences. Furthermore, the abundance of sequences may create a level of “background noise” that can influence and misdirect the design of barcoding assays.

St John’s Wort (*Hypericum perforatum* L.) is one of the world’s leading herbal medicines in terms of global sales. This position in the marketplace has attracted the application of novel, DNA-based, identification techniques since 2004 [11], before the advent of plant DNA barcoding programs. Since then, the identification of medicinal plant material by DNA barcoding has been demonstrated for a significant number of species [1,3–6], including *H. perforatum* [12–16]. The DNA barcodes used for individual medicinal plants have generally been chosen on an ad hoc basis, depending on practical issues such as amplifiability and the need to discriminate between the target species and likely adulterants, rather than fundamental measurements of barcode effectiveness [17–19]. The reason for choosing one particular barcode region is often not made explicit in published work and may be the result of rather idiosyncratic decisions known only to the authors.

This manuscript provides a reasoned approach, informed by years of experience, toward applying DNA-based identification methods to a “target” commercial medicinal plant. A step-by-step process is described using St John’s Wort as a case study, providing a template that can be reproduced by researchers, companies and regulators looking toward the implementation of DNA-based methods for the first time.

## 2. Results

### 2.1. A Scheme for Applied Barcoding

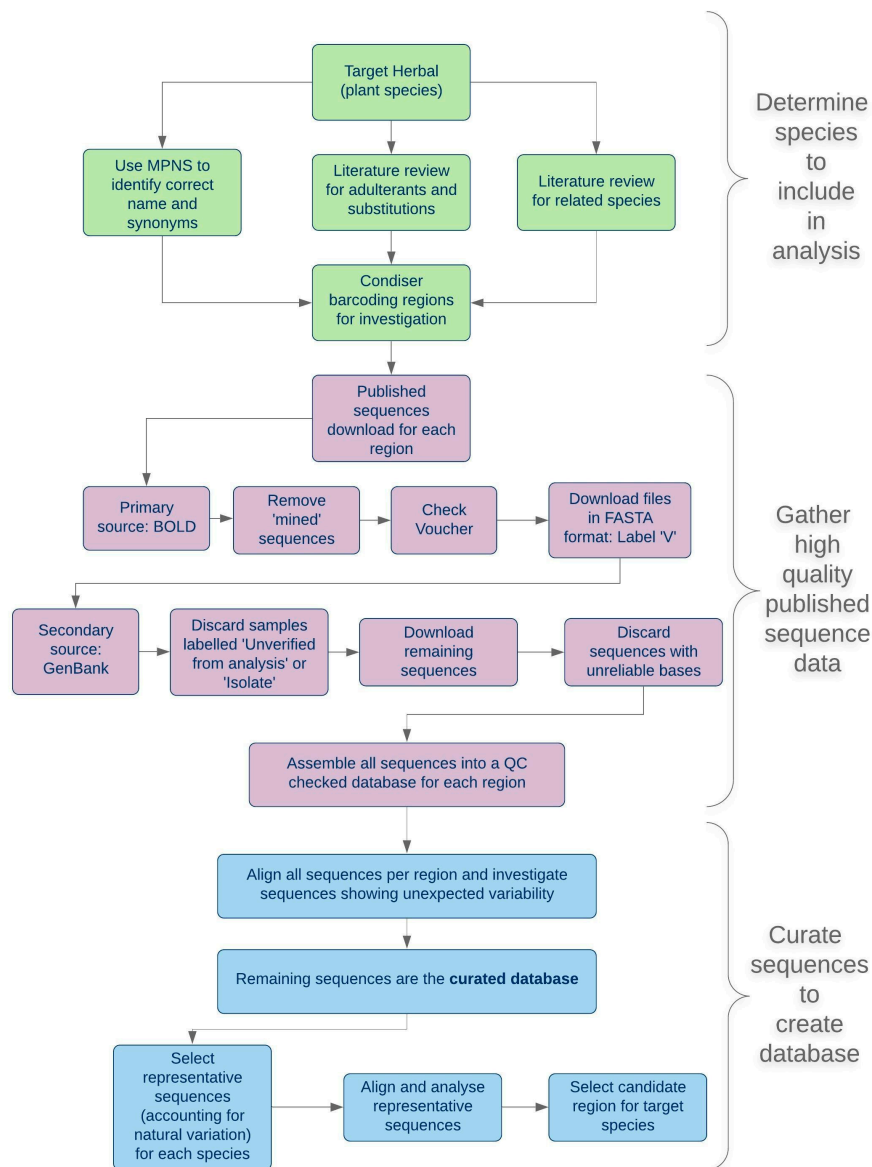
The choice of DNA barcode region and testing platform for a specified target herbal will depend upon practical considerations as well as more theoretical parameters of barcode efficiency. These include background information about the target species and known adulterants, the extent of DNA sequence data for different barcode regions and the availability of reference materials for test development. A flow diagram of a pragmatic approach to choosing a DNA barcode for herbal authentication is shown in Figure 1. Although an increasingly large number of DNA barcode sequences have been published, the quality and veracity of these are often questionable, and databases have numerous examples of sequences assigned to incorrect species [20,21]. For this reason, the decision tree in Figure 1 was developed to avoid some of the common pitfalls and provide a method to select high-quality sequences.

#### 2.1.1. Nomenclature, Taxonomy and Adulteration

The flow diagram proposes that three strands of background information about the medicinal plant of interest should be collated in parallel, prior to consideration of the DNA barcodes: current taxonomic information about the plant species and its relatives; confirmation of the correct botanical nomenclature and medicinal/common names, and investigation of adulteration issues and their underlying causes. Adulteration of herbals could arise from unintended contamination, misidentification during wild-harvesting, incorrect nomenclature, legitimate substitution or deliberate fraud.

There is considerable taxonomic information available about the genus *Hypericum* from both morphological studies captured in Robson’s monograph [22], and complementary chemical [23,24] and molecular studies based primarily on the ITS region [25–27]. This information immediately indicates a problem for DNA barcoding—this is a very large genus comprising some 490 species assigned to 36 Sections [22,28]. However, the main objectives of DNA barcoding a commercial medicinal plant are to successfully identify the target plant and to differentiate it from potential adulterants. The phylogenetic context of a plant provides valuable information about the relationships between target and potential adulterants, particularly the most closely related species. *H. perforatum* and its closest relatives are members of the Section *Hypericum* [29]. These “sister” species provide a “worst-case scenario” to test against, based on the presumption that being able to differentiate between the “target” and a “sister” species will be the most challenging objective. *H. maculatum* is the closest sister species to *H. perforatum*,

grows in close proximity [30] and is a known adulterant [31], so the ability to discriminate between these two species can be regarded as a key requirement of the chosen barcode.



**Figure 1.** Flow diagram of DNA barcode region selection for discrimination of a commercial medicinal species from likely adulterant species. This shows a series of steps and decision points that can be used as a template to guide the choice of a suitable DNA barcode from published information through to laboratory analysis.

The nomenclature of *H. perforatum* and other members of the genus with medicinal properties can be confirmed at the Medicinal Plant Names Services (MPNS) [32]. This valuable resource lists botanical synonyms along with common and medicinal names. For example, *H. perforatum* L. is confirmed as the accepted name and attribution with the highest quality rating for confidence in the taxonomy, and three illegitimate scientific synonyms are noted. Issues around the nomenclature of *H. perforatum* subspecies are discussed in a separate MPNS publication [33]. There are 142 nonscientific names listed with the corresponding medicinal plant reference sources. Fifty medicinal reference sources citing *H. perforatum* are also listed. A number of other *Hypericum* species have medicinal properties and can also be found in this database (see Table 6).

There are several known adulterants of St John's Wort herbal preparations within the genus *Hypericum* [31]. These range from close relatives which grow in similar habitats (e.g., *H. barbatum*, *H. hirsutum*, *H. maculatum*, *H. montanum* and *H. tetraptrum* in Europe [31]) to plants which grow widely in geographical areas where *H. perforatum* is less common (e.g., *H. patulum* in India [34], *H. crux-andreae* in the New World [31] and possibly *H. undulatum* in China [35,36]).

The three strands of background information underpin the next stage, which is to consider possible DNA barcode regions. The initial literature review should include a search for prior recommendations of DNA barcodes for authentication of the target plant. There are relatively few published recommendations for DNA authentication of *Hypericum*, mainly focusing on the nuclear ribosomal internal transcribed spacer region (ITS) [11–13], but also *matK* [15].

### 2.1.2. Selection of a Panel of Potential Adulterant *Hypericum* Species

The key function of a DNA barcode selected for authentication purposes is to discriminate between the correct “target” species and potential adulterant species. The barcode needs only to be unique for the target species. Provided all the nontarget barcodes differ from the target species, it is not essential for them also to differ from each other. Furthermore, the scope of relevant nontarget species can be restricted in order to dramatically simplify the problem. The underlying principle is that if the adulterants are known, it is unnecessary to be concerned with rare or geographically remote relatives of the target species [12]. On this basis, a panel of 20 *Hypericum* species was selected comprising close relatives of *H. perforatum*, common commercial species and known adulterants (Table 6).

### 2.1.3. Public Databases of DNA Barcode Sequences

The next stage is to search publicly available databases of existing DNA barcode sequences to create a collection for further study. The flowchart contrasts two major databases, describing the Barcode of Life Database (BOLD) [37] as the “primary source” and NCBI GenBank [10] as a secondary source, based on the reliability of source material identification and the accuracy of sequence data. The BOLD database is constructed specifically to hold complete records of DNA barcode projects. Full individual specimen records will include herbarium voucher details and photographs, collection details, DNA barcode sequence with supporting electropherogram traces for forward and reverse primers. However, BOLD also skims GenBank for barcode sequences, so it is important to differentiate full records from GenBank records. For example, there are 40 *rbcL* sequences for *H. perforatum* on BOLD but only 22 of these are full records with herbarium sheet photographs and sequence traces (Table 1).

One obvious point is that the more sequences are already publicly available (Table 1), the less *de novo* sequencing work is required in the laboratory. For example, despite the 40 *rbcL* sequences for *H. perforatum*, there is relatively poor coverage of the other panel species. In contrast, while there are also 40 ITS sequences available for *H. perforatum* between the two databases, there is also very good coverage of the other panel species. There are relatively few *psbA-trnH* sequences for most individual species, but the coverage of the panel species is nearly complete. On the other hand, *matK* is very poorly represented in either database. The prior selection of a panel of nontarget plants allows a manageable number of barcode sequences to be collected from the public databases, resulting in the creation of a small database of representative panel barcode sequences that can be analyzed and manually curated.

When these data are in place, a multiple alignment of each region will allow the level of intra- and interspecific variation to be determined. Having multiple accessions of each species at this stage may allow unreliable accessions to be detected and rejected, and consensus or representative sequences of each species to be selected to form the panel database. The objective at this stage is the identification of barcodes that show potential for defining a species-specific reference sequence against which to match test samples. Where there are gaps in the available published data, there will remain a question mark over the applicability of a particular barcode region—this is acceptable and can be addressed further on in the process. The practicalities of each region may be very different from one target species to another and should be confirmed experimentally.

**Table 1.** Availability of *Hypericum* panel barcode sequences in Barcode of Life Database (BOLD) and GenBank. Figures reported for BOLD represent the number of full accessions, with the total number, including those mined from GenBank, shown in parentheses. (Figures from March 2020).

