Advances of biological taxonomy and species identification in Medicinal Plant Species by DNA barcodes

Chong Liu [1]，Zhengyi Gu[1]，Weijun Yang [1]*，Li Yang [2]，Dilnuer[1]

[1.] Xinjiang Institute of Materia Medica/Key Laboratory of Xinjiang Uygur Medicine，Urumqi 830004，China；liu_chong02@163.com

[2.] Pharmacy school of Xinjiang Medical University，Urumqi 830004，China；

Abstract: Medicinal Plant Species taxonomy is authenticated according to morphological features. It is a long-standing problem of mixing authentic species with their adulterants in medicinal preparations. However, DNA barcoding is a new technique that uses a short DNA sequence as a molecular diagnostic for species-level identification, Our purpose is to briefly expose DNA Barcode of Life principles, relevance and universality. Barcode of life framework has greatly evolved, giving rise to a flexible description of DNA barcoding and a larger range of applications. Similarly, a variety of single locus or combined loci have been propose as DNA barcodes for the plant identification, which are the coding regions or non-coding regions in plastids or the nuclear genome, such as *rbcL, matK, rpoB, rpoc1, psbA-trnh, ITS and rbcL+psbA-trnH.*

[Chong Liu，Zhengyi Gu，Weijun Yang，Li Yang，Dilnuer. **Advances of biological taxonomy and species identification in Medicinal Plant Species by DNA barcodes.** Journal of American Science 2011;7(5):147-151]. (ISSN: 1545-1003). http://www.americanscience.org.

Key words: DNA barcoding; ITS2; matK; psbA-trnH

1 Introduction

Traditionally, the Medicinal Plant Species taxonomy is authenticated according to their morphological features. Even until now, morphological keys still are main basis of taxonomy (Wang, 2006). However, there is a long-standing problem of mixing authentic species with their adulterants in medicinal preparations. This problem has consequences for medical safety as well as for the conservation of the authentic species (Ma, 2007; Zhou, 1995; Mao, 2004;). The classical morphological authentication approach was confronted with difficulties due to overly similar traits used for taxonomic characterization (Gao, 2009 and Ching, 1978) and an ever decreasing number of specialists.

However, the term “DNA barcode” for global species identification was first coined by Hebert in 2003 and has gained worldwide attention in the scientific community (Shilin chen, 2010). Recognition of animals, plants and fungi has been performed using this technique. DNA barcode is a short fragment of the genome and species identification using DNA barcoding has been completed in several studies (Gregory, 2005 and Lahaye, 2008) taxonomists (Schindel, 2005) and is likely to usher the taxonomic research into a new era.

Most researchers agree that the mitochondrial gene encoding cytochrome coxidase subunit 1 is a favorable region for use as a DNA barcode in most animal species and even in some fungal species, including those of the groups Ascomycota, Basidiomycota and Chytridiomycota. However, the CO1 gene and other mitoc excluded hondrial genes from plants have limited usefulness for identifying plant species across a wide range of taxa due to the low amounts of variation in the genes, as well as the variable structure of the mitochondrial genome (Pennisi, 2007; Chase, 2005, Fazekas, 2008; Chase, 2007; Daniel; 2006). Thus, screening for single or multiple regions appropriate for DNA barcoding studies in nuclear and plastid genomes in plants has been an important research focus.

2 DNA barcoding as a driving force in the Plant

First, most single-copy genes in the nuclear genome, as well as their introns, have been as barcode candidates because of the lack of universal primers for their amplification[9]. However, with the exception of 5.8S, the internal transcribed spacer (ITS) of nuclear ribosomal DNA and regions of the ITS could be potential barcodes (Lahaye, 2008 and Chase, 2007). Second, extensive studies have focused ongenes and introns of the chloroplast genome. Comparing with animals, CO1 gene evolves slower in plants. Therefore, CO1gene is not appropriate to be the standard of barcode for plant. Recently, the research and application of DNA barcoding in plant is in the exploration stage. So the most important task is to select candidate genes or DNA regions and identify a suitable DNA barcode. There are many researchers did exploration on filtrating the plant barcode, and they try to find the perfect DNA barcode from chloroplast gene.

