Use of *rbc*L sequences for DNA barcoding and authentication of plant drugs used in Traditional Chinese Medicine

Florian Herrmann, Michael Wink*

Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany

Abstract:

Traditional Chinese medicine has become increasingly popular in Europe and North America. There is evidence that quality control in terms of species authentication is sometimes inappropriate. Repeated incidents of adulterations and wrong identification, some even with serious consequences have occurred recently. The necessity of a quality control for TCM drugs to avoid these incidents is given since many years. DNA barcoding was used in this study to authenticate drugs which are often used in Chinese herbal medicine. 37 plants from 28 families were identified using nucleotide sequences of the *rbc*L gene. Only one adulteration could be detected. Both the advantages and limitations of *rbc*L as a marker gene for identification were analysed and discussed. We could show that DNA barcoding is a valid and fast method to identify medicinal herbs, showing some advantages over chemical profiling because of its universal application even for unknown plant species.

*Corresponding author: Michael Wink Institute of Pharmacy and Molecular Biotechnology, Heidelberg University, Im Neuenheimer Feld 364, D-69120 Heidelberg, Germany. Telephone: +49 (0) 62 21 / 54 48 81 email: wink@uni-hd.de

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.319v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

1. Introduction

The complex nomenclature in Traditional Chinese Medicine is an acknowledged yet unsolved problem, which is responsible for potentially fatal confusions (Wu et al., 2007). Currently, 4773 botanicals are listed as used in TCM (Jiangsu New Medicine College Editorial Board 1995). The scientific term “species” often does not correlate with the nomenclature used in TCM. We would like to elaborate this in a few examples. Four main classifications are common in TCM nomenclature.

1. Drug and plant name are identical, e.g. *Panax ginseng* is called ren shen (人参) both as a drug and as a plant species.
2. Drug and plant name differ, even though the drug is derived from a single species. *Ginkgo biloba* is called bai guo (白果) as a drug while the species is referred to as yin xing (银杏).
3. The different parts of a particular species have different names as drugs. *Trichosanthes kirilowii* is a good example where the plant is called gua lou (瓜蒌); the fruit bears the same name, while the seed is called gua lou zi (瓜蒌子), the pericarp gua lou pi (瓜蒌皮) and the root tian hua fen (天花粉).
4. Several plant species can be combined under one drug name, making it impossible to know which exact species has been used. One example is lao guan cao (老鹳草) which can be either *Erodium stephanianum*, *Geranium carolinianum* or *G. wilfordii* (Table 1).

To make things worse, several substitutions are allowed in TCM, so that han fang ji (漢防己), *Stephania tetrandra* can be substituted by mu fang ji (木防己), *Cocculus trilobus* or *C. orbiculatus* or by guang fang ji (廣防己), *Aristolochia fangchi* (Wu et al., 2007). Intoxication with renal failure due to carcinogenic aristolochic acids, as reported in 1993, might be due to such nomenclature difficulties (CFSAN/US FDA 2001). These difficulties are intrinsic problems based on the complex nature of TCM. However, frequent adulterations of expensive drugs with cheap, similar looking species cause additional problems (But et al., 1996, 1994, 1993). Therefore, TCM needs rigorous quality control which allows a reliable authentication of the plant material.

Plant drugs can be authenticated by several methods:

1. Microscopic and macroscopic analysis

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.196v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

2. Identification via the phytochemical profiling
3. Identification via DNA sequences of marker genes

Place of origin, age, season and treatment all affect the chemical profile (Xie and Leung, 2009), while they have no influence on the DNA. Chemical markers are further sensible to severe errors, since they need to be specific for the species, stable during storage and modification processes and should represent the therapeutically relevant compound. Especially the latter is often extremely difficult to achieve, since the active principle is either not known or ignores the other compounds responsible for modifying the pharmaceutical effect (Li et al., 2009). A more holistic approach is chemical profiling by HPLC and mixture NMR (metabolomics).

Here, many of the problems of individual chemical markers are avoided since the profile represents the whole spectrum of compounds (van Beek and Montoro, 2009; Yi et al., 2009; Zhang and Ye, 2009). One major drawback is that the profile is extremely sensitive to origin, age, season and processing the drug went through before being sold on the market and thus no profile is identical.

To better cope with these problems, authentication of the herbs can be accessed from a less variable character, the DNA. The genetic information is not affected by the factors mentioned before but remains constant allowing the reliable identification of a plant (Chang et al., 2006; Ma et al., 2002). The comparison of nucleotide sequences of marker genes is often referred to as DNA barcoding. Using this method, conclusions regarding the relationship between plant families, species and even individuals can be obtained. The choice of the marker gene determines the grade of separation that can be detected. Different DNA methods to identify Chinese medical materials have recently been reviewed (Heubl, 2010; Yip, 2007).