Species	BOLD					GenBank				
	<i>rbcL</i>	<i>matK</i>	<i>trnH-psbA</i>	ITS	ITS2	<i>rbcL</i>	<i>matK</i>	<i>trnH-psbA</i>	ITS	ITS2
<i>H. acmosepalum</i>					1			1	5	
<i>H. androsaemum</i>	5 (6)	1			1	7	1	1	7	4
<i>H. ascyron</i>	1 (2)	1		1	2 (4)	5	3	1	16	3
<i>H. athoum</i>					1			1	4	
<i>H. attenuatum</i>					2				7	
<i>H. barbatum</i>					1				2	
<i>H. calycinum</i>	1	1			2	1	1	1	7	
<i>H. crux-andreae</i>						2		1	2	
<i>H. delphicum</i>					2				6	
<i>H. elegans</i>								1		
<i>H. hirsutum</i>	4 (5)		1	(1)	1 (3)	6		1	5	2
<i>H. japonicum</i>	2				3	9		5	10	7
<i>H. kouytchense</i>					2			1	4	
<i>H. maculatum</i>	7 (8)	(1)		1	1 (9)	11		3	14	3
<i>H. montanum</i>	2 (3)	1		1	2	3		1	6	1
<i>H. olympicum</i>					1			2	9	
<i>H. patulum</i>					2			1	8	2
<i>H. perforatum</i>	22 (40)	3 (6)	(3)	(1)	9 (27)	38	11	7	40	17
<i>H. tetrapterum</i>	2 (4)	1 (2)			4	3	2	2	5	2
<i>H. undulatum</i>	3 (4)	1			1	4	1	1	9	4

#### 2.1.4. Reference Samples for Pilot Studies and Assay Standards

Having identified the panel species, a reference sample set must be established, and each sample verified as a true example of the species that it is intended to represent. Many laboratories will not have the resources or expertise to collect sufficient numbers of verified reference specimens, but may be able to find the relevant species represented in the DNA Bank schemes [38–40] operated by various botanical gardens, with DNA samples available from vouchered herbarium specimens. Once assembled, this reference sample set will provide a “proof of concept” study for candidate barcode regions.

Reference DNA samples of this type were obtained for most panel species (Table S1). The ease of amplification of each barcode region with standard primers differs between species, and this can only be measured experimentally. The amplification of each barcode region also serves to produce sequence data for any species within the reference sample set that was not available in public databases. These data can then be added to the original panel database and provide further validation of the selected barcode region.

#### 2.1.5. Test Samples for Barcode Assay Validation

A set of samples of *H. perforatum* and its closest sister species, *H. maculatum* were collected from different geographical locations around Lithuania (Table S2, Figure S2). These were used as a set of “unknown” samples to confirm the ability of each barcode to successfully distinguish *H. perforatum* samples from their closest relative. Once a set of test criteria were set for chosen barcode regions,

these samples were used as “unknown” samples to determine whether the barcode test agreed with the botanical identification, and whether there was consistency between the barcodes.

## 2.2. The nrITS Barcode Region

### 2.2.1. A Curated Database of ITS Sequences

The availability of published *Hypericum* ITS sequences from 2004 [11] was a major pragmatic reason for the authors choosing this barcode region for developing authentication tests from 2007 onwards. Since then, the number of *Hypericum* sequences available has grown considerably as the data from large-scale phylogenetic studies have been deposited in GenBank. One problem with this abundance of information is that the length and quality of accessions and the reliability of their original identification is inconsistent. As a result, a BLAST search of the database with a genuine *H. perforatum* ITS sequence will often include sequences from a number of different species in the top 50 hits, some having higher scores than genuine *H. perforatum* accessions. In order to filter this “background noise” effect, irrelevant and unreliable sequences were identified using the BLAST distance tree facility to identify and discard obvious outliers (typically singleton accessions with less than 95% identity to any other *H. perforatum* sequence) [20,21]. Tightly clustering groups of sequences unique to individual species (particularly vouchered specimens from a variety of sources) were collected and aligned. Where a small number of sequences showed near to 100% identity, a representative accession was chosen for the curated database. Species represented by larger numbers of accessions were typified by highly similar sequences with infrequent polymorphisms in individual accessions. In these cases, an artificial consensus sequence was constructed to represent the species. Where there were sufficient accessions to detect consistent patterns of variation within the species, consensus sequences representing each group of variants were designed, resulting in three *H. perforatum* ITS subtypes and two for *H. maculatum*. The ITS curated database is shown as a multiple alignment in Figure S1.

The alignment highlights several points of wider relevance. One is the clustering of regions of variation between species into “hotspots” within the ITS1 and ITS2 regions [13]. (The 5.8S coding sequence between the ITS1 and ITS2 regions is completely conserved.) These hotspots are ideal locations in which to detect “mini-barcodes” for targeting by PCR primers [12–14,41]. However, the main boundaries of variation at these hotspots can be seen to fall between Sections or larger clades. In consequence, there is very little variation between the members of the Section *Hypericum*, and most particularly between *H. maculatum* and *H. perforatum*. There are very few positions where both *H. maculatum* subtypes consistently differ from all three of the *H. perforatum* subtypes. Position 87 is unique to *H. perforatum*, position 467 is unique to *H. maculatum*, and 658/9 are not unique to either, but discriminate between the two species.

An advantage of starting with a large sequence dataset is that polymorphisms specific to one or other subtype of *H. perforatum* or *H. maculatum* can be detected. Limited sequence data might lead to the choice of these subtype polymorphisms as key determinants of a target or adulterant species, with subsequent false-positives or negatives. For example, a limited set of *H. perforatum* sequences that were all subtype I might indicate that the polymorphism at position 10 was a unique species marker of *H. perforatum*, but then give false negatives for any subtype II and III *H. perforatum* samples.

### 2.2.2. ITS Barcoding of Reference Samples

The ITS region of all the reference DNA samples was amplified and sequenced successfully, apart from the *H. delphicum* sample 13938. Alignment of the sequences with the *Hypericum* panel ITS database showed that each reference species ITS sequence had the closest match (at least 99% identity) to the corresponding panel sequence, apart from *H. perforatum* sample 13876 (due to a poor quality sequence) and *H. maculatum* sample 13896 (which was not a close match to the reference *H. maculatum* sequences and showed closer resemblance to members of the Section *Ascyreia*).

### 2.3. The ITS2 Barcode Region

#### 2.3.1. ITS2 Sequences

The ITS2 subregion of ITS has been championed as the ideal barcode for medicinal plants [42–44]. Separate ITS2 barcoding projects have been conducted, while ITS2 sequences can also be extracted from full ITS accessions. The ITS2 region seen in Figure S1 contains sufficient variation to differentiate the panel members from *H. perforatum* but it is inevitably more restricted than using the entire ITS region.

#### 2.3.2. Secondary Structure of ITS2 Sequences

One advantage of the ITS2 region is that it is possible to predict the secondary structure of the transcribed spacer RNA and use this information to support one-dimensional sequence alignments [45,46]. The ITS2 Ribosomal RNA database contains direct fold and homology modeled secondary structures for many animal and plant species, including 18 of the *Hypericum* panel. These models were aligned using a secondary structure as well as sequence information (Figure 2).



**Figure 2.** Alignment of panel species ITS2 sequences with RNA secondary structure represented by a single letter (pseudo amino acid) code [47].

The alignment of sequence plus secondary structure ensures that the alignment is more robust and base-pairing information from the RNA structure provides additional points of difference between species. For example, *H. perforatum* differs from *H. maculatum* at positions 31, 41, 219 and 235 and 236 (Figure 2), whereas a simple sequence alignment shows only three SNPs (Figure S1)

It is possible to analyse complementary base pair changes (CBCs) between ITS2 sequences. These are polymorphic sites which are complemented by a second polymorphism to maintain

base-pairing in stems of the RNA secondary structure. These have been hypothesized to follow species boundaries [48–50], but analysis of the number of CBCs between *H. perforatum* and the panel species indicates that they tend to map the clades defined by Meseguer et al. [26] (Table 2). Thus, there are no CBCs that distinguish any of the Section Hypericum or the related Sections in clade E. The species in clade D show one CBC, clade C has two CBCs and B has 4–5 CBCs. The one anomaly is that *H. ascyron* in the Section Roscyna in clade D has five CBCs. This matches the large number of differences seen in the primary sequence between this species and the rest of the panel, which cast doubt on the original botanical identification of these accessions.

**Table 2.** Complementary base pair changes (CBC) in ITS2 RNA secondary structures, relative to the *H. perforatum* ITS2 secondary structure, calculated using the 4Sale program [48,51].

Species	Section	Robson Clade [24]	Meseguer Clade [26]	CBC/Perf
<i>H. delphicum</i>	Adenosepalum	27	E	0
<i>H. montanum</i>	Adenosepalum	27	E	0
<i>H. barbatum</i>	Drosocarpium	13	E	0
<i>H. attenuatum</i>	Hypericum	9	E	0
<i>H. maculatum</i>	Hypericum	9	E	0
<i>H. perforatum</i>	Hypericum	9	E	0
<i>H. tetrapterum</i>	Hypericum	9	E	0
<i>H. undulatum</i>	Hypericum	9	E	0
<i>H. hirsutum</i>	Taenocarpium	18	E	0
<i>H. olympicum</i>	Oligostema	14	E	0
<i>H. acmosepalum</i>	Ascyreia	3	D	1
<i>H. calycinum</i>	Ascyreia	3	D	1
<i>H. kouytchense</i>	Ascyreia	3	D	1
<i>H. patulum</i>	Ascyreia	3	D	1
<i>H. androsaemum</i>	Androsaemum	5	C	2
<i>H. crux-andreae</i>	Myriandra	20	B	4
<i>H. ascyron</i>	Roscyna	7	D	5
<i>H. japonicum</i>	Trignobrathys	30	B	5

## 2.4. The *rbcLa* Barcode Region

### 2.4.1. A Curated Database of *rbcLa* Sequences

As shown in Table 6, there are a limited number of *rbcL* sequences in the BOLD database and not many more in GenBank, apart from *H. perforatum*, which is well represented. The alignment of these sequences shows there to be very little variation between species, particularly those within one Section or closely related sections. A limited haplotype map was constructed, indicating SNPs that were present at the same position in all the accessions from more than one species. Five positions were mapped (180, 300, 372, 379, 492) where the patterns of coinheritance effectively split the panel into two main haplotypes: 1 and 2 (Table 3).

There were sufficient *H. perforatum* accessions to detect two species-related SNPs. The T at position 66 was found in all of the *H. perforatum* accessions, while the T at 263 was found in a large majority. Neither SNP was found in any other accession, particularly the eight *H. maculatum* BOLD accessions. This specific haplotype pattern was designated type 1p. There appeared to be several other species-related SNPs in the *rbcL* barcode, but these are not directly relevant to *H. perforatum* authentication and would require more sequence data to ascertain their significance.



**Table 3.** Haplotype map of the *rbcL* region, showing shared patterns of polymorphism between groups of species. Dots indicate compliance with the consensus sequences.

Species	Type	66	180	263	300	372	378	492
<b>Consensus</b>		<b>C</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>
<i>H. perforatum</i>	1p	T	T	T	G	C	.	.
<i>H. maculatum</i>	1	.	T	.	G	C	.	.
<i>H. tetrapterum</i>	1	.	T	.	G	C	.	.
<i>H. undulatum</i>	1	.	T	.	G	C	.	.
<i>H. montanum</i>	1	.	T	.	G	C	.	.
<i>H. hirsutum</i>	1	.	.	.	G	C	.	.
<i>H. androsaemum</i>	2	.	.	.	.	.	G	T
<i>H. ascyron</i>	2	.	.	.	.	.	G	T
<i>H. calycinum</i>	2	.	.	.	.	.	G	T
<i>H. crux-andreae</i>	2	.	.	.	.	.	G	T
<i>H. japonicum</i>	2	.	.	.	.	.	G	T

#### 2.4.2. Haplotype Mapping of Reference Samples

Using the haplotype map defined from the panel, it was possible to assign the reference samples to haplotypes 1 or 2 (Table 4). Several samples showed an intermediate haplotype not observed in any of the panel accessions, with the three 5' SNPs matching haplotype 2 and the two 3' SNPs matching haplotype 1. These were designated as type 1–2.

