

日本産樹木DNAバーコーディングの現状(2010)

吉村研介・鈴木節子(森林総研)・田中孝尚・鈴木三男(東北大)・神保宇嗣・伊藤元己(東大)・舘田英典(九大)・大谷雅人・勝木俊雄・津村義彦・藤井智之・能城修一・河原孝行・吉丸博志(森林総研)

要旨 : DNA バーコーディングとは、特定の遺伝子領域の短い塩基配列(DNA バーコード)で生物種の同定を行う方法である。2003年に動物分野でミトコンドリア DNA の遺伝子領域 *COI* を対象とし、植物では、当初、葉緑体 DNA の *rbcL* 部分配列と *trnH-psbA* 遺伝子間領域の2領域が提唱されたが、2009年11月に *rbcL* と *matK* 部分配列を植物における DNA バーコードの標準とすることが決められた。筆者らのグループは日本産樹木を対象とする DNA バーコードの開発を、*rbcL* 部分配列、*matK* 部分配列、*trnH-psbA* 遺伝子間領域での解析を行っている。*rbcL* 部分配列は、変異性が低く植物の科や属レベルの分類に有効であった。*matK* 部分配列は *rbcL* 部分配列より変異性は高いが、*trnH-psbA* 遺伝子間領域に比べ種分解能が劣っていた。*rbcL* と *matK* 部分配列は、植物の DNA バーコーディングの基礎として好ましいが、それだけでは分類同定困難な種分類群が存在する。分類同定困難な種分類群については葉緑体 DNA の調査領域を増やすか、ITS 領域等の核 DNA を調べる必要がある。将来的に「証拠標本」「DNA バーコード」の2つのデータベースの構築を進め、同定支援システムと併せて、公開する予定である。
キーワード : DNA バーコーディング, 日本産樹木, *rbcL*, *matK*, *trnH-psbA* 遺伝子間領域

Abstract : DNA barcoding is a taxonomic method that uses a short genetic marker in an organism's DNA. The cytochrome oxidase 1(CO1) mitochondrial gene used in animals. In plants, Consortium for the Barcode of Life(CBOL) Plant Working group recommended a core-barcode consisting two plastid regions, *rbcL* partial sequence and *matK* partial sequence at 2009. We have been collecting herbarium specimens and DNA samples on these woody plant species. DNA sequences are analyzed on *rbcL*, *matK* and *trnH-psbA* as targets of DNA barcoding. The *trnH-psbA* intergenetic spacer region has the most variable information, and *rbcL* has the least. In contrast, the barcode region of *rbcL* is easy PCR amplification and sequence. The combination of *rbcL* + *matK* is proper backbone to the core-barcode dataset, but low discrimination power is far from perfect. It is necessary to using supplementary barcodes that other plastid region or nuclear region.

Keywords : DNA barcoding, Japanese woody plant, *rbcL*, *matK*, *trnH-psbA* spacer

I はじめに

DNA バーコーディングとは、特定の遺伝子領域の短い塩基配列(DNA バーコード)で生物種の同定を行う方法である。2003年に動物において、ミトコンドリア DNA のチトクローム c オキシダーゼサブユニット I 遺伝子(*COI*)の部分配列(648bp)を使うことが、Hebert らによって提唱された(5)。短い DNA 領域ではあるが、全ての種

で単一の領域についての情報を明らかにすることにより、これまで形態が似ていて同一種と同定されていた種が、遺伝学的には別種(隠ぺい種)であることが判明したりしている。また、種同定の特徴となる形態が発現していない時期の個体や種同定不能な部分からでも、種同定が可能になる等の利点がある。全ての生物種のバーコーディングを行うことを目的に、2004年に Consortium

Kensuke YOSHIMURA, Suzuki Setsuko(For. and Forest Prod. Res. Ins., Ibaraki 305-8687), Takahisa Tanaka, Mitsuo Suzuki(Tohoku Univ., Sendai 980-0862), Utsugi Jinbo, Motomi Ito(Univ. of Tokyo, Meguro-ku 153-8902), Hidenori Tachida(Kyushu Univ., Fukuoka 810-8560), Masato Ohtani, Toshio Katsuki, Yoshihiko Tsumura, Tomoyuki Fujii, Shuichi Noshiro, Takayuki Kawahara, Hiroshi Yoshimaru(FFPRI, Ibaraki 305-8687)
 DNA barcoding on Japanese woody plants(2010)

for the Barcode of Life (CBOL)が結成され各種のプロジェクトが世界的に進められている。

植物においては、2005年にKressらが、ミトコンドリアDNAのCOI領域は保存性が高く変異が少ないことから、葉緑体の*rbcL*の部分配列(約700bp)、*trnH-psbA*遺伝子間領域を使うことを提唱した(8)。2007年第2回国際Barcode of Life (BOL)会議では、情報量不足を補うためさらに複数の領域を使う必要があることが複数報告された(6)。CBOL Plant Working Group(2009)は、葉緑体の4つの遺伝子領域と3つの遺伝子間領域を比較検討した結果、植物では*rbcL*部分配列と*matK*部分配列を第一のターゲットとして使うことを提唱し(2)、2009年11月第3回国際BOL会議で、この2つの領域が、植物のDNAバーコードとして国際標準となった(7)。