As discussed above, several aspects require to be considered addressing the complex problem of quality control of TCM. To get reproducible results, a combination of chemical profiling together with the identification of the plant via DNA is crucial to avoid toxic substitutions. We will show in this study the practicability of DNA barcoding to authenticate TCM plants using nucleotide sequences of the chloroplast gene *rbc*L, which is widely used in plant systematics and therefore well represented in GenBank.

2. Material and methods

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.196v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

2.1 Plant material

We analysed 37 herbal drugs purchased in the herbal market of Shanghai, China, belonging to 29 families and 23 orders. Plant samples were deposited at the IPMB, Heidelberg. Authentic species were obtained from the Botanical Garden, Heidelberg, and further 886 DNA sequences were retrieved from the online database GenBank.

2.2 DNA extraction, amplification and sequencing

Chloroplast DNA was extracted from the herbal material using the chloroform extraction method (Doyle and Doyle, 1987). Chloroplast DNA was amplified using a primer pair for ribulose-bisphosphate carboxylase large chain (*rbc*L) obtained from MWG Biotech AG. As forward primer *rbc*L-N (5' ATGTCACCACAAACAGAAACTAAAGC 3') was used, as reverse primer *rbc*L-leg7 (5' TTCRCATGTACCYGCAGTAGCA 3'), obtaining a PCR product of approximately 700 bp length (TRIO Thermoblock Biometra). The PCR-mix contained 5 µl buffer, 1.5 µl nucleotide mix (100µM), 0.5 µl BSA (10mg/ ml), 0.2 µl Taq polymerase (5 units/ µl), 0.5 µl primer *rbc*L-N and 0.5 µl primer *rbc*L-leg7 (concentration: 10 pM/ µl) and 2 µl DNA solution. The temperature program was 94 °C 5 min, 94 °C 43 sec, 50 °C 1 min, 72 °C 2 min (38 times), 72 °C 20 min. The purified PCR products were sequenced on a MegaBace 1000 instrument (GE Healthcare). Dye-terminator sequencing provided reliable >1000 nucleotide long fragments (Olsvik et al., 1993; Kress et al., 2005), sufficient for the 700 bp fragments obtained in the PCR.

2.3 Sequence alignment and data analysis

Clustal W was used to align the sequences (Thompson et al., 1994); the genetic distances were calculated using MEGA 4.0 following the Kimura 2-Parameter (K2P) model (Tamura et al., 2007). BLAST database search was performed as described previously (Altschul et al., 1990); Neighbour-joining (NJ) and Maximum Likelihood (ML) were used to reconstruct phylogenetic tree (Saitou and Nei, 1987).

3. Results

37 herbal drugs traditionally used in TCM were authenticated according to partial nucleotide sequences of *rbc*L. The 37 plants belong to 28 families and 23 orders; they were chosen to represent the high diversity of plants used in TCM and to test the utility of *rbc*L for barcoding of herbal medicine. In 75% of the drugs, species identity could be confirmed by comparison

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.196v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

with authentic DNA sequences (plants or GenBank accessions); in 25% this was possible only at the genus level, which is similar to the findings of Arif et al. (2010) (see table 2). In 8 cases, the interspecific variations of *rbc*L within a genus were too small to allow the distinction of species while in 28 cases, this was possible. One drug (*Fraxinus rhynchophylla*) was substituted with *Arctium lappa*. BLAST search resulted in 100% identity (627/ 627 bp) with 0 gaps. DNA isolation and amplification of the *rbc*L gene sequence was repeated three times with similar results to make sure that there was no mistake in sample processing.

Two examples are given to exemplify this (table 3, 4). *Equisetum hiemale*, (Equisetaceae) could be identified with p-distances within the genus ranging between 0.003 and 0.03. The next family, Lycopodiaceae, has already p-distances of 0.15. The second example is the genus *Coptis*. TCM does not differentiate between the three species used in TCM; *rbc*L did not allow the exact identification since the sequence of *C. chinensis* and *C. deltoidea* was identical. *C. teeta* could be excluded because of a difference at position 498. The p-distances within the genus range between 0.000 and 0.009, within the family Ranunculaceae between 0.03 and 0.04 and within the order between 0.05 and 0.09. The phylogenetic trees of these two examples further visualize and document the relationships (Fig. 1, 2).