**Table 4.** Haplotype map of the *rbcL* region, assigning reference samples to haplotype. Type 1–2 represents an intermediate haplotype between types 1 and 2. Dots indicate identity with the consensus base at that position.

Reference	Type	66	180	263	300	372	378	492
<b>Consensus</b>		<b>C</b>	<b>C</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>	<b>A</b>
<i>H. perforatum</i> 13876	1p	T	T	T	G	C	.	.
<i>H. delphicum</i> 13938	1	.	T	.	G	C	.	.
<i>H. montanum</i> A12F	1	.	T	.	G	C	G	.
<i>H. kouytchense</i> 13866	1–2	.	.	.	.	.	.	.
<i>H. patulum</i> 13908	1–2	.	.	.	.	.	.	.
<i>H. perforatum</i> 13921	1–2	.	.	.	.	.	.	.
<i>H. perforatum</i> 13932	1–2	.	.	.	.	.	.	.
<i>H. androsaemum</i> 13854	1–2	.	.	.	.	.	.	.
<i>H. ascyron</i> 13993	2	.	.	.	.	.	G	T
<i>H. athoum</i> 13923	2	.	.	.	.	.	G	T
<i>H. calycinum</i> 13929	2	.	.	.	.	.	G	T
<i>H. maculatum</i> 13896	2	.	.	.	.	.	G	T

Only one of the *H. perforatum* reference standards perfectly matched the 1p haplotype (*H. perforatum* 13876). The other two *H. perforatum* samples (13921 and 13932) had a type 1–2 haplotype with neither *perforatum*-related SNP present. Other anomalies included the *H. maculatum* reference with a type 2 rather than type 1 haplotype and the *H. androsaemum* reference with a type 1–2 rather than type 2 haplotype.

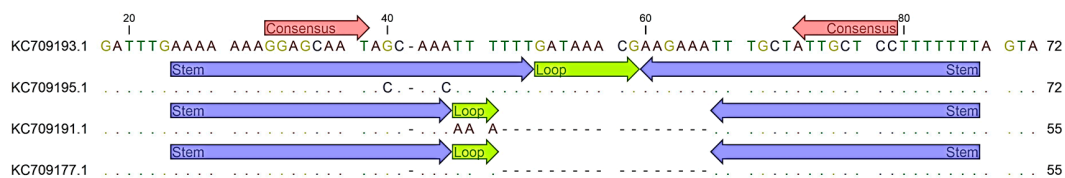
#### 2.5. The *matK* Barcode Region

As noted in Table 1, there are very few *matK* sequences deposited in either BOLD or GenBank. In this situation, it would be necessary to rely on the *matK* sequences obtained from reference samples in the laboratory. However, the *matK* barcode proved to be quite intractable with regard to amplification and sequencing. Several different published primer pairs were tested (see Materials and Methods) but none had good success rates and only four of the 12 DNA standards were able to be sequenced. There was better success with the test sample collection (50% success rate for sequences obtained) but without a range of panel species sequences for comparison, this was regarded as insufficient to proceed.

## 2.6. The *psbA-trnH* Barcode Region

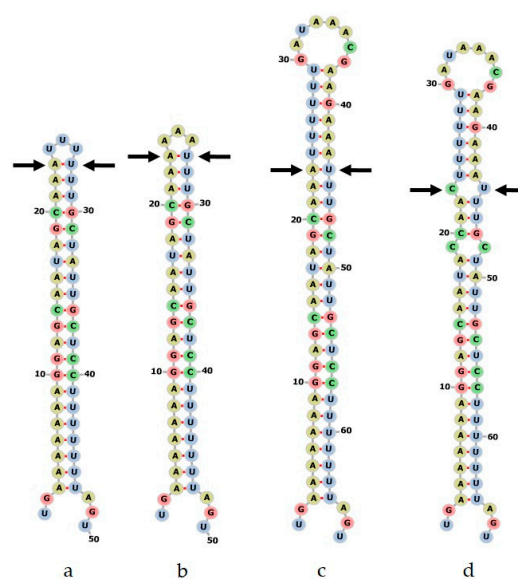
### 2.6.1. A Curated Database of *psbA-trnH* Sequences

There are only a few *Hypericum psbA-trnH* accessions in the BOLD database, but GenBank contains *psbA-trnH* sequences for a large number of *Hypericum* species from a large-scale taxonomic project [26]. Alignment of the panel sequences indicates some of the characteristic features of this region. The 3' intergenic spacer region is characterized by A/T rich sequences with many homopolymeric stretches. Individual accessions of the same species showed random long insertions, making alignment difficult and raising doubts about the reliability of sequences from single accessions representing an entire species. There were, however, some characteristic features of *H. perforatum* that could be used to discriminate this species from all the others. Apart from three SNPs, there were two consistent insertions and a large deletion towards the 5' end of the region. This deletion was found in all six of the complete *H. perforatum* accessions and in some but not all of the other Section *Hypericum* species. On closer inspection, this deletion was found to occur in the loop region of the characteristic stem-loop structure in the 3' UTR of the *psbA* gene [52,53]. Figure 3 shows the sequence at the stem-loop structure in four of the panel sequences, two *H. perforatum* variants plus *H. olympicum* and *H. hirsutum*. The location of the consensus sequence described in [52] is shown along with the full stem and loop sequences relative to the deletion.



**Figure 3.** Alignment of selected *Hypericum* panel *psbA* 3' UTR sequences. The predicted stem-loop RNA structures and consensus sequences [52] are shown. KC709193, *H. olympicum*; KC709195, *H. hirsutum*; KC709191, *H. perforatum* C47; KC709193, *H. perforatum* C22.

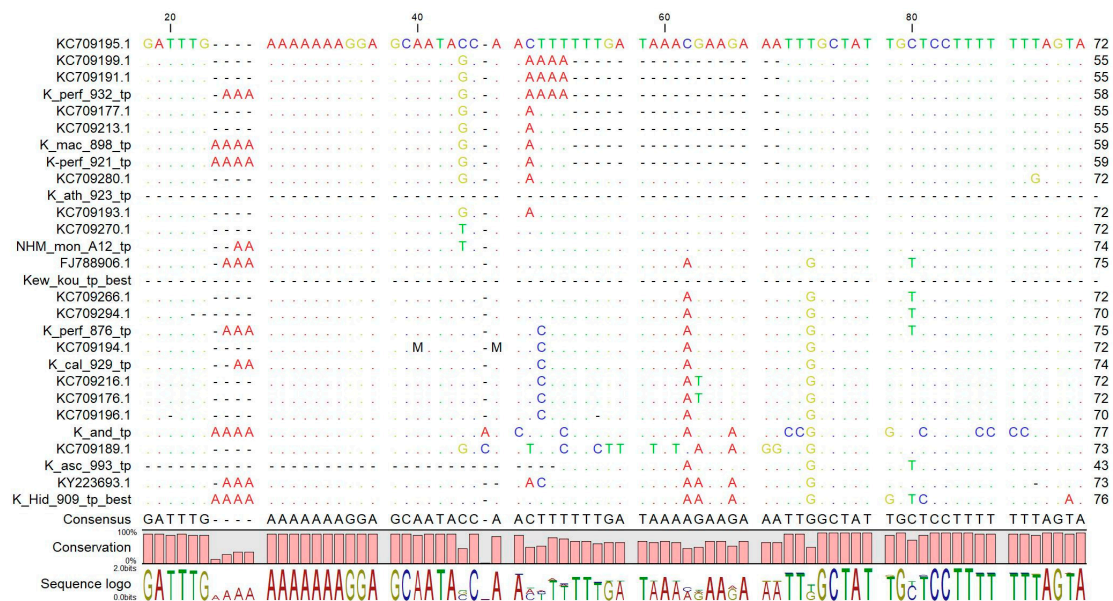
The corresponding secondary structures are shown in Figure 4. The two *H. perforatum* variants can be seen to differ only in the three base loop sequence. The sequences lacking the deletion have an extended stem, with some variation in sequences creating interior loops.



**Figure 4.** Predicted RNA secondary structures of *Hypericum psbA* 3' UTR sequences. a, *H. perforatum* C22; b, *H. perforatum* C47; c, *H. olympicum*; d, *H. hirsutum*. Arrows indicate the equivalent base pair in each structure.

## 2.6.2. Reference and Test *psbA-trnH* Sequences

The reference sample *psbA-trnH* sequences were compared with the database sequences (Figure 5). Most reference sequences matched the features observed in the corresponding panel sequence. Two *H. perforatum* reference samples (13921, 13932) had the characteristic truncated stem-loop (one with AAA in the loop, the other with UUU). In contrast, *H. perforatum* 13876 did not match this pattern and had an extended stem-loop structure. (Note—this is in contrast to the *rbcL* haplotypes, where 13876 had the expected haplotype 1p, but 13921 and 13932 were type 1–2).



**Figure 5.** Alignment of database and reference sample *psbA* 3' UTR RNA stem-loop region. Database sequences are labelled with GenBank accession numbers. Reference samples are labelled according to source (K = Kew; NHM = Natural History Museum), abbreviated species name (“perf” = *H. perforatum* etc.) and abbreviated voucher number as shown in Table S1.

## 2.7. Choosing the Optimal Barcode

Following the survey of suitable barcodes, it was clear that *matK* was not a suitable candidate barcode for two practical reasons: lack of database sequences and poor PCR amplification and sequence analysis. Of the remaining three, *rbcL* was shown to be a suitable discriminator at the genus and Section level, but the key challenge of differentiating between *H. perforatum* and *H. maculatum* was found to rest on just two SNPs. The *psbA-trnH* region showed a more suitable degree of variation between species, but evidence of frequent insertion/deletion events and homopolymer length heterogeneity within species would make the definition of a precise quality standard barcode difficult. The stem-loop region anchors a variable region between two conserved sequences, so could serve as a more useful indicator if this was shown to precisely map the species boundary.

The ITS region starts with the advantage of a large database of sequences and sufficient variation to discriminate *H. perforatum* from all the other members of the panel. The ability to extract ITS2 sequences from the barcode and perform secondary structure analysis provides further confidence in the differentiation of *H. perforatum* and *H. maculatum*. A standard ITS reference barcode for *H. perforatum* for a pharmacopoeial monograph could comprise the two regions, one capturing the two main variable regions of ITS1, and the other covering ITS2 (Figure 6). The lowercase letters define exact bases that must be matched, while the uppercase bases show the remaining sequence [54,55]. There is an overall requirement for a 95% match between a test sequence and the reference, as well as an exact match with defined bases.

```

>Hypericum_perforatum ITS1_reference_sequence
TCGAAACCTTCAAATGACCCGCGAACTAGTTATCCACAAGTCGGGGGTGTCGTGGGcttctgtccccgggcCCCCGtggcggTGGTGGCC
AGGCGGCGCCAAGctcttGGCACGGCTGGcccatcACCTGCCAACAAACAAACCCCGGCGCGGCACGCGCCAAGGAACCTTTGCATCATA
AGAAGgtaaGCTCCCGGCTGTGCCGAAATCGGACAACACGGTGGGGGCTCCTTCTGTTCATAACAATAA

>Hypericum_perforatum ITS2_reference_sequence
ACACATCGTCGCCCCCAAAATCCCGATATCTgcaagagacaatcggaataGGATGGGCGGAAAATGGTCTCCCGTGCCTCCCGTTC
GCGGTTGGCCCAAAATGAGTTCCTGGcaaagcaaAGCCACGACCAGCGGTGGTTGTAAGACCCTCGGTACAAGTCGTGAGCCTTGCATT
GCTCGTAGGGACATGTTGACCCTGAACGTGATCGAGTAACatcgaaacacTCACAAA

```

**Figure 6.** Proposed *H. perforatum* reference sequence from the ITS1 and ITS2 regions. Lower case bases (yellow highlight)—exact match required; Uppercase bases—exact match not required (95% identity required overall). Polymorphic positions in *H. perforatum* are shown with ambiguous single code letters (green highlight).

