Research in Plant Biology, 4(3): 29-35, 2014 ISSN : 2231-5101
www.resplantbiol.com

Mini Review

Research and application of DNA barcode in identification of plant species

Hongying Duan, Feifei Chen, Wenxiao Liu, Chune Zhou and Yanqing Zhou*

College of Life Science, Henan Normal University, Xinxiang, China
*Corresponding author email: yqzhou@htu.cn; dxdhy@126.com

DNA barcode, as a rapid and accurate technique for species discrimination with short DNA fragment, has been hot spot of biotaxonomy in the world, but there are still debates on which DNA region can be used as the standard barcode for land plants. In this article, advantages of DNA barcode, the application and prospect on DNA barcode in plant were described, and the technical process of plant DNA barcode was summarized. Furthermore, some suggestions in research of plant DNA barcode were put forward.

Keywords: species identification, DNA barcode, plant.

Discrimination and identification of species is primary step in taxonomy, and even in all biology research. At present, a variety of characters, such as color, shape, structure, etc. have been used to identify species since Carl Linnaeus conducted system classification of species, however development of taxonomy faces enormous challenges because of limitations in morphological identification and shrinking of morphological taxonomists. Along With development of molecular biology and bioinformatics, "DNA barcode" concept was put forward by Paul Hebert professor (Hebert *et al*., 2003a, 2003b), then the consortium for the barcode of life (CBOL) was set up in order to establish the standard, fast and low cost method for species identification.

Concept and advantage of DNA barcode

DNA barcode is a rapid and accurate technique for species discrimination with short DNA fragment, has been hot spot of biotaxonomy in the world (Sun *et al*., 2011), and has the following advantages (Ning *et al*., 2008; Liu *et al*., 2011a): (1) Many samples could be identified by DNA barcode, such as sample in different growth stages or with different shapes, the partially damaged sample, the processed sample, and so on, so DNA barcode distinctly expands the scope of identified samples. (2) Because of repeatability and high stability, DNA barcode is convenient for non-professional taxonomists to identify species, thus DNA barcode can effectively alleviate lack of talents in species identification by morphological classification. (3) DNA barcode is more accurate, species in genus and family or even most species in dozens of families can be accurately identified only by one or a few gene segments. (4) DNA barcodes

could be uniformly managed and shared by establishing DNA barcode database. Therefore, DNA barcode may be the biggest change in biotaxonomy since the binomial system of nomenclature was established by Carl Linnaeus, not only would complement traditional methods of species identification, but also possibly make evaluation of samples automate and standardize. In addition, the application system of DNA barcode is easily established in a shorter time, and is more utilizable (Ren and Chen, 2010), thus DNA barcode would have broad application prospects in life science, forensic science, ecology, biology, medicine, epidemiology, evolutionary biology, biogeography, conservation biology, food industry, and other fields.

Research of candidate DNA barcode in plant

Since mitochondrial cytochrome c oxidase subunit I (COI) was proposed to be as DNA barcode for animal discrimination by Hebert (Hebert *et al*., 2003a, 2003b), *COI* gene has been widely used to identify insects, fish, birds and other animals (Hebert *et al*., 2004a, 2004b; Ward *et al*., 2005; Kerr *et al*., 2007; Tavares and Baker, 2008), yet the evolution rate of *COI* gene in plant is far slower than that in animal, which makes *COI* gene only apply to identification of some algae (Saunders, 2005; Evans *et al*., 2007). Furthermore, totally different from genome evolution of animal, many plants could cross each other, which would cause larger difference between plants on the species level. Therefore, the standard sequence of plant DNA barcode should be further explored, many researchers tried to find DNA barcode from chloroplast genome and nuclear genome (Cho *et al*., 2004; Chase *et al*., 2005), and CBOL Plant Working Group (CBOL-PWG) suggested that combination of genes should be adopted. Theses proposed genes consist of coding genes and non-coding regions, in which coding genes are mainly *rpoB*, *rpoC1*, *matK*, *rbcL* and *UPA*, the non-coding regions are *trnH-psbA*, *atpF-atpH*, *psbK-psbI* and *ITS* (Kress *et al*., 2005; Kress and Erickson, 2007; Pennisi, 2007).