4. Discussion

The genetic authentication via DNA barcoding is an important aspect of quality control to increase the safety of TCM drugs. Unfortunately, substitutions and adulterations with cheaper plants are a well-known phenomenon in TCM (Yip et al., 2007). DNA barcoding is increasingly used to identify these substitutes and adulterations (Heubl, 2010; Guo et al., 2011; Li et al., 2012; Lu et al., 2005). In one study by Mihalov et al. (2000), soybean was detected as an adulteration in *P. ginseng* preparations. We could also detect an alteration in our sample of *Fraxinus rhynchophylla*. Instead, *Arctium lappa* was present, as BLAST and comparison of the gene sequence unequivocally revealed.

However, adulterations are not the only problem quality control of TCM has to face. The more common problem lies in the system of TCM itself. As explained in the Introduction, the complex nomenclature of TCM plants can be responsible for unintentional substitutions with fatal consequences (Wu et al., 2007).

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.196v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

A major problem every DNA barcoding approach of TCM drugs has to face is the often problematic condition of the DNA (Heubl, 2010). Due to the various processing methods such as drying, steaming, bleaching etc TCM drugs have to undergo before being sold, the DNA is often badly damaged. Additionally, the secondary metabolites such as flavonoids, tannins or alkaloids inhibit the PCR amplification of the selected marker gene. Intercalating substances can disturb PCR resulting in mismatched base pairs and erroneous DNA sequences.Therefore, rigorous purification of the DNA is essential to reduce the secondary metabolites before a successful amplification can be tried (Shahzadi et al., 2010; Ribeiro and Lovato, 2007).

The challenge is to discover a DNA marker that is general enough not to raise false alarm but specific enough to discover all adulterations. Furthermore, the marker must be universal to cover the large variety of plant species applied in TCM. And, last but not least, the DNA must ideally exist in many copies to increase the chance of detection in TCM drugs which are usually dried and grounded and thus often contain degraded DNA. Several studies showed that a marker fulfilling all these requirements hardly exists and we need to live with certain restrictions (Rubinoff et al., 2006; Yip et al., 2007). However, the level of identification can be directed by carefully choosing the adequate target region of the genome. Promising for the identification of herbal medicine are especially chloroplast genes since chloroplasts contain many copies of the same gene and thus increase the chances of successful detection (Chase et al., 1993). The interspecific variations change from plant species to plant species which means the decision for the right marker gene can not be absolute but needs to be adapted to the particular situation (Song et al., 2005; Yip et al., 2007).

Nevertheless, two marker genes, *rbc*L and ITS (a commonly used nuclear marker), are widely used in DNA barcoding. Several examples of successful identification of TCM drugs both of *rbc*L (Mihalov et al., 2000; Song et al., 2005) and ITS (Chen et al., 2010; Gao et al., 2010; Yang et al., 2007; Lu et al., 2005) were published recently.

To choose the right gene, we have to remember the nature of TCM. Quite often, several closely related members of the same genus are used as one drug; a marker gene distinguishing between these species or even subspecies might raise false alarm. ITS is useful to detect a plant at the species level, while *rbc*L is in 75% of the cases precise enough to determine the species but can not distinguish drugs on the species level in the remaining 25% of the cases (Chase et al., 1993; Yip et al., 2007; Arif et al., 2010). This disadvantage of *rbc*L for

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.196v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

phylogenetic research might be an advantage in quality control of herbal medicine. It is important to detect adulterations beyond doubt, but the exact species within a genus is only of secondary importance, since TCM itself does not differentiate to that level and also phytochemical profiles are similar between closely related species. *rbc*L can fulfil these requirements successfully and can guarantee the safety of the drug, as we could demonstrate in our study. The level of detection of *rbc*L is high enough to discover possibly toxic substitutions within the traditional system of TCM, such as *Aristolochia* species as substitutes for *Stephania*. Adulterations can be identified using database search since an extensive library of most families and even most genera exists already for these two marker genes. Furthermore, the *rbc*L marker is present in many copies in each plant cell, making a successful amplification more probable than for nuclear genes.

In our study we have demonstrated the utility of *rbc*L as marker for DNA barcoding, but have to point out its limitations as well. Since genetic information does not cover the morphology, chemical profile, quality control should always try to consider different techniques. It is advisable to establish a TCM library of all *rbc*L sequences for international use to allow rapid detection since authentic species examples are rather difficult to obtain outside of Asia. This would improve the acceptance of TCM internationally beyond the image of traditional medicine.

Acknowledgement:
We thank our Chinese colleague Dr. Zhe Zhou for verifying the Chinese characters and TCM drug names.

References:

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *Journal of Molecular Biology* 215: 403-410.

Arif IA, Bakir MA, Khan HA, Al Farhan AH, Al Homaidan AA, Bahkali AH, Al Sadoon M, Shobrak M. 2010. A brief review of molecular techniques to assess plant diversity. *International Journal of Molecular Sciences*. 11: 2079-2096.