### 2.7.1. ITS Barcoding of Test Samples

Based on this choice, a collection of *H. perforatum* and *H. maculatum* specimens from Lithuania were tested against the ITS reference barcode [56]. The ITS regions of the test sample collection were amplified and sequenced with a success rate of just over 70%. Eleven of the fourteen *H. perforatum* labeled samples matched the reference ITS barcode, while the remaining three did not meet the criteria and would be rejected in a quality control situation. None of the 12 samples labeled as *H. maculatum* matched the *H. perforatum* ITS reference barcode and showed closest similarity to *H. maculatum* ITS barcode sequences. (Table 5).

**Table 5.** Test plant samples of *H. perforatum* and *H. maculatum* [56]. The sample number indicates the preliminary botanical identification of the specimen. The primary identification is by matching the ITS sequence to the *H. perforatum* reference standard. Supporting evidence is provided by the *psbA-trnH* stem-loop (L = long; S = short; Sa = short with AAA loop) and *rbcL* haplotype (1(p) shows one *H. perforatum*-related SNP, 1p shows two SNPs). The final assignment (ID) is coded as M = *H. maculatum*; m = possible *H. maculatum*; P = *H. perforatum*; p = possible *H. perforatum*; U = undetermined.

Sample No.	DNA No.	ITS Match to <i>H. perforatum</i>	<i>trnH-psbA</i> Stem	<i>rbcL</i> Type	ID
mac 01	007	No	L	1	M
mac 02	011	No	L	1(p)	M
mac 03	012	No	L	1p	M
mac 04	018	No	L	1p	M
mac 05	030		L	1p	m
mac 06	045	No	L	1-2	m
mac 07	001		L	1	m
mac 08	031	No	L	1p	M
mac 09	027	No	L	1	M
mac 10	022		L	1(p)	m
mac 11	014		S	1p	p
mac 12	034	No	L	1p	M
mac 13	025	No	L	1p	M
mac 14	019	No	L	1	M
mac 15	036	No	s	1p	m
mac 16	023	No	L	1	M
perf 01	005		S	1p	p
perf 02	043	Yes	S	1-2	P
perf 03	017	No	Sa	1	U
perf 04	026	Yes	S	1p	P
perf 05	016		S	1p	p
perf 06	038		L	1p	U
perf 07	044	Yes	S	1p	P
perf 08	035	Yes	S	1	P

Table 5. Cont.

Sample No.	DNA No.	ITS Match to <i>H. perforatum</i>	<i>trnH-psbA</i> Stem	<i>rbcL</i> Type	ID
perf 09	029	No	S	1p	U
perf 10	015		S	1p	p
perf 11	032	Yes	S	1	P
perf 12	041		S	1p	p
perf 13	028	No	Sa	1p	U
perf 14	024		S	1p	p
perf 15	013		L	1p	U
perf 16	037	Yes	S	1p	P
perf 17	039	Yes	S	1p	P
perf 18	020	Yes	Sa	1p	P
perf 19	042	Yes	Sa	1p	P
perf 20	040	Yes	Sa	1p	P
perf 21	033	Yes	S	1p	P
perf 22	021		L	1(p)	U

### 2.7.2. Supporting Evidence from *rbcL* Haplotype and *psbA-trnH* Stem-Loop

The failure to amplify and sequence all of the test sample ITS barcodes was one reason to seek supporting evidence from the *rbcL* and *psbA-trnH* barcodes. These were amplified and characterized for their *rbcL* haplotype (see Tables 3 and 4) and *psbA-trnH* stem-loop length (see Figures 3 and 4). These characters are recorded in Table 5. Nearly all of the “perf” samples had the 1p *rbcL* haplotype, but the majority of “mac” samples also showed this haplotype. The apparent specificity of the 1p haplotype to *H. perforatum* was called into question by these results. Two samples showed an anomalous 1–2 *rbcL* haplotype (mac06 and perf02) despite having ITS sequences consistent with *H. maculatum* and *H. perforatum* respectively.

The *psbA* stem-loop structure showed a better correlation with the botanical identification and ITS sequences of samples. Of 22 nominal *H. perforatum* samples, the identity of 12 was supported by the ITS and *psbA* barcodes. Another five without ITS sequences were designated as likely *H. perforatum*, as was one of the *H. maculatum*-labeled samples. The identity of the remaining six remained undetermined, largely due to the failure of the ITS barcoding. The identity of 10 of the 16 *H. maculatum* samples was confirmed by ITS + *psbA*, with another five being consistent with *H. maculatum* based on *psbA* stem length.

## 3. Discussion

### 3.1. Applied Barcoding for Herbals

The flowchart/decision tree shown in Figure 1 distills many years of experience designing DNA barcode quality control tests for the herbals industry [7] and in developing reference barcodes for pharmacopoeial monographs [54,55]. The flow effectively starts with the question—what exactly is the “target”? Is it a single species or are several species legitimately traded under one common name? Is there confusion in the trade about the correct species name(s) and is the taxonomy clear? Is adulteration a problem, what are the causes and are the adulterant species known?

This background knowledge lays the foundation for the next stage, which is to identify a suitable “adulterant panel”. In this case study, a panel of 20 *Hypericum* species was chosen. This is larger than would be required for many herbals and reflects the large size of the *Hypericum* genus and the fact that several different adulterant *Hypericum* species are known or suspected [31]. Twenty *Hypericum* species in the panel proved to be manageable in terms of the manual collection of database sequences, sequencing of reference samples and interpretation of multiple sequence alignments. Careful consideration of the composition of the panel should ensure not only that all likely adulterant

species can be differentiated from the target, but also that unknown adulterants will be highly unlikely to fortuitously match the *H. perforatum* reference barcode sequence.

For many herbals, there may just be one or a few problematic species to discriminate from the target plant. This obviously makes the design of the panel easier, though the process of choosing a suitable barcode may still depend upon the recognition of a small number of discriminatory polymorphisms if the adulterant is a close relative. There are situations where the known adulterant is not a close relative or may not even be congeneric [57]. If the aim is restricted to discriminating between a target and specific unrelated adulterants, finding a discriminating barcode should be straightforward, but it must be noted that the resulting reference barcode sequence will not necessarily be unique to the target species, nor indeed allow discrimination of any other adulterants.

### 3.2. Choosing a Barcode for Species Identification

The choice of a suitable authentication barcode will depend on several factors. At the DNA level, the key elements are the ability to discriminate between the target species and the adulterant panel, and the reliability of that discrimination. This discrimination will typically result from small regions of variation (mini- or microcodes [58–60]) that consistently map the species boundary between target and panel, rather than a calculation of overall genetic distances. Beyond the choice of the barcode sequence itself, the identification of variable minibarcode regions may also be an objective of the study in order to design authentication assays based on conventional, multiplex or qPCR primers [12–14]. For example, the qPCR primers designed to authenticate *H. perforatum* samples were designed using the same adulterant panel of *Hypericum* ITS sequences to target regions of essential bases (Figure 6) and tested against the same reference DNA collection [12]. The recent recommendations for validation of qualitative real-time PCR assays for diagnostic identification present a valuable opportunity to standardize such assays [61] but it should be recognized that such assays are a substitute for full-length DNA barcode identification, with the primers/probes acting as “minibarcode readers” [61]. The accuracy of such assays depends not only upon the discriminatory ability of the targeted minibarcode regions but also the specificity of design and performance of the primers/probes.

One factor that informs the choice of barcode is that the resultant identification assay should be fit for purpose and fulfill the needs of the industry and its regulators [7]. Chemical tests for herbal quality are typically based on simple HPTLC banding patterns rather than complex analytical profiles and our model for DNA testing has emulated this approach by picking out a small number of key informative SNPs. Indeed, the difference between a target and adulterant could be a single SNP if it is reliably known to be present in 100% of target samples and 0% of adulterant samples; this could still allow the design of a rapid high-resolution melt curve (HRM) authentication assay [62].

#### 3.2.1. Defining a Reference Barcode Sequence: Sensitivity and Specificity

As with any other diagnostic test, the reliability of a test based on matching a DNA barcode to a reference standard will depend upon the *specificity/sensitivity* of the reference barcode sequence and the *accuracy* and *precision* of the identification assay. The specificity of a reference standard barcode is measured by the proportion of false-positive identifications and is related to the number of characteristic features that discriminate between the target species and its adulterants. The sensitivity is a measure of the proportion of false negatives and reflects the extent to which intraspecific variation is captured within the reference sequence. The sensitivity and specificity of the reference sequence are therefore heavily dependent on the number and breadth of available target and adulterant sequences. The more barcode sequences collected from different laboratories and geographical locations, the more confidence can be placed in the barcode as a unique and consistent identifier of the target species. In this study, the ITS region had a starting advantage of having large numbers of *H. perforatum* sequences, a reasonable number from closest relative *H. maculatum* and a broad spread from other species (Table 1). This increased the confidence that the informative sites stipulated in the reference barcode were

representative and characteristic of the species as a whole, and also allowed unreliable sequences to be identified and ignored. The flowchart emphasizes the value of checking the sources of database sequences, particularly those on GenBank [20,21], and where possible matching sequences against conspecific accessions.

The large number of *H. perforatum* ITS sequences also allowed consistent sites of intraspecific variation to be identified (Figure S1) [12]. Four subspecies of *H. perforatum* are recognized by Robson (*perforatum*, *chinense*, *veronense* and *songaricum*) [29] but are rarely recorded in the published literature or database accessions [33]. Molecular phylogenetic studies have characterized two distinct gene pools in European *H. perforatum* populations, though their relationship to the subspecies is not clear [63]. This highlights the need to collect as many sequences and reference samples as possible in order to ensure that any intraspecific variation that might legitimately be found in commercial trade is captured and incorporated into the reference barcode sequence using ambiguity codes (Figure 6). There is also a requirement to curate reference sequences, as sequence data steadily accumulate and new target plant populations reach the market. The possibility of the barcode becoming outdated is guarded against by regular iterations of the process described in Figure 1, and by an active stakeholder community using the reference barcode in routine testing protocols and reporting back anomalous results. This feedback system is already in place for other pharmacopoeial test methods.

### 3.2.2. Matching a Reference Barcode Sequence: Accuracy and Precision

In a diagnostic test based on matching test sample barcode sequences to a reference sequence, the accuracy and precision of the process start with the “wet laboratory” procedures of DNA extraction, PCR amplification and DNA sequence analysis. A flow chart to ensure that good quality sequence data is obtained from test samples has been recommended, along with proposals for the optimal application of DNA testing in herbal drug supply chains as an upstream triage system to complement chemical testing [7]. The subsequent process of matching a test barcode sequence to a reference sequence determines the accuracy of the identification.

In the *Hypericum* examples described in this paper, the definition of reference barcodes and matching of samples has been largely conducted by manual inspection of pairwise or multiple alignments. In the case of *psbA-trnH*, the definition of a single characteristic feature (long/short 3′*psbA* RNA stem) and subsequent matching of test samples is straightforward. The *rbcl* 5-base core haplotypes with two possible *H. perforatum*-specific SNPs was also straightforward to recognize and read manually. The ITS barcode was more complex to analyse, both in terms of defining the reference sequence and then matching the test samples. Several polymorphic sites were recognized that discriminated between members of the Section *Hypericum* and other sections, but far fewer were found that differentiated *H. perforatum* from its closest relatives. There is only one SNP in ITS1 and three in ITS2 that consistently distinguish all of the three subtypes of *H. perforatum* from *H. maculatum*, compared to seven SNPs that characterize the three *H. perforatum* subtypes. This exemplifies the requirement for a diagnostic identification assay to specify essential discriminatory bases rather than a measure of overall genetic divergence.