In order to obtain the universal DNA barcode of plants, many researchers explored actively in the larger scope of taxa, such as angiosperms, even the whole land plants, and put forward a variety of barcode fragments or their combination (Chen *et al*., 2009). In view of research results of DNA barcode in plant identification, CBOL-PWG recommended *rbcL* and *matK* as the universal DNA barcode in land plants (Hollingsworth *et al*., 2009). Whereas, there are some limitations in the above research, for example sampling is mainly in larger taxa, and is insufficient on the genus or species level, so the intraspecific variation would be underestimated. Furthermore, sister groups is not included in some analysis, which may make differences between species overvalued. Subsequently, some researchers evaluated candidate DNA fragments by dense sampling in different families or genera again (Sass *et al*., 2007; Newmaster *et al*., 2008; Zhang *et al*., 2009). In the third international conference of DNA barcode which was held in Mexico in 2009, CBOL-PWG decided *rbcL* and *matK* as the core barcode of plant DNA barcode, and suggested *trnH-psbA* and *ITS* as supplementary barcode of plant DNA barcode (Ren and Chen, 2010).

Screening of plant DNA barcode

Many researchers compared

and evaluated these candidate DNA barcodes, and found that combination of *ITS* and *trnH-psbA* was relatively more suitable to species identification in Birch alder genera, Mallow, Ginseng genus, Plum genera and other genera and families (Ren *et al*., 2010; Wang *et al*., 2011; Zuo *et al*., 2011; Yang *et al*., 2012). *ITS* alone or combined with *rbcL* and *matK* had very high resolution, and could identify species in Lysimachia (Zhang *et al*., 2012), simultaneously, the combination of *ITS* and *MatK* could also identify species in Orchids (Xiang *et al*., 2011). Furthermore, the applicability of *rbcL*, *matK*, *trnH-psbA*, *trnL-F* and *ITS* in Pacltitaxel was evaluated, but *ITS*, *trnL-F*, alone or their combination was suitable for species identification in Pacltitaxel (Liu *et al*., 2011b). Therefore, the above studies show that *ITS* alone or combined with other DNA fragments is applicable to identify plant species in a lot of genera and families.

In addition, Chen compared seven candidate DNA barcodes and analyzed 6600 *ITS2* sequences of 4800 plant species in 753 genera, and found the identification efficiency of *ITS2* on the species level was as high as 92.7% (Chen *et al*., 2010), moreover the identification ability of *ITS2* in Rosaceae, Leguminosae, Euphorbiaceae, Rutaceae, Paris linn, and Compositae was in the scope of 78% to 100% (Pang and Chen, 2009; Song *et al*., 2009; Luo *et al*., 2010; Gao *et al*., 2011). As a result, CBOL-PWG in China suggested that *ITS2* should be as candidate DNA barcode to identify species in the wider range of plant taxa (Li *et al*., 2011). Species taxonomy of some genera in Meteoriaceae was studied, and results showed that the identification effect of *ITS2* was the best, yet combination of plasmid gene and nuclear gene was more advantageous to identify species (Zhao *et al*., 2010), moreover the combination of *ITS2*, *matK* and *rbcL* also had higher identification efficiency (92%) in palms (Arecaceae) (Jeanson *et al*., 2011). However, some inconsistent results were also obtained, for example, ten kinds of molecular markers in moss identification on the family level were evaluated, but *ITS2* was not suitable for identification of moss plants (Liu *et al*., 2010), and each species in the neotropical genus Psiguria (Cucurbitaceae) had their own unique barcode (Steele *et al*., 2010).

In short, the same candidate DNA barcode or the same combination of candidate barcode presents different identification ability in different taxa, so selection and evaluation of DNA barcode in the wider range of plant taxa might be research focus of plant DNA barcode now.