But PP. 1993. Need for correct identification of herbs in herbal poisoning. *Lancet.* 341: 637.

But PP. 1994. Herbal poisoning caused by adulterants or erroneous substitutes. *The Journal of Tropical Medicine and Hygiene*. 97: 371-374.

But PP, Tomlinson B, Cheung KO, Yong SP, Szeto ML, Lee CK. 1996. Adulterants of herbal products can cause poisoning. *British Medical Journal*. 313: 117.

CFSAN/US FDA, Office of Nutritional Products, Labelling and Dietary Supplements. 2001. Letter to industry associations regarding safety concerns related to the use of botanical products containing aristolochic acid. *Available at http://www.fda.gov/Food/DietarySupplements/Alerts/ucm096374.htm* (accessed 25 february 2011).

Chang WT, Thissen U, Ehlert KA, Koek MM, Jellema RH, Hankemeier T, van der Greef J, Wang M. 2006. Effects of growth conditions and processing on *Rehmannia glutinosa* using fingerprint strategy. *Planta Medica*. 72: 458-467.

Chase MW, Soltis DE, Olmstead RG, Morgan D, Les DH, Mishler BD, Duvall MR, Price RA, Hills HG, Qiu YL. 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from plastid gene *rbc*L. *Annals of the Missouri Botanical Garden*. 80: 528-580.

Chen S, Yao H, Han J, Liu C, Song J, Shi L, Zhu Y, Ma X, Gao T, Pang X, Luo K, Li Y, Li X, Jia X, Lin Y, Leon C. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. *PLoS ONE*. 5: 1.

Doyle JJ, Doyle JL. 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin.* 19: 11–15.

Gao T, Yao H, Song J, Liu C, Zhu Y, Ma X, Pang X, Xu H, Chen S. 2010. Identification of medicinal plants in the family Fabaceae using a potential DNA barcode ITS2. *Journal of Ethnopharmacology*. 130: 116-121.

Guo X, Wang X, Su W, Zhang G, Zhou R. 2011. DNA barcodes for discriminating the medicinal plant *Scutellaria baicalensis* (Lamiaceae) and its adulterants. *Biological & Pharmaceutical Bulletin*. 34: 1198-1203.

Heubl G. 2010. New aspects of DNA-based authentication of Chinese medicinal plants by molecular biological techniques. *Planta Medica*. 76: 1963-1974.

Jiangsu New Medicine College Editorial Board. 1995. *A Grand Dictionary of Chinese Medicinal Herbs, 3rd Ed*. Shanghai: Shanghai Science Technology Publishing Co., 2754.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. *Proceedings of the National Academy of Sciences of the United States of America*. 102: 8369-8374.

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.196v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

Li M, Au KY, Lam H, Cheng L, Jiang RW, But PP, Shaw PC. 2012. Identification of Baiying (Herba Solani Lyrati) commodity and its toxic substitute Xungufeng (Herba Aristolochiae Mollissimae) using DNA barcoding and chemical profiling techniques. *Food Chemistry*. 135: 1653-1658.

Li S, Han Q, Qiao C, Song J, Cheng CL, Xu H. 2008. Chemical markers for the quality control of herbal medicines: an overview. *Chinese Medicine*. 3: 7.

Lu KT, Lo CF, Chang HC, Lin JH. 2005. Identification of *Saposhnikoviae* radix in concentrated Chinese medicine preparations by nested PCR and DNA sequencing methods. *Journal of Food and Drug Analysis*. 13: 219-224.

Ma XQ, Shi Q, Duan JA, Dong TT, Tsim KW. 2002. Chemical analysis of Radix *Astragali* (Huangqi) in China: a comparison with its adulterations and seasonal variations. *Journal of Agricultural and Food Chemistry*. 50: 4861-4866.

Mihalov JJ, Marderosian AD, Pierce JC. 2000. DNA identification of commercial ginseng samples. *Journal of Agricultural and Food Chemistry*. 48: 3744-3752.

Olsvik O, Wahlberg J, Petterson B, Uhlén M, Popovic T, Wachsmuth IK, Fields PI. 1993. Use of automated sequencing of polymerase chain reaction-generated amplicons to identify three types of cholera toxin subunit B in *Vibrio cholerae* O1 strains. *Journal of Clinical Microbiology*. 31: 22-25.

Ribeiro RA, Lovato MB. 2007. Comparative analysis of different DNA extraction protocols in fresh and herbarium specimens of the genus Dalbergia. *Genetics and Molecular Research*. 6: 173-187.

Rubinoff D, Cameron S, Will K. 2006. Are plant DNA barcodes a search for the Holy Grail? *Trends in Ecology & Evolution*. 21: 1-2.