日本においては、植物全体について組織だったDNAバーコーディングは行われていない。そのため、筆者らのグループは日本産樹木を対象にDNAバーコーディングを2008年より進めている。

II 材料と方法

筆者らの研究グループの内、東北大学と森林総研木材特性研究領域は、1998年より国内樹種の木材標本・押し葉試料を、全国各地で28か所で採集をしており、これを主な試料とした。2008年より、これらのDNA抽出を開始し、その他の収集した試料も併せて、DNA解析を進めている。これまでのところ137科、369属、979種、1145種・亜種・変種、6169個体のDNA抽出を行った。過去に収集したDNA用試料は、シリカゲルで乾燥し常温保存していた。そのためDNAが劣化してDNA解析に使えない試料もあった。2008年以降は、新鮮なサンプルを冷凍保存し随時DNA抽出するようにしている。DNA試料は、当初2007年の情報に基づき*rbcL*部分配列、*trnH-psbA*遺伝子間領域について解析を進めた。その後、*rbcL*部分配列と*matK*部分配列が国際標準として決められたので、*matK*部分配列の解析も進めている。

III 結果と考察

一部のサンプルがシーケンスまで至らず欠測値となったため、それらのサンプルを除いた集計値を表-1にまとめた。*rbcL*部分配列は、129科、350属、861種、5141個体シーケンスが明らかになった。*rbcL*部分配列は、PCRの増幅効率が良く、良質なDNAならほとんど全て増幅することが可能であった。DNA抽出を行う前の保存状況が悪い試料があったため、供試個体全体に対するシーケンスの成功率は全体で83%であった。

*matK*部分配列は、91科、231属、556種、2282個体のシーケンスを明らかにした。現在使用しているプラ

イマーは増幅効率が悪く、比較的質の良いDNAでも66%程度の増幅効率であった。質の悪いDNAだと極端に増幅効率が落ちる傾向があった。PCRの増幅に成功したものは、シーケンスは比較的順調であった。また、現在のプライマーでは裸子植物等は増幅できない。

*trnH-psbA*遺伝子間領域は、92科、255属、612種、3032個体のシーケンスを明らかにした。PCRの増幅効率は86%と*rbcL*部分配列より若干落ちる程度であったが、シーケンスの成功率は、全体で70%と*rbcL*と比べて低かった。この遺伝子間領域は、polyAやpolyTの単塩基反復配列が多く含まれ、PCRの増幅が見られても、スリッピングなどのため良好なシーケンスが得られない場合がある。

表-2に種同定能力を示した。*rbcL*部分配列では、861種のうち400種(46%)が、種同定に成功していたが、半数以上の種で近縁種と同一シーケンスがあり同定不能であった。*rbcL*は比較的保存性が高いこと、対象とした塩基配列が599bpと短いことのため、属レベルの同定には非常に有効であるが、分類群によっては種レベルの同定が不可能な場合が多い。

*matK*部分配列は、556種のうち335種(60%)が、種同定可能であった。この領域は、遺伝子領域であるにもかかわらず変異性が高く、系統関係の研究でよく利用されているが、今回使用した樹種群では種同定能はあまり高くなかった。また、この領域は変異性が高すぎて、単一のプライマー組では、増幅できないため、簡易にDNAシーケンスを行うDNAバーコードとしては欠点がある。しかしながら、プライマーに関しては現在、研究が盛んに行われており(4,9)、今後の改善が期待される。

*trnH-psbA*遺伝子間領域は、612種のうち453種(74%)が、種同定可能であった。この領域は、変異性は高いが塩基配列の挿入・欠失が多く、長さが145bpから700bp以上と多岐にわたっていた。長さが700bp以上だと1回のシーケンス操作で全て読み取れないためデータを取るのが困難である。変異性が高すぎて種内多型が数多く見られるため、DNAバーコーディングによる種同定を目的とした場合に、誤同定の可能性が高くなる。