Operation of plant DNA barcode technology

Obtaining of candidate barcode sequence

PCR primers of candidate DNA barcode could be designed on the basis of primers used in some studies (Kress *et al*., 2005; Chase *et al*., 2007; Song *et al*., 2009) or the related resources. In order to obtain target sequence by PCR amplification, it is necessary to optimize PCR reaction conditions and primer combination. The amplification efficiency and sequencing success rate of candidate barcode should be compared, and the evaluation of sequencing quality can be performed with Sequencher, CodonCode and others (Chen *et al*., 2009). The ease or complexity about sequence alignment of candidate barcode could be analyzed with Clustal W software (Newmaster *et al*., 2006), simultaneously target sequence should be aligned in Genbank database by BLAST to contrast reliability of sequence information.

In addition, the acquisition of plant material should base on traditional morphological taxonomy, as far as possible cover the whole area of species, and sampling in the population level should be relatively dense (Stoeckle and Hebert, 2008).

Screening of candidate barcode sequence

In order to explore application of DNA barcode in identification and classification of plant species, degree of interspecific and intraspecific variation and identification efficiency of candidate barcode need to be compared and evaluated. Generally, the genetic distance of candidate sequence in interspecies and intraspecies is calculated with Kimura-2-parameter distance (K2P) model which is suggested by CBOL (Meyer and Paulay, 2005; Lahaye *et al*., 2008), which would contribute to resolution and appropriateness of different candidate sequences. Furthermore, the intraspecific and interspecific differences of candidate barcode is tested with Wilcoxon signed rank tests (Lahaye *et al*., 2008), and the variation of candidate barcode in interspecies and intraspecies is analyzed by Taxon DNA software (Meyer and Paulay, 2005). Then, the pros and cons of candidate DNA barcode would be determined according to its resolution and suitability.

Evaluation on identification efficiency of candidate barcode sequence

The identification efficiency of candidate barcode is generally evaluated according to analysis of molecular evolutionary genetics, for example, the molecular phylogenetic tree established with MEGA4.0 or PAUP software would test whether different individuals of the same species can clustering together (Chen *et al*., 2009). Furthermore, characteristics of barcode sequence in the same species groups, such as GC content, variable site, information site and others, should be calculated to study special feature of candidate barcode sequence, which would contribute to explore application of DNA barcode in identification and classification of plant species.

In addition, DNA barcode sequence and its sequence diagram, species information, gathering information and others are submitted to DNA barcode database (Ratnasingham and Hebert, 2007), which would provide basis for species identification.

Prospects

DNA barcode is a rapid and accurate identification technology, its successful application in identification of plant species would be significant. However, there are some problems in screening and research of plant DNA barcode, (1) Because the phenomenon of hybridization between plants is widespread, it would be more difficult to obtain the universal DNA barcode in plants, and the identification of different species groups might need different DNA barcode. Furthermore, primers of some candidate barcodes have poor universality, even different primers might be suitable for different groups, thus it is necessary to explore primer combination of DNA barcode for different taxa. (2) Although the feasibility of DNA barcode in animal classification is constantly validated, lower difference between species in the same genus has cast doubt on taxonomic identification of species. In addition, species which is rapidly differentiated and hybrids would increase difficulty of species identification. At present, the biggest controversy in DNA barcode research is whether DNA barcode

could be applicable to these closely related species or rapidly differentiated species. (3) The accuracy of experimental materials is crucial because DNA barcode is based on traditional taxonomy, furthermore species identification would appear chaos if the impact of distribution on species is ignored.

Therefore, the identification ability of candidate barcode or their combination should be comprehensively compared and evaluated by plant taxonomy, molecular biology, bioinformatics and other methods, and the applicability of DNA barcode in classification of these closely related species should be further explored, which would provide theory basis for screening of plant DNA barcode and identification of plant species.

Acknowledgment

This research was kindly supported by Personnel Training United Fund of NSFC-Henan province (U1304304), in China.

References

Chase MW, Cowan RS, Hollingsworth PM. 2007. A proposal for a standardised protocol to barcode all land plants. Taxon. 56(2): 295-299.

Chase MW, Salamin N, Wilkinson M, Dunwell JM, Kesana-kurthi RP, Haidar N, Savolainen V. 2005. Land plants and DNA barcodes: short-term and long-term goals. Philos Trans R Soc Lond B Biol Sci. 360: 1889-1895.

Chen SL, Song JY, Yao H, Shi LC, Luo K, Han JP. 2009. Strategy and key technique of identification of chinese herbal medicine using DNA barcoding. Chin J Nat Med. 7 (5): 322-327.