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

Shahzadi I, Ahmed R, Hassan A, Shah MM. 2010. Optimization of DNA extraction from seeds and fresh leaf tissues of wild marigold (*Tagetes minuta*) for polymerase chain reaction analysis. *Genetics and Molecular Research*. 9: 386-393.

Song J, Yao H, Li Y, Li X, Lin Y, Liu C, Han J, Xie C, Chen S. 2009. Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. *Journal of Ethnopharmacology*. 124: 434-439.

Tamura K, Dudley J, Nei M, Kumar S. 2007. MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*. 24: 1596-1599.

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. 22: 4673-4680.

Van Beek TA, Montoro P. 2009. Chemical analysis and quality control of *Ginkgo biloba* leaves, extracts, and phytopharmaceuticals. *Journal of Chromatography A.* 1216. 2002-2032.

Wu KM, Farrelly JG, Upton R, Chen J. 2007. Complexities of herbal nomenclature system in traditional Chinese medicine (TCM): Lessons learned from the misuse of *Aristolochia*-related species and the importance of pharmaceutical name during botanical drug development. *Phytomedicine*. 14: 273-279.

337 Xie PS, Leung AY. 2009. Understanding the traditional aspect of Chinese medicine in order to achieve
338 meaningful quality control of Chinese materia medica. *Journal of Chromatography A*. 1216: 1933-
339 1940.
340
341 Yang ZY, Chao Z, Huo KK, Xie H, Tian ZP, Pan SL. 2007. ITS sequence analysis used for molecular
342 identification of the *Bupleurum* species from northwestern China. *Phytomedicine*. 14: 416-423.
343
344 Yi L, Liang Y, Wu H, Yuan D. 2009. The analysis of Radix Angelicae Sinensis (Danggui). *Journal of*
345 *Chromatography A*. 1216: 1991-2001.
346
347 Yip PY, Chau CF, Mak CY, Kwan HS. 2007. DNA methods for identification of Chinese medicinal
348 materials. *Chinese Medicine*. 2: 9.
349
350 Zhang Q, Ye M. 2009. Chemical analysis of the Chinese herbal medicine Gan-Cao (licorice). *Journal*
351 *of Chromatography A*. 1216: 1954-1969.

Table 1: Examples for the complex nomenclature in Traditional Chinese Medicine

Case	Scientific name of the plant	Chinese name of the plant	Part used	Chinese name of the drug
Case 1	*Panax ginseng*	ren shen 人参	root	ren shen 人参
Case 2	*Ginkgo biloba*	yin xing 银杏	seed	bai guo 白果
Case 3	*Trichosanthes kirilowii*	gua lou 瓜蒌	fruit	gua lou 瓜蒌
			seed	gua lou zi 瓜蒌子
			roasted seed	chao gua lou zi 炒瓜蒌子
			pericarpum	gua lou pi 瓜蒌皮
			root	tian hua fen 天花粉
Case 4	*Erodium stephanianum*	mang niu er miao 牻牛儿苗	herb	lao guan cao 老鹳草
	Geranium wilfordii	lao guan cao 老鹳草	herb	lao guan cao 老鹳草
	Geranium carolinianum	ye lao guan cao 野老鹳草	herb	lao guan cao 老鹳草

PeerJ PrePrints | http://dx.doi.org/10.7287/peerj.preprints.196v1 | CC-BY 3.0 Open Access | received: 12 Jan 2014, published: 12 Jan 2014

Table 2: Plants studied and identified by *rbc*L.