2.1 Single fragment DNA barcode in plant

ITS2 ribosome transcribed spacer gene states in RNA gene 5.85 and 28S compartment. Its evolutive speed is faster, so that the more evolution information can be acquired. In 2010, Shiling Chen et al (2010) show the study on filtrating DNA barcode sequence in 8557 medical plant samples. The samples were about Angiosperms, gymnosperms, ferns, mosses, lichens, algae and fungi. Date show compared seven candidate DNA barcodes (psbA-trnH, matK, rbcL, rpoC1, ycf5,ITS2, and ITS) from medicinal plant species. According to the result, ITS2 sequence has the best performance, the discrimination ability of ITS2 in more than 6600 plant samples belonging to 4800 species from 753 distinct genera and found that the rate of successful identification with the ITS2 was 92.7% at the species level. At the same year, Xiaohui Pang et al (2010) show adopted ITS2 sequence identify 888、1410、1183 samples form Angiosperms, gymnosperms and Euphorbiaceae, and the identification rates are 73%、100%、90%. Therefore, The ITS2 region can be potentially used as a standard DNA barcode to identify medicinal plants and their closely related species.

Comparing with other chloroplast cording gene, *matk* evolutionary speed is faster, and it is more appropriate to PCR amplification matk gene segments. In 2008, Labaye at al (2008) show had 100% success by matk gene to PCR amplification in 1667 plant materials. However, Kress et al (2007) argued that if other Family plants have this high success rate on amplification, because the Labaye's 96% using materials are Orchidaceae. The different branch groups of *matk* gene are difficult to be extended and sequencing , because the *matk* primer is not universal. According to the result in table1, Fazekas et al (2008) had 87.6% success rate by using 10 pairs primers to amplification 251 samples form 32 genera plants. *Matk* single segment has only 56.0% success on identifying the different plants, and there are 25.5% the original forward and reverse sequencing are less than 80%. In 2007, Sass and Kress et al (Kress, 2007 and Sass, 2007) did not acquire the expected identification result in Cycadales, and its correct identification rate are 24% and 14.6% separately. Therefore, the design and effective identification in *matk* gene universal primer will be important for next stage.

Another plastid DNA region proposed is the non-coding *psbA-trnH* spacer (Kress, 2005 and Shaw, 2007). This region is one of the most variable non-coding regions of the plastid genome in angiosperms in terms of having the highest percentages of variable sites (Shaw, 2007).This variation means that this inter-genic spacer can offer high levels of species discrimination. Kress et al (Kress, 2005) compared 10 loci for authenticating closely related species in 7 plant families and 99 species belonging to 88 genera in 53 families, they found that the extended length of *trnH-psbA* is 247-1221 bp, and the space region is 119-1094 bp, they reported that the *psbA-trnH* spacer and the internal transcribed spacer could be used as a pair of potential barcodes for identifying widely divergent angiosperm taxa. Moreover, Kress et al (Kress, 2005) and Fazekas et al shown the results that *trnH-psbA* has the highest success ratio of extension and the highest exactness ratio of identification in the alternative segments. *trnH-psbA* also has the best performance on distinguishing the relate plant. Specifically, the identification ratio of *trnH-psbA* is above 90% in Orchidaceae, and 70% in Compsoneura (Kress, 2005). However, Sass et al extended two bands in Cycadales plant by using the primers from Kress. Therefore, *trnH-psbA* space region cannot be the appropriate barcode in Cycadales, and the variance of segments in *trnH-psbA* is not enough for the identification in Heracleum.L and Glyceria.

The protein encoding plastid gene *rbcL* has been proposed as a potential plant barcode by several sets of researchers (Chase, 2005 and Newmaster, 2006), usually in conjunction with one or more other markers. One benefit of this region is the large amount of existing information—there are more than 10,000 *rbcL* sequences already in GenBank (Chase, 2005 and Newmaster, 2006). However, many of these are unvouchered or erroneously identified, and none has electropherogram trace files available, so all of these would have to be repeated to meet the standards for an official "DNA barcode" designation in GenBank. Furthermore, studies by Chase & al. (2005) and Newmaster & al. (2006), which demonstrated a fair degree of success in discriminating species, used nearly entire *rbcL* sequences (at least 1300 bp long). An ideal DNA barcoding region should be short enough to amplify from degraded DNA and analysed via single-pass sequencing. One possibility is to develop primer sets for short portions of this gene to produce a barcode of appropriate length, but our attempts to develop universal primers to achieve this have been unsuccessful to date.