Chen SL, Yao H, Han JP, Liu C, Song JY, Shi LC, Zhu YJ, Ma XY, Gao T, Pang XH, Luo K, Li Y, Li XW, Jia XC, Lin YL, Leon C. 2010. Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. 5 (1): e8613.

Cho Y, Mower JP, Qiu YL, Palmer JD. 2004. Mitochondrial substitution rates are extraordinarily elevated and variable in a genus of flowering plant. Proc Natl Acad Sci USA. 101: 17741-17746.

Evans KM, Wortley AH, Mann DG. 2007. An assessment of potential diatom ''barcode'' genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relation-ships in *Sellaphora* (Bacillariophyta). Protist. 158(3): 349-364.

Gao T, Sun Z, Yao H, Song J, Zhu Y, Ma X, Chen S. 2011. Identification of Fabaceae plants using the DNA barcode matK. Planta Med. 77 (1): 92-94.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003a. Biological identifications through DNA barcodes. Proc Biol Sci. 270(1512): 313-321.

Hebert PDN, RatnasinghamS, Dewaard JR. 2003b. Barcoding animal life: cytochrome coxidase subunit 1 divergences among closely related species. Proc Biol Sci. 270 (1): 96-99.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator. Proc Natl Acad Sci USA. 101 (41): 14812-14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004b. Identification of birds through DNA barcodes. PLoS Biol. 2 (10): e312.

Hollingsworth PM, Forrest LL, Spouge JL, Hajibabaei M, Ratnasingham S, and others in CBOL Plant Working Group. 2009. A DNA barcode for land plants. Proc Natl Acad Sci USA. 106(31):

12794-12797.

Jeanson ML, Labat JN, Little DP. 2011. DNA barcoding: a new tool for palm taxonomists?Ann Bot. 108(8): 1445-1451.

Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PD. 2007. Comprehensive DNA barcode coverage of North American birds. Mol Ecol Notes. 7(4): 535-543.

Kress WJ, Erickson DL. 2007. A two-locus global DNA barcode for land plants: the coding rbcL gene complements the non-coding trnH-psbA spacer region. PLoS One. 2 (6): e508.

Kress WJ, Wurdack KJ, Zimmer EA, Weigt LA, Janzen DH. 2005. Use of DNA barcodes to identify flowering plants. Proc Natl Acad Sci USA. 102 (23): 8369-8374.

Lahaye R, Van der Bank M, Bogarin D, Warner J, Pupulin F, Gigot G, Maurin O, Duthoit S, Barraclough TG, Savolainen V. 2008. DNA barcoding the floras of biodiversity hotspots. Proc Natl Acad Sci USA. 105 (8): 2923-2928.

Li DZ, Gao LM, Li HT, Wang H, Ge XJ, Liu JQ, and others in China Plant BOL Group. 2011. Comparative analysis of a large dataset indicates that internal transcribed spacer (ITS) should be incorporated into the core barcode for seed plants. 108(49): 19641-19646.

Liu J, Möller M, Gao LM, Zhang DQ, Li DZ. 2011b. DNA barcoding for the discrimination of Eurasian yews (Taxus L., Taxaceae) and the discovery of cryptic species. Mol Ecol Resour. 11(1): 89-100.

Liu Y, Yan HF, Cao T, Ge XJ. 2010. Evaluation of ten plant barcodes in Bryophyta (Mosses). J Syst Evol. 48: 36-46.

Liu YJ, Liu Y, Huang YJ, Long CL. 2011a. Progress and application of DNA barcoding technique in plants. J Plant Resources Environment. 20(1): 74-82, 93.

Luo K, Chen SL, Chen KL, Song JY, Yao H, Ma XY, Zhu YJ, Pang XH, Yu H, Li XW, Liu Z. 2010. Assessment of candidate plant DNA barcodes using the Rutaceae family. Sci China Life Sci. 53(6): 701-708.

Meyer CP, Paulay G. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biol. 3(12): e422.

Newmaster SG, Fazekas AJ, Ragupathy S. 2006. DNA barcoding in land plants: evaluation of rbcL in a multigene tiered approach. Can J Bot. 84: 5-341.