This diagnostic approach to identification is analogous to a botanical key for identification of plants by matching distinguishing morphological features. A number of barcode identification algorithms based on diagnostic methods have been developed [64–68]. These recognize short sequence strings [68] (“diagnostic distinguishers” [67] or “distinguishing subsequences” [66]) in reference sequences which are then matched to query sequences. These are applicable to single barcode regions or entire genomes, and may be independent strings or located within their surrounding context [65]. Comparison of these diagnostic algorithms with pairwise distance [18,69], similarity, hierarchical clustering and phylogenetic tree-based methods indicates superior identification accuracy [64,65]. The reference sequence matching method described here is effectively a simplified diagnostic method in which the reference database is a single reference sequence and the diagnostic distinguishers are the prescribed essential bases. The requirement for overall similarity with the entire barcode sequence can then be

viewed as contextualizing the diagnostic strings rather than using the similarity threshold to identify the species directly.

### 3.3. Recommended Barcodes for *H. perforatum* Authentication

The choice of ITS as the primary barcode was not dictated solely by the number of available sequences. The other barcodes showed suboptimal characteristics that have been noted for many other plant groups. The *rbcL* barcode had insufficient variation to discriminate between species or even Sections of the genus. The three haplotypes noted were based on five SNPs. In combination, these appeared to differentiate three clades containing species from related Sections, but the expectation that *H. perforatum* and *H. maculatum* samples would always show a type 1 haplotype was undermined by anomalous results obtained with three of the reference samples (Kew 13896, 13921 and 13932) and two of the test samples (mac06 and perf02). The two apparent *H. perforatum*-specific SNPs proved to be even more unreliable when tested against the Lithuanian collection, with the majority of *H. maculatum* samples showing the *H. perforatum* type 1p haplotype.

The *matK* region was ruled out for pragmatic reasons, as it proved difficult to amplify and sequence with several different published primer pairs. This may explain the paucity of published sequences which also weighed against the suitability of adopting this barcode for *H. perforatum* authentication. The third accepted plastid barcode, *psbA-trnH*, also showed traits that have been reported in other studies of this type. The barcode showed sufficient variation to distinguish between the panel species and *H. perforatum*, but even in this small sample set, there were several examples of random insertions and deletions within single accessions when compared to conspecific sequences. This behavior has been noted elsewhere [69–71] and a number of authors have focused on the specific features of the *psbA* 3' UTR [52,53,71,72]. In this study, a shortening of the predicted RNA stem-loop structure in this region appeared to be characteristic of *H. perforatum* (though not unique to this species; Figures 3–5). The obvious deletion in the *H. perforatum psbA* 3'UTR sequence could form the basis of a rapid authentication assay if shown to be reliable. Although anomalous results were found with the reference samples, this feature could discriminate between the “perf” and “mac” samples in the test collection with reasonable consistency (14/16 *H. maculatum* samples showed the long stem character, while 20/22 *H. perforatum* samples showed the short stem character). Interestingly, three of the “perf” samples that did not match the *H. perforatum* ITS reference (perf 03, 09 and 13) still showed the *psbA* short stem character.

### 3.4. Current Trends in Applied Barcoding

The disparities between the ITS, *psbA* and *rbcL* barcodes in certain specimens may reflect the complex apomictic and sexual reproduction systems in *H. perforatum* [63,73–75], its frequent hybridization with related species (*H. maculatum*, *H. undulatum* and *H. tetrapterum*) and variable ploidy [30,73,74,76–78]. These are all factors likely to contribute to lower levels of success in species discrimination in DNA barcoding studies [18]. There is also the possibility of chloroplast capture leading to disparities between nuclear and plastid phylogenetic relationships [79–83]. One way to resolve this would be to sequence the entire plastid genomes of *H. perforatum* and related species with a view to discovering more effective genetic markers [84,85] or to use the entire plastid genome as a “super-barcode” [85,86].

Next-Generation Sequencing (NGS) technologies have been applied to herbal drug authentication using genome skimming or amplicon metabarcoding approaches [1,8,87–92], including *H. perforatum* commercial products [16]. The quality of a herbal product is determined by the correct identity of its ingredients and its purity, measured by % contamination by a range of inorganic and biological contaminants. As described in this paper, DNA barcoding provides a direct method to confirm target species identity and, less frequently, to detect specific toxic adulterants [57,93–96]. A major advantage of metabarcoding assays is that they provide a “What’s In My Pot?” (WIMP) analysis [97] which determines both identity and purity. A key issue for the development of these assays is



to meet the accuracy requirements of two quite discrete criteria: (i) the accuracy of the taxonomy prediction algorithms used in metabarcoding sequence analysis pipelines [98] and (ii) the quantitation of contamination by nontarget adulterants [6,99].

#### 4. Materials and Methods

##### 4.1. A panel of Potential Adulterant *Hypericum* Species

A panel of *Hypericum* species meeting some or all of the following criteria was selected:

1. Close relatives of *H. perforatum*;
2. Common in commercial trade as ornamental or medicinal plants;
3. Reported as adulterants of *H. perforatum*.

The selected panel comprized just twenty species (Table 6) and covers several close relatives from the Section *Hypericum* (*H. attenuatum*, *H. elegans*, *H. maculatum*, *H. tetrapterum*, *H. undulatum*), a number from the sections *Ascyreia* (*H. acmosepalum*, *H. calycinum*, *H. kouytchense*, *H. patulum*) and *Adenosepalum* (*H. athoum*, *H. delphicum*, *H. montanum*) and single representatives from the sections *Androsaemum* (*H. androsaemum*), *Drosocarpium* (*H. barbatum*), *Myriandra* (*H. crux-andreae*), *Oligostema* (*H. olympicum*), *Roscyna* (*H. ascyron*), *Taeniocarpium* (*H. hirsutum*) and *Trignobrathys* (*H. japonicum*) [12]. Table 6 indicates the criteria met for selection of each species as known or potential adulterants and/or close relatives.

**Table 6.** *Hypericum* species included in the restricted panel. The criteria for choosing these species are indicated in the columns: Rel, close relatives of *H. perforatum* in the Section *Hypericum*; MPNS, medicinal plants listed in the Medicinal Plants Names Service with at least one medicinal plant reference [32]; Cult, ornamental plants listed in the RHS horticultural database <http://apps.rhs.org.uk/horticulturaldatabase/index.asp>; Adult, plants identified as adulterants of commercial St John's Wort herbal products [31].

Species	Attribution	Section	Rel	MPNS	Cult	Adult
<i>H. acmosepalum</i>	N.Robson	Ascyreia	-	-	-	Y
<i>H. androsaemum</i>	L.	Androsaemum	-	Y	Y	Y
<i>H. ascyron</i>	L.	Roscyna	-	Y	-	Y
<i>H. athoum</i>	Boiss. & Orph	Adenosepalum	-	-	Y	-
<i>H. attenuatum</i>	Fisch. ex Choisy	Hypericum	Y	-	-	-
<i>H. barbatum</i>	Jacq.	Drosocarpium	-	-	-	Y
<i>H. calycinum</i>	L.	Ascyreia	-	Y	Y	-
<i>H. crux-andreae</i>	(L.) Crantz	Myriandra	-	Y	-	Y
<i>H. delphicum</i>	Boiss. & Heldr.	Adenosepalum	-	-	-	-
<i>H. elegans</i>	Stephan ex Willd.	Hypericum	Y	Y	-	-
<i>H. hirsutum</i>	L.	Taeniocarpium	-	-	-	Y
<i>H. japonicum</i>	Thunb.	Trignobrathys	-	Y	Y	-
<i>H. kouytchense</i>	H.Lev	Ascyreia	-	-	Y	-
<i>H. maculatum</i>	Crantz	Hypericum	Y	Y	-	Y
<i>H. montanum</i>	L.	Adenosepalum	-	-	-	Y
<i>H. olympicum</i>	L.	Oligostema	-	-	Y	-
<i>H. patulum</i>	Thunb.	Ascyreia	-	Y	Y	Y
<i>H. perforatum</i>	L.	Hypericum	-	Y	Y	-
<i>H. tetrapterum</i>	Fr.	Hypericum	Y	Y	Y	Y
<i>H. undulatum</i>	Schousb. ex Willd.	Hypericum	Y	-	-	Y

#### 4.2. Plant and DNA Materials

DNA reference samples were obtained from vouchered specimens available in The Royal Botanic Gardens, Kew, DNA Bank, <https://www.kew.org/data/dnaBank/>. A sample of *H. montanum* DNA was provided by Dr Mark Carine (The Natural History Museum, London). These are listed in Table S1. Dried leaf samples of *H. perforatum* and *H. maculatum* plants collected from the wild were used as the test collection (Table S2). Voucher specimens of each field accession are deposited in the Herbarium, the Institute of Botany/BILAS, Vilnius, Lithuania [56] and a map showing the location of their collection is shown as Figure S2.

#### 4.3. DNA Extraction and Amplification of DNA Barcode Regions

DNA extractions were carried out using the Qiagen DNeasy Plant Mini Kit according to the manufacturer's instructions, starting with 0.02 g dried leaf material.

PCR reactions consisted of Green GoTaq® Flexi Buffer (Promega, Madison, WI, USA; 1×), MgCl<sub>2</sub> (2.5 mM), GoTaq® DNA Polymerase (Promega; 1.25 Units), relevant primers (0.1 μM each), dNTPs (0.1 μM each), and template DNA (0.7–1 μg) made up to a final volume 50 μL with nuclease-free water in 0.2 mL polypropylene tubes (Starlab, Milton Keynes, UK). The Applied Biosystems GeneAmp PCR System 9700 thermal cycler (Applied Biosystems, Foster City, CA, USA) was used with differing programs (Table 7).

Reactions without template DNA were utilized as controls. PCR products were run on 3% (*w/v*) agarose, 0.5 × TBE gels with 2 μL SYBRsafe™ (Invitrogen, Carlsbad, CA, USA) DNA stain at 90 V for ~30 min and analyzed in a BioRad Illuminator with ChemiDocXRS Camera and Quantity One software [100].

The PCR primers and conditions used to amplify each barcode region are shown in Table 7.

**Table 7.** Primers and programs used for amplification of barcode regions.

Region	Primer Sequence	Program
ITS	ITS1- TCCGTAGGTGAACCTGCGG ITS4-TCCTCCGCTTATTGATATGC [101]	7 min at 95 °C initial denaturation step, 30 cycles consisting of 1 min at 95 °C, 30 s at 60 °C and 1 min at 72 °C, final extension period of 7 min at 72 °C.
trnH-psbA	trnHf_05-CGCGCATGGTGGATTCACAATCC psbA3_f-GTTATGCATGAACGTAATGCTC [102]	5 min at 95 °C initial denaturation step, 35 cycles consisting of 1 min at 95 °C, 30 s at touchdown temperature and 1 min at 72 °C, final extension period of 7 min at 72 °C. Touchdown temperature began at 58 °C, reduced by 1 °C per cycle until 48 °C, then continued at 48 °C for the remainder of the program.
rbcL	rbcLa_f -ATGTCACCACAAAACAGAAAC rbcLa_rev-GTAAAATCAAGTCCACCRCG [103] 390F-CGATCTATTCAATCAATATTTTC 1326R-TCTAGCACACGAAAGTCGAAGT 2.1-CCTATCCATCTGGAAATCTTAG 2.1a-ATCCATCTGGAAATCTTAGTTC X F- TAAATTTACGATCAATTCATTC	5 min 95 °C initial denaturation step, 35 cycles consisting of 30 s at 95 °C, 20 s at 52 °C and 50 s at 72 °C, with a final extension period of 5 min at 72 °C. Initial "touch-up" program, 5 min 94 °C initial denaturation step, 5 cycles consisting of 30 s at 94 °C, 40 s at 44 °C and 40 s at 72 °C, followed by 30 cycles consisting of 30 s at 94 °C, 40 s at 46 °C and 40 s at 72 °C, with a final extension period of 3 min at 72 °C.
matK	5-GTTCTAGCACAAGAAAGTCG 3.2-CTTCTCTGTAAAGAATTC 3F_KIM f-CGTACAGTACTTTGTGTTTACGAG 1R_KIM r -ACCCAGTCCATCTGGAAATCTTGGTTC [104–106]	The second amplification contained 2 μL of the initial PCR product diluted 1:200 as the DNA template. Second matK program: 5 min 94 °C initial denaturation step, 35 cycles consisting of 30 s at 95 °C, 20 s at 46 °C and 40 s at 72 °C, with a final extension period of 3 min at 72 °C.