2.2 Combined fragment DNA barcode in plant

In higher plants, the mitochondrial genome evolves much more slowly than in animals. The COI-region is thus inappropriate for plant species distinction (Shilin Chen, 2010). The CBOL plant working group (PWG) agrees that plant barcoding will be multilocus, with one "anchor" (i.e. universal across the plant kingdom) and "identifiers" to distinguish closely related species.

Table 1 Plant DNA Barcodes Identification Rate

Gene	Species	Numbers	Identification Rate	Reference	Time
ITS2	Medicinal plant	8557	92.70%	Silin chen et al	2010
	tragacanth	41	100%	T gao et al	2010
	Gymnosperms	888	73%	pang XH et al	2010
	Rosaceae	1410	100%	pang XH et al	2010
	Euphorbiaceae	1183	90%	pang XH et al	2010
matK	Cycadales		24%	Sass et al	2007
	Relative plant	96	14.60%	Kress and Erickson	2007
	Compsoneura	8	failure	Newmaster	2008
	Relative plant	251	56.00%	Fazekas et al	2008
		1667	100%	Lahaye et al	2008b
trnH-psbA	Relative plant	99	92%	Kress et al	2005
	Pteridophytes Gymnosperms Angiosperm	251	100%	Fazeka et al	2008
	Orbanchaceae	4	100%	Han JP et al	2010
	Compsoneura		70%	Newmaster et al	2008
	Orchidaceae	1667	90%	Lahay et al	2008b
	Heracleum.L Glyceria	87	failure	Whipple et al Logacheva et al	2007 2008
rbcL+trnH-psbA	Relative plant	96	88%	Kress, Erickson et al	2007
	Gymnosperms Angiosperm	251	64%	Fazekas et al	2008
matK + atpF-atpH + psbK-psbl	Monocotyledoneae	101	93.10%	Lahaye et al	2008a
matK +atpF-atpH +trnH-psb	Monocotyledoneae	101	89.30%	Lahaye et al	2008a

There are many researches indicate that the identification ratio of single segment DNA barcode is low. Therefore, barcode selection should not only consider single segment DNA barcode, but also the multiple segments component. Kress and Erichson said (Fazekas, 2008) that *rbcL+trnH-psbA* can be the segment to create the barcode for all land plant. Chase et al emphasized *Matk* is a key factor to identify advanced plant, *rpoC1+rpoB+matk* and *rpoC1+rpoB+rbcL+trnH-psbA*. In 2008, Newmaster et al (2008) did the sampling study on Compsoneura, and did the selection test on 7 chloroplast DNA segments （*UPA*、*rpoB*、*rpoC1*、*accD*、*rbcL*、*matk* and *trnH-psbA*）. The result shows that *matk+trnH-psbA* perform better in identification. In May 2009, CBOL analyzed 907 samples of 550 plants: 445 samples were come from Angiosperm, 38 samples were come from Gymnosperms, and 67 were come from cryptogamous plants. The selection of 7 DNA barcodes of plant were compared the candidate sequence（*atpF-atpH*、*matK*、*rbcL*、*rpoB*、*rpoC1*、*psbK-psb1*、*psbA-rbcL*）, then they suggested subdue *matK*+ *rbcL* combination to be the currency barcode for land plant. However, CBOL thought this combination is not perfect, because its success ratio of identification on species was 72%. Ren BQ et al (Ren BQ et al,2009) examination team focused on Betulales as their research objective in 2009. They did sampling analysis on 131 of all species

in the world (26 species), and evaluated 4 DNA segments（*ITS*、*rbcL*、*matK*、*psbA-trnH*）. The result shown that *ITS* +*psbA-trnH* combination achieved 88% for species-level identification and this was the highest ratio in all subjects. November 2009, Global DNA barcode, the third international meeting in Mexico, not only rbcL and matk, but also *psbA-trnH* sequence gained the regard and agreement form experts. CBOL advised that the currency and ability of identification *ITS* and *psbA-trnH* is very impartant.