A. Identification: Species level

No	IPMB Accession number	Genbank Accession number	Species, Family, Order
1	P6839 / 05	JF949994	*Arctium lappa*, Asteraceae, Asterales
2	P6843 / 09	JF949995	*Belamcanda chinensis*, Iridaceae, Asparagales
3	P6883 / 49	JF949996	*Berberis bealei*, Berberidaceae, Ranunculales
4	P6846 / 12	JF949997	*Capsella bursa-pastoris*, Brassicaceae, Brassicales
5	P6859 / 25	JF949998	*Cyrtomium fortunei*, Dryopteridaceae, Polypodiales
6	P6860 / 26	JF949999	*Dendrobium loddigesii*, Orchidaceae, Asparagales
7	P6863 / 29	JF950000	*Eclipta prostrata*, Asteraceae, Asterales
8	P6864 / 30	JF950001	*Ephedra sinica*, Ephedraceae, Gnetales
9	P6865 / 31	JF950002	*Epimedium koreanum*, Berberidaceae, Ranunculales
10	P6866 / 32	JF950003	*Equisetum hiemale*, Equisetaceae, Equisetales
11	P6894 / 60	JF950004	*Fallopia japonica*, Polygonaceae, Caryophyllales
12	P6872 / 38	JF950005	*Ginkgo biloba*, Ginkgoaceae, Ginkgoales
13	P6875 / 41	JF950006	*Houttuynia cordata*, Saururaceae, Piperales
14	P6879 / 45	JF950007	*Kadsura longipedunculata*, Schisandraceae, Austrobaileyales
15	P6882 / 48	JF950008	*Magnolia officinalis*, Magnoliaceae, Magnoliales
16	P6885 / 51	JF950009	*Ophioglossum vulgatum*, Ophioglossaceae, Ophioglossales
17	P8088 / 81	JF950028	*Panax ginseng*, Araliaceae, Apiales
18	P6888 / 54	JF950010	*Paris polyphylla*, Melanthiaceae, Liliales
19	P6891 / 57	JF950011	*Platycladus orientalis*, Cupressaceae, Pinales
20	P6893 / 59	JF950012	*Polygonum aviculare*, Polygonaceae, Caryophyllales
21	P6896 / 62	JF950013	*Prunella vulgaris*, Lamiaceae, Lamiales
22	P6897 / 63	JF950014	*Punica granatum*, Lythraceae, Myrtales
23	P6898 / 64	JF950015	*Rheum officinale*, Polygonaceae, Caryophyllales
24	P6901 / 67	JF950016	*Sanguisorba officinalis*, Rosaceae, Rosales
25	P6903 / 69	JF950017	*Scutellaria baicalensis*, Lamiaceae, Lamiales
26	P6904 / 70	JF950018	*Selaginella tamariscina*, Selaginellaceae, Selaginellales
27	P6908 / 74	JF950019	*Taraxacum officinale*, Asteraceae, Asterales
28	P6910 / 76	JF950020	*Verbena officinalis*, Verbenaceae, Lamiales

B. Identification: Genus level

29	P6844 / 10	JF950021	*Bupleurum chinense*, Apiaceae, Apiales
30	P6849 / 15	JF950022	*Centella asiatica*, Apiaceae, Apiales
31	P6853 / 19	JF950023	*Cinnamomum cassia*, Lauraceae, Laurales
32	P6855 / 21	JF950024	*Coptis chinensis*, Ranunculaceae, Ranunculales
33	P6873 / 39	JF950025	*Glycyrrhiza inflata*, Fabaceae, Fabales
34	P6886 / 52	JF950026	*Paeonia lactiflora*, Paeoniaceae, Saxifragales
35	P6887 / 53	JF950030	*Panax notoginseng*, Araliaceae, Apiales
36	P6892 / 58	JF950027	*Polygonatum kingianum*, Ruscaceae, Asparagales

C. Identification: Substitution of the TCM drug *Fraxinus rhynchophylla* with *Arctium lappa*

37	P6871 / 37	*Arctium lappa*, Oleaceae, Lamiales

372 Table 3: Phylogeny of *Equisetum hiemale*, Equisetaceae, with *Lycopodium* and *Polypodium*
373 as outgroups
374

	Species	Family	Order	Genbank Accession number
1	*Equisetum hiemale* (TCM)	Equisetaceae	Equisetales	JF950003
2	*Equisetum hiemale*	Equisetaceae	Equisetales	EU677110
3	*Equisetum hiemale* Bot. Garden Heidelberg	Equisetaceae	Equisetales	-
4	*Equisetum arvense*	Equisetaceae	Equisetales	L11053
5	*Equisetum bogotense*	Equisetaceae	Equisetales	AY226139
6	*Equisetum diffusum*	Equisetaceae	Equisetales	AY226141
7	*Equisetum fluviatile*	Equisetaceae	Equisetales	DQ463101
8	*Equisetum palustre*	Equisetaceae	Equisetales	GQ248601
9	*Equisetum pratense*	Equisetaceae	Equisetales	AY226137
10	*Equisetum sylvaticum*	Equisetaceae	Equisetales	AY226136
11	*Equisetum telmateia*	Equisetaceae	Equisetales	AF313580
12	*Equisetum variegatum*	Equisetaceae	Equisetales	AY226134
13	*Equisetum x ferrissii*	Equisetaceae	Equisetales	AF313579
14	*Lycopodium annotinum*	Lycopodiaceae	Lycopodiales	EU352290
15	*Polypodium scouleri*	Polypodiaceae	Filicales	FJ825693