2つの配列領域を併せて解析した結果では、*rbcL+trnH-psbA*の組み合わせが608種のうち464種(76%)と、*matK+trnH-psbA*の組み合わせが410種のうち310種(76%)がそれぞれ高い同定率を示した。3領域を併せて解析した結果では、410種のうち311種(76%)が同定可能であり、*matK+trnH-psbA*の組み合わせとほとんど変わらなかった。

本研究では、国際標準である*rbcL+matK*の組み合わ

せは、551種のうち369種(63%)と種同定能が低かった。CBOLが国際標準と決めた際、550種で72%の種同定能があるとされ(2)、カナダで植物種全体269属436種で調べた結果では、93%の種同定能を示している(1)。本研究の場合、種同定能が低かったのは、日本産木本植物のみを対象としているため、近縁種が比較的多く含まれているためではないかと思われる。しかしながら、*rbcL*部分配列は、全植物種において安定的にPCR増幅、シーケンスが可能であり、属レベルの同定に有効で有ること。*matK*部分配列は、PCR増幅に現在のところ困難な場合があるが、葉緑体DNAの遺伝子領域の中では、もっとも変異があること。*matK+trnH-psbA*遺伝子間領域は、良好なシーケンスを取ることが困難な場合があることにより、*rbcL+matK*の組み合わせは、妥当な選択であるといえる。

葉緑体DNAでは、種分化が被子植物では母親から、裸子植物では父親から遺伝するため、浸透交雑が進んでいる種群では同定できない。また、葉緑体DNAは、核DNAに比べると保存性が高いため種分化が進んでいない分類群では情報量が少なく種同定には十分でないことが以前から指摘されている(6)。

葉緑体DNAを用いた植物DNAバーコードが、種同定にそぐわない例としてサクラ属(*Cerasus*)の4種1変種の*rbcL*、*matK*の部分配列でNJ法を用いた系統樹を図-1に示した。この系統樹の場合、サクラ属は大きく3つのクラスタに分かれたが、各クラスタ全てに、4種1変種が属してしまい、種内変異はあるが、属内で同じ種内変異を共有しているため、種の分類同定には使用できないことが判る。試料数を増やし、種数を増やしても、日本国内でのサクラ属は、個体変異は見られるが、種間、地域間で傾向が見られなかった。このような分類群では、種同定のためには核DNAなどの情報をあわせて解析す

る必要が出てくる。現在、核DNAのITS2などの領域等が、補助的なDNAバーコーディング領域候補としてあげられている(2,6)。

一方、図に示さないが、主に樹形で変種レベルに分類されるカヤ(*Torreya nucifera*)、チャボガヤ(*Torreya nucifera v. radicans*)は、*rbcL*、*trnH-psbA*ともに明確に区別された。一方、同じく樹形を主な区別点としているイヌガヤ(*Cephalotaxus harringtonia*)、ハイイヌガヤ(*Cephalotaxus harringtonia v. nana*)については、区別できるような変異はDNAバーコードでは見られなかった。チャボガヤとカヤは、葉の形態などでは同定が不可能なため、植物DNAバーコードが有効であるといえる。

IV おわりに

2009年11月にCBOL Plant Working Groupが、植物におけるDNAバーコード領域として*rbcL*部分配列と*matK*部分配列の2つの領域を標準とし、それで足りない部分を、*trnH-psbA*遺伝子間領域、核DNAのITS領域などで補うことを決定している(2)。

補助的な領域については、まだ決定されておらず、現在探索が行われている最中である。

DNAバーコードは、証拠標本とDNA試料がセットとして保存され、データベース化されることにより、同定等に疑問が生じたときに、いつでも遡って検証し直すことができる利点がある。

本研究を行っている最中でも、DNAシーケンスの結果で、同定を誤っていたことが判明した試料が、数は少ないが存在した。同定者の熟練度や、花や果実のついていない個体、開葉が進んでない個体等、サンプルの採取時期、採取部位等の問題で、同定が困難であったり、同定を間違えることは起こりがちであり、DNAが抽出できさえすれば、同定が可能であるDNAバーコードは有効なツールになりうる。

表-1. 個体別、種別のPCR増幅、シーケンス成功率

Table 1. PCR amplification and sequence success of three plastid markers

配列領域	供試個体数	PCR増幅個体数	(%)	シーケンス	(%)
<i>rbcL</i>	6169	5663	91.8	5141	83.3
<i>matK</i>	3693	2453	66.4	2282	61.8
<i>trnH-psbA</i>	4348	3744	86.1	3032	69.8
	供試種数				
<i>rbcL</i>	979	951	97.1	863	88.2
<i>matK</i>	787	596	75.7	556	70.7
<i>trnH-psbA</i>	707	657	90.0	612	75.7