The full results of this study will be published elsewhere, and analyses of the efficacy of intensively trialled regions in part 2 are still ongoing. However, as a "research update", we summarise the current state of play. Two plastid gene regions, partial *rpoC1* and *rpoB*, performed well as barcoding regions in terms of being amplifiable with a limited range of PCR conditions and primer sets and, although not particularly rapidly evolving, were able to discriminate among species in many groups of organisms. A third gene region, *matK*, showed much higher levels of sequence variation and provided better species discrimination, but work is still underway to improve PCR primer sets to enhance its 'universality'. The greatest level of species discrimination was achieved when all three regions are combined (*rpoC1*, *rpoB* and *matK*), and this represents one option as a standard DNA barcode for plants.

A second option we present is *rpoC1*, *matK*, and *psbAtrnH*. This option substitutes for the relatively conserved coding region, *rpoB*, the previously mentioned, highly length-variable, non-coding intergenic spacer, *psbA-trnH*. The benefits to doing this are that additional species level resolution may be obtained, while at least part of the plant barcode (the sequences from *rpoC1* and *matK*) will be comparable and alignable across broad evolutionary distances. The downside of this approach relates to the introduction of bioinformatics challenges and problems with degraded tissue due to the variation and often larger size of the *psbA-trnH* spacer.

3 Barcode Candidates in Medicinal Plant

DNA barcoding can be regarded as a tremendous tool to accelerate species discovery and initiate new species descriptions. Moreover, it re-opens the debate on species concepts (Fitzhugh, 2006; Rubinoff 2006b; Balakrishnan, 2007; Miller, 2007 and Vogler, 2007). In addition, DNA barcodes will be a useful and powerful tool for nonprofessional users such as customs officers, traditional drug producers and managers and forensic specialists. Therefore, a rapid and simple DNA barcoding identification system, even an imperfect one, is likely to be welcomed. The search for and development of herbal medicines is rapidly increasing worldwide, so practical and accurate authentication resources are urgently needed (Sucher, 2008; Yao, 2009; Song, 2009 and Shaw, 1997). Chen et al shows the potential for a DNA barcoding technique to become a standard for the authentication of medicinal plants and their adulterants.

Acknowledgements

This work is supported by the Pillar Project of Science & Technology of Xinjiang province (200933123).

Referece:

1. Wang L., He Z., Chin. Wild Plant Resour., 25, 1—4 (2006).
2. Ma X. Q., Tan C. H., Zhu D. Y., Gang D. R., Xiao P. G., J. Ethnopharmacol.,113, 15—34 (2007).
3. Zhou R., Zhou T., China J. Chin. Materia Medica, 20, 199—200(1995).
4. Mao J., Li L., Chin. Tradit. Herbal Drugs, 35, 1066—1067 (2004).
5. Gao Z. P., Su Y. L., Wu J. H., J. Beijing Univ. Tradit. Chin. Med., 32, 568—570 (2009).
6. Ching R. C., Acta Phytotaxonomica Sin., 16, 1—19 (1978).
7. Shilin Chen, Hui Yao, Jianping Han.etal. Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. PlosONE, 2010,5(1),e8613
8. Gregory TR (2005). DNA barcoding does not compete with taxonomy. Nature 434, 1067–1067.
9. Lahaye R, van der Bank M, Bogarin D, et al(2008). DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA 105, 2923–2928
10. Schindel DE, Miller SE. 2005. DNA barcoding a useful tool for taxonomists. Nature 435: 17
11. Pennisi E (2007) Wanted: a barcode for plants. Science 318: 190–191.
12. Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesanakurthi RP, et al.(2005) Land plants and DNA barcodes: short-term and long-term goals. PhilosTrans R Soc Lond B Biol Sci 360: 1889–1895.
13. Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, et al. (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS ONE 3: e2802.
14. Chase MW, Cowan RS, Hollingsworth PM, van den Berg C, Madrin˜a´n S, et al. (2007) A proposal for a standardised protocol to barcode all land plants. Taxon 56: 295–299.
15. Daniel R,Stephen C,Kipling W, A Genomic Perspective on the Short comings of Mitochondrial DNA for "Barcoding" Identification. The American Genetic Association, 2006:97(6):581–594