375

376 Table 4: Phylogeny of *Coptis chinensis*, Ranunculaceae, and other families of the
377 Ranunculales
378

	Species	Family	Genbank Accession number
1	*Coptis chinensis* (TCM)	Ranunculaceae	JF950024
2	*Coptis chinensis*	Ranunculaceae	AB163775
3	*Coptis deltoidea*	Ranunculaceae	AB163774
4	*Coptis teeta*	Ranunculaceae	AB163773
5	*Coptis aspleniifolia*	Ranunculaceae	AB163777
6	*Coptis japonica_var._anemonifolia*	Ranunculaceae	AB163764
7	*Coptis japonica_var._major*	Ranunculaceae	AB163765
8	*Coptis laciniata*	Ranunculaceae	AB163778
9	*Coptis lutescens*	Ranunculaceae	AB163766
10	*Coptis occidentalis*	Ranunculaceae	AB163779
11	*Coptis omeiensis*	Ranunculaceae	AB163776
12	*Coptis quinquefolia*	Ranunculaceae	AB163770
13	*Coptis quinquesecta*	Ranunculaceae	AB163772
14	*Coptis ramosa*	Ranunculaceae	AB163769
15	*Coptis trifolia*	Ranunculaceae	AF093730
16	*Coptis trifoliolata*	Ranunculaceae	AB163768
17	*Aconitum napellus*	Ranunculaceae	EU053898
18	*Anemone hupehensis*	Ranunculaceae	FJ626577
19	*Aquilegia vulgaris*	Ranunculaceae	FJ449851
20	*Clematis montana*	Ranunculaceae	FJ449855
21	*Delphinium bonvalotii*	Ranunculaceae	FJ626583
22	*Glaucidium palmatum*	Ranunculaceae	L75848
23	*Hydrastis canadensis*	Ranunculaceae	L75849
24	*Ranunculus japonicus*	Ranunculaceae	FJ449862
25	*Thalictrum simplex*	Ranunculaceae	FJ449863
26	*Berberis bealei*	Berberidaceae	FJ449858
27	*Circaeaster agrestis*	Circaeasteraceae	FJ626607
28	*Euptelea pleiosperma*	Eupteleaceae	AY048174
29	*Lardizabala biternata*	Lardizabalaceae	D85693
30	*Menispermum dauricum*	Menispermaceae	FJ026493
31	*Papaver rhoeas*	Papaveraceae	FJ626614

379
380
381

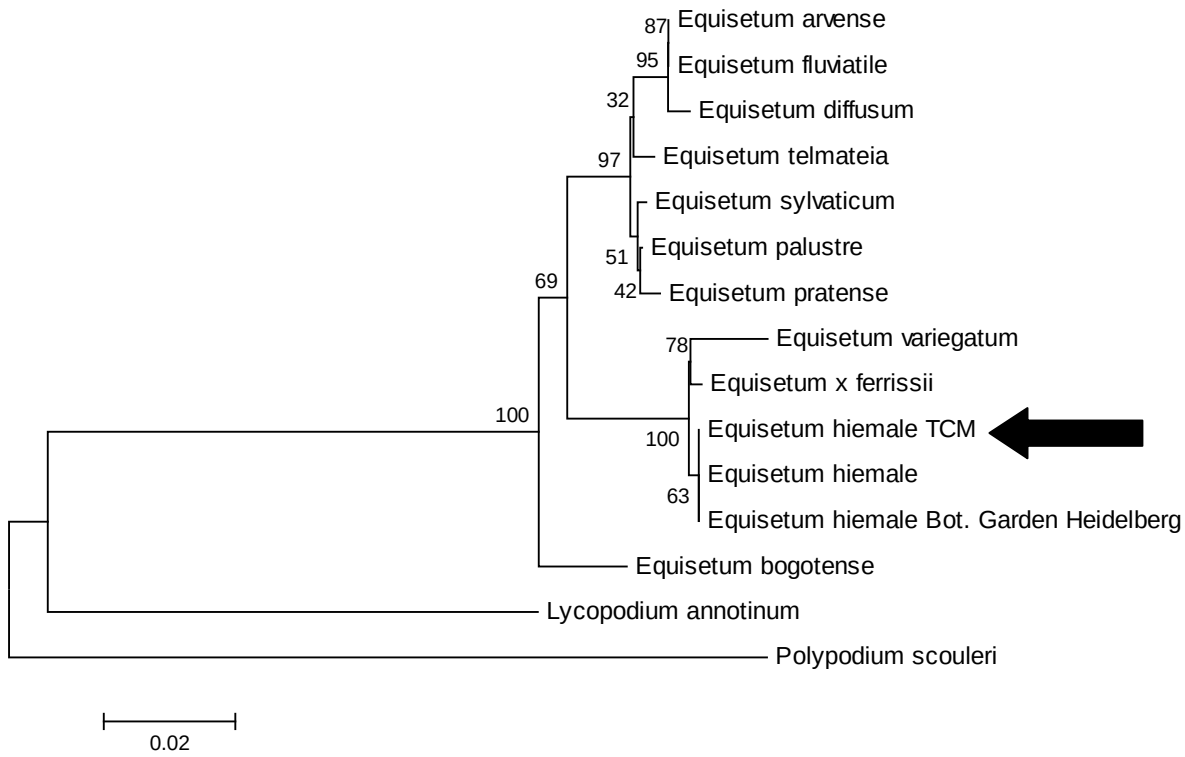

Fig. 1: Phylogenetic NJ tree of *E. hiemale* based on nucleotide sequences of the *rbc*L gene with bootstrap values