#### 4.4. DNA Sequence Analysis of Barcode Amplicons

Samples were either sequenced "in-house" or sent to an external sequence provider (Macrogen Europe B.V., Amsterdam, The Netherlands)

For in-house sequencing, preliminary PCR reactions were purified using QuickStep™ 2 PCR Purification Kit (EdgeBio, San Jose, CA, USA) and the DNA quantified using a Qubit® Fluorometer and Quant-iT™ dsDNA BR Assay Kit (Invitrogen, Carlsbad, CA, USA).

Cycle sequencing reactions were conducted using the BigDye<sup>®</sup> Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems, Foster City, CA, USA). Reactions consisted of Ready Reaction Premix (2.5×; ABI), BigDye Sequencing Buffer (5×; ABI), sequencing primer (3.2 pM; VHBio, Gateshead UK or IDT, Leuven, Belgium) template PCR product (5–20 ng) and nuclease-free water. The sequencing program comprised; 1 min at 96 °C initial denaturation, 25 cycles consisting of 10 s at 96 °C, 5 s at 50 °C, 4 min at 60 °C. Extension products were purified using Performa<sup>®</sup> DTR Gel Filtration Cartridges (EdgeBio, Maryland, USA), 10 µL Hi Di formamide was added and the sample thoroughly vortexed.

Products were analyzed on the ABI Prism<sup>™</sup> 310 Genetic Analyzer (Applied Biosystems, Foster City, CA, USA), using a 47 cm capillary and Performance Optimised Polymer 6 (Applied Biosystems, Foster City, CA, USA). The run module used consisted of a 30 s injection at 2.0 kV, followed by electrophoresis running at 50 °C and 15 kV for 36 min. Sequence Analysis 5.2 (Applied Biosystems, Foster City, CA, USA) software was used to collect data, with Basecaller 310POP6, to create the output AB1 file.

#### 4.5. Computer Analysis of Barcode Regions

Contig assembly of sequencing traces was performed using the CLC Main Workbench (Qiagen, Hilden, Germany). At least three reads in forward and reverse directions were assembled and conflicts resolved by manual inspection of traces.

Multiple alignment of sequences was also conducted on the CLC platform using the “Slow (accurate)” settings.

ITS2 secondary structure was analyzed using the tools available on the ITS2 database <http://its2.bioapps.biozentrum.uni-wuerzburg.de/>. The alignment of ITS2 sequences plus secondary structures was further analyzed and converted to single letter code using the 4Sale tools <http://4sale.bioapps.biozentrum.uni-wuerzburg.de/>.

The *psbA-trnH* secondary structure predictions were created with the RNAfold tools available at <http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi> using the default minimum free energy (MEF) and partition function setting.

**Supplementary Materials:** The following are available online at <http://www.mdpi.com/2223-7747/9/9/1150/s1>, Figure S1: Multiple alignment of the ITS barcode region of *Hypericum* panel species. Table S1: List of reference DNA samples from DNA banks. Table S2: Collection of *H. perforatum* and *H. maculatum* plant leaf samples from Lithuania. Figure S1. Alignment of the adulteration panel of *Hypericum* ITS sequences.

**Author Contributions:** A.S. and C.H. conceived and designed the experiments; C.H. and C.L.-W. performed the experiments; A.S. and C.H. analyzed the data; A.S. and C.H. wrote the paper. All authors have read and agreed to the published version of the manuscript.

**Funding:** This research was funded by De Montfort University via discretionary HEIF3 funding and a PhD bursary awarded to C.H.

**Acknowledgments:** The authors would like to thank Asta Judzentiene, University of Vilnius for supplying the collection of Lithuanian *H. maculatum* and *H. perforatum* plant samples, Mark Carine, Natural History Museum, London and Monique Simmonds, Royal Botanic Gardens, Kew for *Hypericum* reference samples. The authors acknowledge the award of a De Montfort University PhD Bursary to C.H.

**Conflicts of Interest:** The authors declare no conflict of interest.

## References

1. Parveen, I.; Gafner, S.; Techen, N.; Murch, S.J.; Khan, I.A. DNA Barcoding for the identification of botanicals in herbal medicine and dietary supplements: Strengths and limitations. *Planta Medica* **2016**, *82*, 1225–1235. [[CrossRef](#)] [[PubMed](#)]
2. Techen, N.; Parveen, I.; Pan, Z.; Khan, I.A. DNA barcoding of medicinal plant material for identification. *Curr. Opin. Biotechnol.* **2014**, *25*, 103–110. [[CrossRef](#)] [[PubMed](#)]
3. Mishra, P.; Kumar, A.; Nagireddy, A.; Mani, D.N.; Shukla, A.K.; Tiwari, R.; Sundaresan, V. DNA barcoding: An efficient tool to overcome authentication challenges in the herbal market. *Plant Biotechnol. J.* **2016**, *14*, 8–21. [[CrossRef](#)] [[PubMed](#)]

4. Han, J.; Pang, X.; Liao, B.; Yao, H.; Song, J.; Chen, S. An authenticity survey of herbal medicines from markets in China using DNA barcoding. *Sci. Rep.* **2016**. [[CrossRef](#)] [[PubMed](#)]
5. Pawar, R.S.; Handy, S.M.; Cheng, R.; Shyong, N.; Grundel, E. Assessment of the authenticity of herbal dietary supplements: Comparison of chemical and DNA barcoding methods. *Planta Medica* **2017**, *83*. [[CrossRef](#)]
6. Raclariu, A.; Heinrich, M.; Ichim, M.; Boer, H. Benefits and Limitations of DNA Barcoding and Metabarcoding in Herbal Product Authentication. *Phytochem. Anal.* **2017**. [[CrossRef](#)]
7. Sgamma, T.; Lockie-Williams, C.; Kreuzer, M.; Williams, S.; Scheyhing, U.; Koch, E.; Slater, A.; Howard, C.; Mishra, P.; Kumar, A.; et al. DNA Barcoding for industrial quality assurance. *Planta Medica* **2017**, *14*, 8–21. [[CrossRef](#)]
8. Ivanova, N.V. Authentication of herbal supplements using next-generation sequencing. *PLoS ONE* **2016**, *11*, e0156426. [[CrossRef](#)] [[PubMed](#)]
9. Moraes, D.F.; Still, D.W.; Lum, M.R.; Hirsch, A.M. DNA-based authentication of botanicals and plant-derived dietary supplements: Where have we been and where are we going? *Planta Medica* **2015**, *81*, 687–695. [[CrossRef](#)]
10. Benson, D.A.; Cavanaugh, M.; Clark, K.; Karsch-Mizrachi, I.; Ostell, J.; Pruitt, K.D.; Sayers, E.W. GenBank. *Nucleic Acids Res.* **2018**, *46*, D41–D47. [[CrossRef](#)]
11. Crockett, S.L.; Douglas, A.W.; Scheffler, B.E.; Khan, I.A. Genetic profiling of *Hypericum* (St. John's wort) species by nuclear ribosomal ITS sequence analysis. *Planta Medica* **2004**, *70*, 929–935. [[CrossRef](#)] [[PubMed](#)]
12. Howard, C.; Hill, E.; Kreuzer, M.; Mali, P.; Masiero, E.; Slater, A.; Sgamma, T. DNA Authentication of St John's wort (*Hypericum perforatum* L.) commercial products targeting the ITS region. *Genes* **2019**, *10*, 286. [[CrossRef](#)] [[PubMed](#)]
13. Howard, C.; Bremner, P.D.; Fowler, M.R.; Isodo, B.; Scott, N.W.; Slater, A. Molecular identification of *Hypericum perforatum* by PCR amplification of the ITS and 5.8S rDNA region. *Planta Medica* **2009**, *75*, 864–869. [[CrossRef](#)] [[PubMed](#)]
14. Howard, C.; Socratous, E.; Williams, S.; Graham, E.; Fowler, M.R.; Scott, N.W.; Bremner, P.D.; Slater, A. PlantID - DNA-based identification of multiple medicinal plants in complex mixtures. *Chin. Med.* **2012**, *7*, 18. [[CrossRef](#)] [[PubMed](#)]
15. Costa, J.; Campos, B.; Amaral, J.S.; Nunes, E.M.; Oliveira, B.M.; Mafra, I. HRM analysis targeting ITS1 and matK loci as potential DNA mini-barcodes for the authentication of *Hypericum perforatum* and *Hypericum androsaemum* in herbal infusions. *Food Control* **2016**, *61*, 105–114. [[CrossRef](#)]
16. Raclariu, A.C.; Paltinean, R.; Vlase, L.; Labarre, A.; Manzanilla, V.; Ichim, M.C.; Crisan, G.; Brysting, A.K.; De Boer, H. Comparative authentication of *Hypericum perforatum* herbal products using DNA metabarcoding, TLC and HPLC-MS. *Sci. Rep.* **2017**, *7*, 1291. [[CrossRef](#)]
17. Chase, M.; Cowan, R.; Hollingsworth, P. A proposal for a standardised protocol to barcode all land plants. *Taxon* **2007**, *56*, 295–299. [[CrossRef](#)]
18. Hollingsworth, P.M.; Graham, S.W.; Little, D.P. Choosing and using a plant DNA barcode. *PLoS ONE* **2011**. [[CrossRef](#)]
19. Hollingsworth, M.L.; Clark, A.; Forrest, L.L.; Richardson, J.; Pennington, R.; Long, D.G.; Cowan, R.; Chase, M.W.; Gaudeul, M.; Hollingsworth, P.M. Selecting barcoding loci for plants: Evaluation of seven candidate loci with species level sampling in three divergent groups of land plants. *Mol. Ecol. Resour.* **2009**, *9*. [[CrossRef](#)]
20. Banchi, E.; Ametrano, C.G.; Greco, S.; Stanković, D.; Muggia, L.; Pallavicini, A. PLANiTS: A curated sequence reference dataset for plant ITS DNA metabarcoding. *Database* **2020**, *2020*. [[CrossRef](#)]
21. Pentinsaari, M.; Ratnasingham, S.; Miller, S.E.; Hebert, P.D.N. BOLD and GenBank revisited—Do identification errors arise in the lab or in the sequence libraries? *PLoS ONE* **2020**, *15*, e0231814. [[CrossRef](#)] [[PubMed](#)]
22. Carine, M.A.; Christenhusz, M.J.M. About this volume: The monograph of *Hypericum* by Norman Robson. *Phytotaxa* **2013**, *4*, 1–4. [[CrossRef](#)]
23. Nürk, N.M.; Crockett, S.L. Morphological and Phytochemical Diversity among *Hypericum* Species of the Mediterranean Basin. *Med. Aromat. Plant Sci. Biotechnol.* **2011**, *5*, 14–28. [[PubMed](#)]
24. Crockett, S.L.; Robson, N.K. Taxonomy and chemotaxonomy of the Genus *hypericum*. *Med. Aromat. Plant Sci. Biotechnol.* **2011**, *5*, 1–13.