16. Xiaohui Pang etal. Use of barcodes to identify gymnosperms、Pteridophytes、Angiosperm. (2010)
17. Lahaye R,van der Bank M,Bogarin D,etal. DNA barcoding the floras of biodiversity hotspots[J]Proceedings of the National Acadeiny ofSciences, USA, 2008, 105:2923-2928
18. Kress W J. Efickson D L. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region[J]. PloS ONE,2007,2(6):e508.
19. Fazekas A J, Burgess K S, Kesan akurti P R, et a1. Multiple multilocns DNA barcodes from the plastid genome discriminate plant species equally well[J]. PloS One,2008,3:e2802.
20. Sass C, Little D P,Stevenson D W, et a1. DNA barcoding in the cycadales :testing the potential of proposed barcoding markers for species identifcation ofcycads[J], PloS One,2007,2:e115.
21. Kress, W.J., Wurdack, K.J., Zimmer, E.A., Weigt, L.A. & Janzen, D.H. 2005. Use of DNA barcodes to identify flowering plants. Proc. Natl. Acad. Sci. U.S.A. 102: 8369–8374.
22. Shaw, J., Lickey, E.B., Schilling, E.E. & Small, R.L. 2007. Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms.The tortoise and the hare III. Amer. J. Bot. 94: 275–288.
23. Kress W J,Wurdack K J,Zimmer E A, et a1.Use of DNA barcodes to identify flowering plants[J]. Proceedings of the National Academy of Sciences,USA, 2005,102:8369-8374.
24. Stoeckle M. Taxonomy DNA and the barcode of life[J]. Bioscience,2003,53:2-3.
25. Chase, M.W., Salamin, N., Wilkinson, M., Dunwell, J.M., Kesanakurthi, R.P., Haider, N. & Savolainen, V. 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos. Trans., Ser. B 360: 1889–1895
26. Newmaster, S.G., Fazekas, A.J. & Ragupathy, S. 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Canad. J. Bot. 84: 335–341.
27. Newmaster S G, Fazekas A J, Steeves R A D, et a1. Testing candidate plant barcode regions in the Myfisticaceae[J]. Molecular Ecology Resources, 2008, 8: 480-490.
28. Ren BQ, Xiang XG, Chen ZD. Species identification of Alus(Betulaceae) using nrDNA and cpDNA genetic markers[J]. Mol Ecol Resource,2010, doi:10.1111 /j.1755-0988.2009.02815.x
29. Fitzhugh, K., 2006. DNA barcoding: an instance of technology-driven science BioScience 56 (6), 462–463.
30. Rubinoff, D., 2006b. DNA barcoding evolves into the familiar. Cons. Biol. 20, 1548–1549.
31. Balakrishnan, R., 2007. Species concepts, species boundaries and species identification: a view from the tropics. Syst. Biol. 54 (4), 689–693.
32. Miller, S.E., 2007. DNA barcoding and the renaissance of taxonomy. Proc. Natl. Acad.Sci. U.S.A. 104, 4775–4776.
33. Vogler, A.P., Monaghan, M.T., 2007. Recent advances in DNA taxonomy. J. Zool. Syst. Evol. Res. 45 (1), 1–10.
34. Sucher NJ, Carles MC (2008) Genome-based approaches to the authentication of medicinal plants. Planta Med 74: 603–623.
35. Yao H, Song JY, Ma XY, Liu C, Li Y, et al. (2009) Identification of Dendrobium species by a candidate DNA barcode sequence: the chloroplast psbA-trnH intergenic region. Planta Med 75: 667–669.
36. Song JY, Yao H, Li Y, Li XW, Lin YL, et al. (2009) Authentication of the family Polygonaceae in Chinese pharmacopoeia by DNA barcoding technique. J Ethnopharm 124: 434–439.
37. Shaw D, Leon C, Kolev S, Murray V (1997) Traditional remedies and food supplements: a 5-year toxicological study (1991-1995). Drug Saf 17: 342–356.

4/18/2011