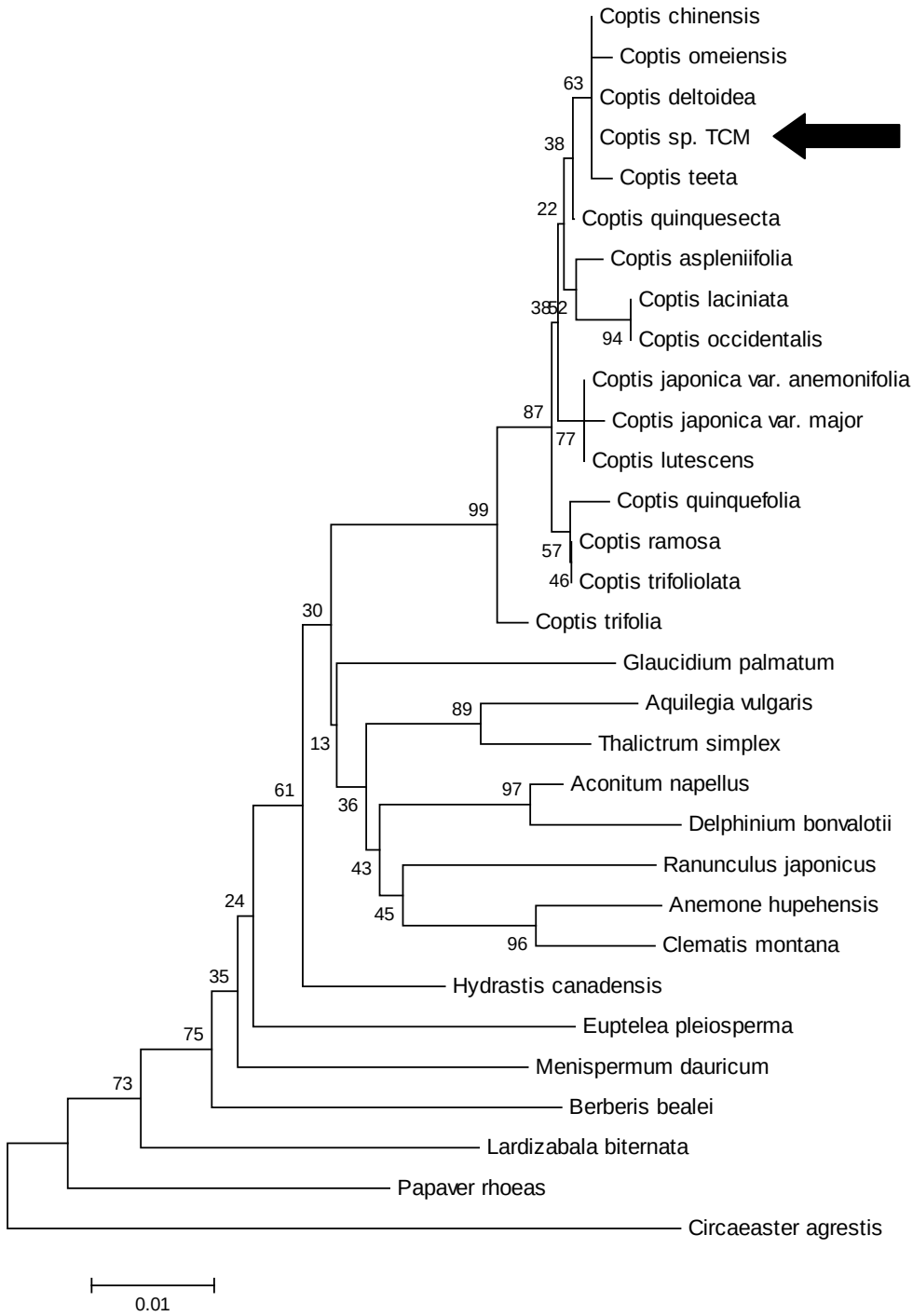

Fig. 2: Phylogenetic NJ tree of *C. chinensis* based on nucleotide sequences of the *rbc*L gene with bootstrap values