表-2. 配列領域による種同定の確率			
Table 2. Discrimination success for three markers and locus combination			
	種同定 可能	種同定 不能	同定率 (%)
<i>rbcL</i> + <i>matK</i> + <i>trnH-psbA</i>	311	99	75.85
<i>matK</i> + <i>trnH-psbA</i>	310	100	75.61
<i>rbcL</i> + <i>trnH-psbA</i>	464	144	76.32
<i>rbcL</i> + <i>matK</i>	349	202	63.34
<i>trnH-psbA</i>	453	159	74.02
<i>matK</i>	335	221	60.25
<i>rbcL</i>	400	461	46.46

引用文献

- (1) BURGESS, K.S., FAZEKAS, A.J., KESANAKURTI, P.R., GRAHAM, S.W., HUSBAND, B.C., NEWMASER, S.G., PERCY, D.M., HAJIBABAEI, M., BARRETT, S.C.H. (2011) Discriminating plant species in a local temperate flora using the *rbcL*+*matK* DNA barcode. *Meth. Ecol. Evol.* 2:333-340
- (2) CBOL PLANT WORKING GROUP (2009) A DNA barcode for land plants. *PNAS* 106:12794-12797.
- (3) 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. (2011) Validation of the ITS2 Region as a Novel DNA Barcode for Identifying Medicinal Plant Species. *PLoS One* 5: e8613
- (4) DUNNING, L.T., SAVOLAINEN, V. (2010) Broad-scale amplification of *matK* for DNA barcoding plants, a technical note. *Bot. J. Linn. Soc.* 164:1-9.
- (5) HEBERT, P.D.N., CYWINSKA, A., BALL, S.L., DEWAARD, J.R. (2003) Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B. Biol. Sci.* 270: 313-321.
- (6) HOLLINGSWORTH, P.M. (2008) DNA barcoding plants in biodiversity hot spots: Progress and outstanding questions. *Heredity* 101:1-2.
- (7) HOLLINGSWORTH, P. M., GRAHAM, S. W., Little, D.

P. (2011) Choosing and Using a Plant DNA Barcode. *PLoS One*. 6: e19524

(8) KRESS, W.J., ERICKSON, D.L. (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: e508.

(9) LI, Y., GAO, L., POUDEL, R.C., LI, D., FORREST, A. (2011) High universality of *matK* primers for barcoding gymnosperms. *J. of Systematics and Evolution* 49 :169-175 .

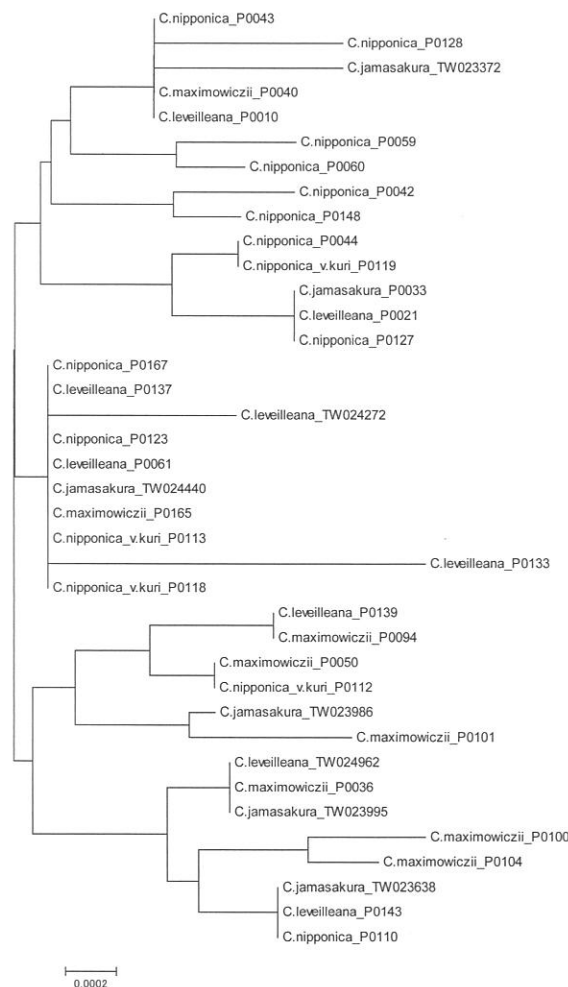


図-1. サクラ属 (*Cerasus*) の *rbcL* と *matK* の系統樹
 Fig.1. A taxon identification tree generated using neighbor-joining analysis for 4 *Cerasus* species from *rbcL* and *matK* partial sequences
C. jamasakura (ヤマザクラ), *C. leveilleana* (カスミザクラ), *C. maximowiczii* (ミヤマザクラ), *C. nipponica* (タカネザクラ), *C. nipponica v. kurilensis* (チシマザクラ)