25. Nürk, N.M.; Madriñán, S.; Carine, M.A.; Chase, M.W.; Blattner, F.R. Molecular phylogenetics and morphological evolution of St. John's wort (*Hypericum*; *Hypericaceae*). *Mol. Phylog. Evol.* **2013**, *66*, 1–16. [[CrossRef](#)]
26. Meseguer, A.; Aldasoro, J.; Sanmartín, I. Bayesian inference of phylogeny, morphology and range evolution reveals a complex evolutionary history in St. John's wort (*Hypericum*). *Mol. Phylog. Evol.* **2013**, *67*, 379–403. [[CrossRef](#)]
27. Pilepić, K.H.; Balić, M.; Blažina, N. Estimation of phylogenetic relationships among some *Hypericum* (*Hypericaceae*) species using internal transcribed spacer sequences. *Plant Biosyst. Int. J. Deal. Asp. Plant Biol.* **2011**, *145*, 81–87. [[CrossRef](#)]
28. ROBSON, N.K.B. And then came molecular phylogenetics—Reactions to a monographic study of *Hypericum* (*Hypericaceae*). *Phytotaxa* **2016**, *255*, 181–198. [[CrossRef](#)]
29. Robson, N. Studies in the genus *Hypericum* L. (*Guttiferae*) 4(2). Section 9. *Hypericum* sensu lato (part 2): Subsection 1. *Hypericum* series 1. *Hypericum*. *Bull. Nat. Hist. Mus. Bot. Ser.* **2002**, *32*. [[CrossRef](#)]
30. Brutovská, R.; Čellárová, E.; Schubert, I. Cytogenetic characterization of three *Hypericum* species by in situ hybridization. *Theor. Appl. Genet.* **2000**, *101*, 46–50. [[CrossRef](#)]
31. McCutcheon, A. Adulteration of *Hypericum perforatum*. *Bot. Adulterants Bull.* **2017**, 1–9.
32. Dauncey, E.; Edwards, S. Kew's Medicinal Plant Names Services (MPNS) enable effective information retrieval and communication. *Planta Medica* **2015**, *81*. [[CrossRef](#)]
33. Dauncey, E.; Irving, J.; Allkin, R. A review of issues of nomenclature and taxonomy of *Hypericum perforatum* L. and Kew's Medicinal Plant Names Services. *J. Pharm. Pharmacol.* **2019**, *71*, 4–14. [[CrossRef](#)]
34. Mitra, S.; Kannan, R. A note on unintentional adulterations in *Ayurvedic* herbs. *Ethnobot. Leaflets* **2007**, *11*, 11–15.
35. Huck-Pezzei, V.; Bittner, L.; Pallua, J.; Sonderegger, H.; Abel, G.; Popp, M.; Bonn, G.; Huck, C. A chromatographic and spectroscopic analytical platform for the characterization of St John's wort extract adulterations. *Anal. Methods* **2012**, *5*. [[CrossRef](#)]
36. Frommenwiler, D.A.A.; Reich, E.; Sudberg, S.; Sharaf, M.H.; Bzhelyansky, A.; Lucas, B. St. John's wort versus counterfeit St. John's wort: An HPTLC study. *J. AOAC Int.* **2016**, *99*, 1204–1212. [[CrossRef](#)]
37. Ratnasingham, S.; Hebert, P.D. Bold: The barcode of life data system (<http://www.barcodinglife.org>). *Mol. Ecol. Notes* **2007**. [[CrossRef](#)]
38. Hodkinson, T.R.; Waldren, S.; Parnell, J.A.N.; Kelleher, C.T.; Salamin, K.; Salamin, N. DNA banking for plant breeding, biotechnology and biodiversity evaluation. *J. Plant Res.* **2007**, *120*, 17–29. [[CrossRef](#)]
39. Gemeinholzer, B.; Dröge, G.; Zetzsche, H.; Haszprunar, G.; Klenk, H.-P.; Güntsch, A.; Berendsohn, W.G.; Wägele, J.-W. The DNA bank network: The start from a german initiative. *Biopreserv. Biobank* **2011**, *9*, 51–55. [[CrossRef](#)]
40. Droege, G.; Barker, K.; Astrin, J.J.; Bartels, P.; Butler, C.; Cantrill, D.; Coddington, J.; Forest, F.; Gemeinholzer, B.; Hobern, D.; et al. The Global Genome Biodiversity Network (GGBN) data portal. *Nucleic Acids Res.* **2014**, *42*, D607–D612. [[CrossRef](#)]
41. Kazi, T.; Hussain, N.; Bremner, P.; Slater, A.; Howard, C. The application of a DNA-based identification technique to over-the-counter herbal medicines. *Fitoterapia* **2013**, *87*, 27–30. [[CrossRef](#)] [[PubMed](#)]
42. Chen, S.; Pang, X.; Song, J.; Shi, L.; Yao, H.; Han, J.; Leon, C. A renaissance in herbal medicine identification: From morphology to DNA. *Biotechnol. Adv.* **2014**, *32*, 1237–1244. [[CrossRef](#)] [[PubMed](#)]
43. Chen, S.; Yao, H.; Han, J.; Liu, C.; Song, J.; Shi, L.; Zhu, Y.; Ma, X.; Gao, T.; Pang, X.; et al. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE* **2010**, *5*, e8613. [[CrossRef](#)] [[PubMed](#)]
44. Yao, H.; Song, J.; Liu, C.; Luo, K.; Han, J.; Li, Y.; Pang, X.; Xu, H.; Zhu, Y.; Xiao, P.; et al. Use of ITS2 region as the universal DNA barcode for plants and animals. *PLoS ONE* **2010**. [[CrossRef](#)] [[PubMed](#)]
45. Zhang, W.; Yuan, Y.; Yang, S.; Huang, J.; Huang, L. ITS2 secondary structure improves discrimination between medicinal “Mu Tong” species when using DNA barcoding. *PLoS ONE* **2015**, *10*, e0131185. [[CrossRef](#)] [[PubMed](#)]
46. Koetschan, C.; Hackl, T.; Müller, T.; Wolf, M.; Förster, F.; Schultz, J. ITS2 Database IV: Interactive taxon sampling for internal transcribed spacer 2 based phylogenies. *Mol. Phylogenet. Evol.* **2012**. [[CrossRef](#)]

47. Wolf, M.; Koetschan, C.; Müller, T. ITS2, 18S, 16S or any other RNA—Simply aligning sequences and their individual secondary structures simultaneously by an automatic approach. *Gene* **2014**, *546*, 145–149. [[CrossRef](#)]
48. Schultz, J.; Wolf, M. ITS2 sequence–structure analysis in phylogenetics: A how-to manual for molecular systematics. *Mol. Phylogenet. Evol.* **2009**. [[CrossRef](#)]
49. Müller, T.; Philippi, N.; Dandekar, T.; Schultz, J.; Wolf, M. Distinguishing species. *RNA* **2007**, *13*, 1469–1472. [[CrossRef](#)]
50. Wolf, M.; Chen, S.; Song, J.; Ankenbrand, M.; Müller, T. Compensatory base changes in ITS2 secondary structures correlate with the biological species concept despite intragenomic variability in ITS2 sequences—A proof of concept. *PLoS ONE* **2013**, *8*, e66726. [[CrossRef](#)]
51. Seibel, P.N.; Müller, T.; Dandekar, T.; Wolf, M. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. *BMC Res. Notes* **2008**, *1*, 91. [[CrossRef](#)] [[PubMed](#)]
52. Štorchová, H.; Olson, M. The architecture of the chloroplast psbA-trnH non-coding region in angiosperms. *Plant Syst. Evol.* **2007**, *268*, 235–256. [[CrossRef](#)]
53. Degtjareva, G.; Logacheva, M.; Samigullin, T.; Terentieva, E.; Valiejo-Roman, C. Organization of chloroplast psbA-trnH intergenic spacer in dicotyledonous angiosperms of the family umbelliferae. *Biochemistry (Moscow)* **2012**, *77*, 1056–1064. [[CrossRef](#)]
54. Deoxyribonucleic Acid (DNA) Based identification techniques for herbal drugs. In *British Pharmacopoeia*; British Pharmacopoeia Commission: London, UK, 2017; Appendix XI V.
55. DNA barcoding as a tool for botanical identification. In *British Pharmacopoeia*; British Pharmacopoeia Supplementary Chapter SC VII D; British Pharmacopoeia Commission: London, UK, 2017.
56. Radusiene, J.; Judzentiene, A.; Bernotiene, G. Essential oil composition and variability of *Hypericum perforatum* L. growing in Lithuania. *Biochem. Syst. Ecol.* **2005**, *33*, 113–124. [[CrossRef](#)]
57. Sgamma, T.; Masiero, E.; Mali, P.; Mahat, M.; Slater, A. Sequence-specific detection of aristolochia DNA—A simple test for contamination of herbal products. *Front. Plant Sci.* **2018**, *9*. [[CrossRef](#)] [[PubMed](#)]
58. Little, D. Authentication of Ginkgo biloba herbal dietary supplements using DNA barcoding. *Genome* **2014**. [[CrossRef](#)]
59. Hajibabaei, M.; McKenna, C. DNA mini-barcodes. In *Methods in Molecular Biology*; Humana Press: Clifton, NJ, USA, 2012; pp. 339–353. [[CrossRef](#)]
60. Summerbell, R.; Lévesque, C.; Seifert, K.; Bovers, M.; Fell, J.; Diaz, M.; Boekhout, T.; de Hoog, G.; Stalpers, J.; Crous, P. Microcoding: The second step in DNA barcoding. *Phil. Transact. R. Soc. London Ser. B Biol. Sci.* **2005**, *360*, 1897–1903. [[CrossRef](#)]
61. Newmaster, S.G.; Shanmughanandhan, D.; Kesanakurti, P.; Shehata, H.; Faller, A.; Noce, I.; Lee, J.; Rudzinski, P.; Lu, Z.; Zhang, Y.; et al. Recommendations for validation of real-time PCR methods for molecular diagnostic identification of botanicals. *J. AOAC Int.* **2019**. [[CrossRef](#)]
62. Masiero, E.; Banik, D.; Abson, J.; Greene, P.; Slater, A.; Sgamma, T. Genus-specific real-time PCR and HRM assays to distinguish liriopoe from ophiopogon samples. *Plants* **2017**, *6*, 53. [[CrossRef](#)]
63. Koch, M.A.; Scheriau, C.; Betzin, A.; Hohmann, N.; Sharbel, T.F. Evolution of cryptic gene pools in *Hypericum perforatum*: The influence of reproductive system and gene flow. *Ann. Bot.* **2013**, *111*, 1083–1094. [[CrossRef](#)]
64. Little, D.P.; Stevenson, D. A comparison of algorithms for the identification of specimens using DNA barcodes: Examples from gymnosperms. *Cladistics* **2007**, *23*. [[CrossRef](#)]
65. Little, D.P. DNA Barcode sequence identification incorporating taxonomic hierarchy and within taxon variability. *PLoS ONE* **2011**, *6*, e20552. [[CrossRef](#)]
66. Gibbs, M.J.; Armstrong, J.S.; Gibbs, A.J. Individual sequences in large sets of gene sequences may be distinguished efficiently by combinations of shared sub-sequences. *BMC Bioinform.* **2005**, *6*, 90. [[CrossRef](#)] [[PubMed](#)]
67. DasGupta, B.; Konwar, K.; Mandoiu, I.; Shvartsman, A. DNA-BAR: Distinguisher selection for DNA barcoding. *Bioinformatics* **2005**, *21*, 3424–3426. [[CrossRef](#)] [[PubMed](#)]
68. Rash, S.; Gusfield, D. String barcoding: Uncovering optimal virus signatures. In Proceedings of the 6th Annual International Conference on Computational Biology, Washington, DC, USA, 18–21 April 2002; pp. 254–261. [[CrossRef](#)]