Newmaster SG, Fazekas AJ, Steevesr AD, Janovec J. 2008. Testing candidate plant barcode regions in the Myristicaceae. Mol Ecol Resour. 8(3): 480-490.

Ning SP, Yan HF, Hao G, Ge XJ. 2008. Current advances of DNA barcoding study in plants. Biodiversity Science. 16 (5): 417-425.

Pang X, Chen SL. 2009. Using DNA barcodes to identify Rosaceae. Planta Med. 75(4): 417.

Pennisi E. 2007. Taxonomy. Wanted: a barcode for plants. Science. 318 (5848): 190-191.

Ratnasingham S, Hebert PDN. 2007. Bold: the barcode of life data system (www.barcodinglife.org). Molecular Ecology Notes. 7: 355-364.

Ren BQ, Chen ZD. 2010. DNA barcoding plant life. Chin Bulletin Botany. 45 (1): 1-12.

Ren BQ, Xiang XG, Chen ZD. 2010. Species identification of Alnus (Betulaceae) using nrDNA and cpDNA genetic markers. Mol Ecol Resour. 10(4): 594-605.

Sass C, Little DP, Stevenson DW, Specht CD. 2007. DNA barcoding in the

Cycadales: testing the potential of proposed barcoding markers for species identification of Cycads. PLoS One. 2(11): e1154.

Saunders GW. 2005. Applying DNA barcoding to red macroalgae: a preliminary appraisal holds promise for future applications. Philos Trans R Soc Lond B Biol Sci. 360(1462): 1879-1888.

Song JY, 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. J Ethnopharmacol. 124 (3): 434-439.

Steele PR, Friar LM, Gilbert LE, Jansen RK. 2010. Molecular systematics of the neotropical genus Psiguria (Cucurbitaceae): implications for phylogeny and species identification. Am J Bot. 97(1): 156-173.

Stoeckle MY, Hebert PDN. 2008. Bar code of life: DNA tags help classify animals: inspired by commercial bar codes, DNA tags could provide a quick, inexpensive way to iden-tify species. Sci Am. 299 (4): 82-88.

Sun Y, Hao HC, Xue CY, Li DZ. 2011. DNA Barcoding for Plant: chances always go with challenge. J MUC(Nat Sci Edi.). 20(2): 71-75.

Tavares ES, Baker AJ. 2008. Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evol Biol. 8: 81-81.

Wang K, Chen KJ, Liu Z, Chen SL. 2011. Screening of universal DNA Barcodes for Malvaceae plants. Chin Bulletin Botany. 46(3): 276-284.

Ward RD, Zemlak TS, Innes BH, Last PR, Hebert PD. 2005. DNA barcoding Australia's fish species. Philos Trans R Soc Lond B Biol Sci. 360(1462): 1847-1857.

Xiang XG, Hu H, Wang W, Jin XH. 2011. DNA barcoding of the recently evolved genus Holcoglossum (Orchidaceae: Aeridinae): a test of DNA barcode candidates. Mol Ecol Resour. 11(6): 1012-21.

Yang JB, Wang YP, Möller M, Gao LM, Wu D. 2012. Applying plant DNA barcodes to identify species of Parnassia (Parnassiaceae). Mol Ecol Resour. 12(2): 267-275.

Zhang CY, Wang FY, Yan HF, Hao G, Hu CM, Ge XJ. 2012. Testing DNA barcoding in closely related groups of Lysimachia L. (Myrsinaceae). Mol Ecol Resour. 12(1): 98-108.

Zhang JM, Wang JX, Xia T, Zhou SL. 2009. DNA barcoding: species delimitation in tree peonies. Sci China C Life Sci. 52(6) : 568-578.

Zhao LJ, Jia Y, Zhou SL, Du GL. 2010. The Prel iminary study on DNA barcoding of Mosses - a case of part of genera of Meteoriaceae. Acta Botanica Yunnanica. 32(3): 239-249.

Zuo Y, Chen Z, Kondo K, Funamoto T, Wen J, Zhou S. 2011. DNA barcoding of Panax species. Planta Med. 77(2): 182-187.

(Received: 24.4.2014; Accepted : 20.6.2014)