69. Liu, C.; Liang, D.; Gao, T.; Pang, X.; Song, J.; Yao, H.; Han, J.; Liu, Z.; Guan, X.; Jiang, K.; et al. PTIGS-IdIt, a system for species identification by DNA sequences of the psbA-trnH intergenic spacer region. *BMC Bioinform.* **2011**, *12* (Suppl. 13), S4. [[CrossRef](#)] [[PubMed](#)]
70. Whitlock, A.B.; Hale, A.M.; Groff, P.A. Intraspecific inversions pose a challenge for the trnH-psbA plant DNA barcode. *PLoS ONE* **2010**. [[CrossRef](#)] [[PubMed](#)]
71. Hernández-Ledesma, P.; Bárcenas, R.T. Phylogenetic utility of the trnH-psbA IGR and stem-loop diversity of the 3' UTR in *Cactaceae* (*Caryophyllales*). *Plant Syst. Evol.* **2017**, *303*, 299–315. [[CrossRef](#)]
72. Bain, J.F.; Jansen, R.K. A chloroplast DNA hairpin structure provides useful phylogenetic data within tribe Senecioneae (Asteraceae). *Can. J. Bot.* **2006**, *84*, 862–868. [[CrossRef](#)]
73. Molins, M.; Corral, J.M.; Aliyu, O.; Koch, M.A.; Betzin, A.; Maron, J.L.; Sharbel, T.F. Biogeographic variation in genetic variability, apomixis expression and ploidy of St. John's wort (*Hypericum perforatum*) across its native and introduced range. *Ann. Bot.* **2014**, *113*, 417–427. [[CrossRef](#)]
74. Barcaccia, G.; Arzenton, F.; Sharbel, T.; Varotto, S.; Parrini, P.; Lucchin, M. Genetic diversity and reproductive biology in ecotypes of the facultative apomict *Hypericum perforatum* L. *Heredity* **2006**, *96*, 322–334. [[CrossRef](#)]
75. Murch, S.J.; Saxena, P.K. St. John's wort (*Hypericum perforatum* L.): Challenges and strategies for production of chemically-consistent plants. *Can J. Plant Sci.* **2006**, *86*, 765–771. [[CrossRef](#)]
76. Pank, F.; Matzk, F.; Kästner, U.; Blüthner, W.; de Garcia, F.E.; Meister, A.; Ryschka, U.; Schumann, G. Reproductive diversity and strategies for breeding in St. John's wort (*Hypericum perforatum* L.). *Euphytica* **2003**, *134*, 77–84. [[CrossRef](#)]
77. Matzk, F.; Hammer, K.; Schubert, I. Coevolution of apomixis and genome size within the genus *Hypericum*. *Sex. Plant Reprod.* **2003**, *16*, 51–58. [[CrossRef](#)]
78. Galla, G.; Barcaccia, G.; Schallau, A.; Molins, P.M.; Bäumllein, H.; Sharbel, T. The cytohistological basis of apospory in *Hypericum perforatum* L. *Sex. Plant Reprod.* **2011**, *24*, 47–61. [[CrossRef](#)]
79. Tsitrone, A.; Kirkpatrick, M.; Levin, D.A. A model for chloroplast capture. *Evolution* **2003**, *57*, 1776–1782. [[CrossRef](#)] [[PubMed](#)]
80. Mariac, C.; Scarcelli, N.; Pouzadou, J.; Barnaud, A.; Billot, C.; Faye, A.; Kougbéadjou, A.; Maillol, V.; Martin, G.; Sabot, F.; et al. Cost-effective enrichment hybridization capture of chloroplast genomes at deep multiplexing levels for population genetics and phylogeography studies. *Mol. Ecol. Resour.* **2014**. [[CrossRef](#)]
81. Stegemann, S.; Keuthe, M.; Greiner, S.; Bock, R. Horizontal transfer of chloroplast genomes between plant species. *Proc Natl. Acad. Sci USA* **2012**, *109*, 2434–2438. [[CrossRef](#)]
82. Liu, B.-B.; Campbell, C.S.; Hong, D.-Y.; Wen, J. Phylogenetic relationships and chloroplast capture in the *Amelanchier-Malacomeles-Peraphyllum* clade (Maleae, Rosaceae): Evidence from chloroplast genome and nuclear ribosomal DNA data using genome skimming. *Mol. Phylogenet. Evol.* **2020**, *147*, 106784. [[CrossRef](#)]
83. Milne, R.I.; Davies, C.; Prickett, R.; Inns, L.H.; Chamberlain, D.F. Phylogeny of *Rhododendron* subgenus *Hymenanthes* based on chloroplast DNA markers: Between-lineage hybridisation during adaptive radiation? *Plant Syst. Evol.* **2010**, *285*. [[CrossRef](#)]
84. Kreuzer, M.; Howard, C.; Adhikari, B.; Pendry, C.A.; Hawkins, J.A. Phylogenomic approaches to DNA barcoding of herbal medicines: Developing clade-specific diagnostic characters for berberis. *Front. Plant Sci.* **2019**, *10*. [[CrossRef](#)]
85. Vaughn, J.N.; Chaluvadi, S.R.; Tushar; Rangan, L.; Bennetzen, J.L. Whole plastome sequences from five ginger species facilitate marker development and define limits to barcode methodology. *PLoS ONE* **2014**. [[CrossRef](#)] [[PubMed](#)]
86. Li, X.; Yang, Y.; Henry, R.J.; Rossetto, M.; Wang, Y.; Chen, S. Plant DNA barcoding: From gene to genome. *Biol. Rev.* **2014**. [[CrossRef](#)] [[PubMed](#)]
87. Zhokhova, E.; Rodionov, A.; Povydysh, M.; Goncharov, Y.M.; Protasova, Y.A.; Yakovlev, G. Current state and prospects of DNA barcoding and DNA fingerprinting in the analysis of the quality of plant raw materials and plant-derived drugs. *Biol. Bull. Rev.* **2019**, *9*, 301–314. [[CrossRef](#)]
88. Seethapathy, G.; Raclariu-Manolica, A.-C.; Anmarkrud, J.; Wangensteen, H.; de Boer, H.J. DNA metabarcoding authentication of ayurvedic herbal products on the European market Raises concerns of quality and fidelity. *Front. Plant Sci.* **2019**, *10*, 68. [[CrossRef](#)] [[PubMed](#)]
89. Raclariu, A.; Tebrencu, C.; Ichim, M.; Ciupercă, O.; Brysting, A.; de Boer, H. What's in the box? Authentication of Echinacea herbal products using DNA metabarcoding and HPTLC. *Phytomedicine* **2018**. [[CrossRef](#)]

90. Omelchenko, D.O.; Speranskaya, A.S.; Ayginin, A.A.; Khafizov, K.; Krinitsina, A.A.; Fedotova, A.V.; Pozdyshev, D.V.; Shtratnikova, V.Y.; Kupriyanova, E.V.; Shipulin, G.A.; et al. Improved protocols of ITS1-based metabarcoding and their application in the analysis of plant-containing products. *Genes* **2019**, *10*, 122. [[CrossRef](#)]
91. Hoban, C.L.; Musgrave, I.F.; Coghlan, M.L.; Power, M.W.P.; Byard, R.W.; Nash, C.; Farrington, R.; Maker, G.; Crighton, E.; Trengove, R.; et al. Adulterants and contaminants in psychotropic herbal medicines detected with mass spectrometry and next-generation DNA sequencing. *Pharm. Med.* **2018**, *32*, 429–444. [[CrossRef](#)]
92. Coghlan, M.L.; Maker, G.; Crighton, E.; Haile, J.; Murray, D.C.; White, N.E.; Byard, R.W.; Bellgard, M.I.; Mullaney, I.; Trengove, R.; et al. Combined DNA, toxicological and heavy metal analyses provides an auditing toolkit to improve pharmacovigilance of traditional Chinese medicine (TCM). *Sci. Rep.* **2015**, *5*, 17475. [[CrossRef](#)]
93. Yang, P.; Li, X.; Zhou, H.; Hu, H.; Zhang, H.; Sun, W.; Wang, Y.; Yao, H. Molecular identification of chinese materia medica and its adulterants using its2 and psba-trnh barcodes: A case study on rhizoma menispermi. *Chin. Med.* **2014**, *5*. [[CrossRef](#)]
94. Li, M.; Au, K.-Y.; Lam, H.; Cheng, L.; But, P.; Shaw, P.-C. Molecular identification and cytotoxicity study of herbal medicinal materials that are confused by *Aristolochia* herbs. *Food Chem.* **2014**, *147*, 332–339. [[CrossRef](#)]
95. Wu, L.; Wang, B.; Zhao, M.; Liu, W.; Zhang, P.; Shi, Y.; Xiong, C.; Wang, P.; Sun, W.; Chen, S. Rapid Identification of officinal akebiae caulis and its toxic adulterant *Aristolochiae manshuriensis* caulis (*Aristolochia manshuriensis*) By loop-mediated isothermal amplification. *Front. Plant Sci.* **2016**, *7*, 887. [[CrossRef](#)] [[PubMed](#)]
96. Wu, L.; Sun, W.; Wang, B.; Zhao, H.; Li, Y.; Cai, S.; Xiang, L.; Zhu, Y.; Yao, H.; Song, J.; et al. An integrated system for identifying the hidden assassins in traditional medicines containing aristolochic acids. *Sci. Rep.* **2015**, *5*, 11318. [[CrossRef](#)] [[PubMed](#)]
97. Juul, S.; Izquierdo, F.; Hurst, A.; Dai, X.; Wright, A.; Kulesha, E.; Pettett, R.; Turner, D.J. What's in my pot? Real-time species identification on the MinIONTM. *BioRxiv* **2015**. [[CrossRef](#)]
98. Edgar, R.C. Accuracy of taxonomy prediction for 16S rRNA and fungal ITS sequences. *PeerJ* **2018**, *6*, e4652. [[CrossRef](#)] [[PubMed](#)]
99. Staats, M.; Arulandhu, A.J.; Gravendeel, B.; Holst-Jensen, A.; Scholtens, I.; Peelen, T.; Prins, T.W.; Kok, E. Advances in DNA metabarcoding for food and wildlife forensic species identification. *Anal. Bioanal. Chem.* **2016**, *408*, 4615–4630. [[CrossRef](#)] [[PubMed](#)]
100. Howard, C. The Development of Deoxyribonucleic Acid (DNA) Based Methods for the Identification and Authentication of Medicinal Plant Material. Ph.D. Thesis, De Montfort University, Leicester, UK, 2010.
101. White, T.J.; Bruns, T.; Lee, S.; Taylor, J.W.; Innis, M.A.; Gelfand, D.H.; Sninsky, J.J.; White, T.J. *Amplification and Direct Sequencing of Fungal Ribosomal Rna Genes for Phylogenetics*; Academic Press Inc.: Cambridge, MA, USA, 1990.
102. Kress, J.W.; Wurdack, K.J.; Zimmer, E.A.; Weigt, L.A.; Janzen, D.H. Use of DNA barcodes to identify flowering plants. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 8369–8374. [[CrossRef](#)]
103. Lahaye, R.; van der Bank, M.; Bogarin, D.; Warner, J.; Pupulin, F.; Gigot, G.; Maurin, O.; Duthoit, S.; Barraclough, T.G.; Savolainen, V. DNA barcoding the floras of biodiversity hotspots. *Proc. Natl. Acad. Sci. USA* **2008**, *105*, 2923–2928. [[CrossRef](#)]
104. Kim, W.; Ji, Y.; Choi, G.; Kang, Y.; Yang, S.; Moon, B. Molecular identification and phylogenetic analysis of important medicinal plant species in genus *Paeonia* based on rDNA-ITS, matK, and rbcL DNA barcode sequences. *Genet. Mol. Res. GMR* **2016**, *15*. [[CrossRef](#)]
105. Fazekas, A.J.; Burgess, K.S.; Kesanakurti, P.R.; Graham, S.W.; Newmaster, S.G.; Husband, B.C.; Percy, D.M.; Hajibabaei, M.; Barrett, S.C. Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. *PLoS ONE* **2008**, *3*, e2802. [[CrossRef](#)]
106. Chatrou, L.; Grayer, R.; Chase, M. Molecular phylogenetics of Caryophyllales based on nuclear 18S rDNA and plastid rbcL, atpB, and matK DNA sequences. *Am. J. Bot.* **2002**, *89*, 132–144. [[CrossRef](#)]

