

系統解析における *matK* 塩基配列の有効性

— バラ科サクラ亜科の解析 —

遠藤 康弘

(東京大学総合研究資料館)

はじめに

DNA の塩基配列による系統解析は、制限酵素断片長多型解析 RFLP のように DNA の変異を間接的に知る方法と比べて費用はかかるものの、ひとたび塩基配列を決定すれば情報を蓄積することができる。そのため解析したい分類群を同時に全て集める必要がないなど、利点が多い。たとえば多くの研究者が協力すれば、数千という分類群の系統を解析することも可能である。実際、上位分類群を対象にしてリブ्रोースビスリン酸カルボキシラーゼの大サブユニットをコードする遺伝子 *rbcL* の塩基配列を用いた系統解析の例がある (Chase *et al.*, 1993)。これから初めて系統解析を行う者にとっては、1300を超える分類群について *rbcL* の塩基配列が DNA のデータバンクに登録されているので、解析したい分類群の大まかな系統的位置が簡単にわかる。そのため系統解析において *rbcL* は標準的な遺伝子となっている。

陸上植物では *rbcL* の塩基配列には挿入や欠損などがなく、塩基配列の位置合わせ (アライメント alignment) や塩基配列の編集 (通常、全自動塩基配列決定装置で決定される塩基配列は完全でないので既知の塩基配列と比較・編集しなければならない) が容易である反面、近縁種間では変異が非常に少なく、分類群によっては亜科や科のレベルでも十分な解像度が得られないことがある (Chase *et al.*, 1993; Soltis *et al.*, 1995)。そこでより近縁な種群の系統を調べるために *rbcL* より進化速度の速い、すなわち変化の多い遺伝子が多く研究者によって探索されており、その候補の一つとして *matK* が最近注目されている。

matK は葉緑体ゲノム中の *trnK* 遺伝子の中にある蛋白質をコードしていると考えられる領域 open reading frame (ORF) であり、*trnK* 遺伝子の約2500塩基対のイントロンの中で約1500塩基対 (約500アミノ酸) を占めている (Sugita *et al.*, 1985)。この ORF は、自分自身を含むイントロンを切り取る マチュラーゼ maturase という酵素をコードしていることがわかり、*matK* とよばれることになった (Neuhaus & Link, 1987)。この遺伝子の進化速度は *rbcL* より 2-3 倍程度速いことが知られており、近縁種間の系統解析に

有効な遺伝子と考えられている。Johnson & Soltis (1994) は狭義のユキノシタ科 Saxifragaceae s.s. の系統解析に、Steele & Vilgalys (1994) はハナシノブ科 Polemoniaceae の属間の系統解析にこの遺伝子の塩基配列を用いている。また現在ウキクサ科 Lemnaceae (Les *et al.*, 1994)、ウコギ科 Araliaceae、セリ科 Apiaceae (Plunkett *et al.*, 1994) などの系統解析に用いられ始めているものの、まだ *matK* の解析が *rbcL* より優れている点を確認する十分な研究が行われているとはいえない。本稿ではバラ科サクラ亜科を解析例として *matK* と *rbcL* で構築した分岐図の比較を行うことによって系統解析における *matK* の塩基配列の有効性について考察する。

バラ科の果実の多様性

バラ科 Rosaceae は種子植物の中で花の形態や果実の形態について最も多様性の高い科の1つである。バラ状果 cynarrodium、ナシ状果 pome、核果 drupe、さく果 capsule、袋果 follicle、そう果 achene などの果実の形態は系統を反映しているとしてバラ科における重要な分類形質とされている (Cronquist, 1981; Hutchinson, 1964)。その中で、サクラ亜科 Prunoideae は核果を持つという形質でまとめられているグループである。しかし亜科が違っていてもヤマブキ連 Kerrieae やシロヤマブキ連 Rhodotypeae など、従来そう果であるといわれてきたものの、内果皮が硬壁化するという核果状の果実を持つものがある。サクラ亜科の果実の形態の多様性を理解するためには系統関係を基礎にした比較形態学的研究が必要となる。

解像度や信頼性の問題のみならず形態の多様性を議論する上で形態形質に基づく分岐図を用いることは循環論に陥るため、形態形質を用いない分子形質に基づく分岐図を用いることによって真の「比較」形態学的研究が可能になると考えられる。

バラ科サクラ亜科における系統解析的研究

バラ科全体の系統学的研究には Kalkman (1988) の形態形質を用いた研究がある。Kalkman はバラ科を連を単位とした22のグループに分け、14の形質を取り上げて分岐分類学的解析を行ったが、彼は原始形質 plesiomorphy と派生形質 apomorphy を先に決めて解析しており、また解析自身にも最大節約樹を完全に数え上げていないなどの欠点がみられる。彼と同じ形質のマトリックスを使い、厳密合意樹 strict consensus tree を計算したところ、ほとんどの枝が多岐になっており、互いの系統関係は議論できないことがわかった。DNA を用いた研究では Morgan *et al.* (1994) の *rbcL* の塩基配列を用いた系統解析の例がある。しかし彼らの解析ではサクラ亜科については5種しか扱っておらず、樹長や消失指数 decay index (Bremer, 1988) が小さいのでサクラ亜科の系統関係は全くわからない。これは *rbcL* では塩基情報が十分多くないことを示しており、この問題を解決するにはより多くの情報を他の遺伝子配列に求めなければならない。

材料—解析の範囲

サクラ亜科とその周辺属の解析をするためにサクラ亜科については各属から1種以上をとり上げ、また周辺属では Moragn *et al.* (1994) の分岐図を参考にコゴメウツギ *Stephanandra incisa* (Thunb.) Zabel, ヤマブキ *Kerria japonica* (L.) DC., シロヤマブキ *Rhodotypos scandence* (Thunb.) Makino, クサボケ *Chaenomeles japonica* (Thunb.) Lindl. ex Spach を選定し、外群としてはノイバラ *Rosa multiflora* Thunb. とミツバツチグリ *Potentilla freyniana* Bornm. を選択した。サクラ属 *Prunus* は、たいへん大きな属であるため、ふつう、ウメ亜属

subgen. *Armeniaca*, スモモ亜属 subgen. *Prunus*, モモ亜属 subgen. *Amygdalus*, サクラ亜属 subgen. *Cerasus*, ウワミズザクラ亜属 subgen. *Padus*, バクチノキ亜属 subgen. *Laurocerasus* に分類される (大橋ら, 1989)。今回の解析ではこの6亜属のうちから4亜属 (モモ亜属, サクラ亜属, ウワズミザクラ亜属, バクチノキ亜属) をとり上げ解析した。セイヨウバクチノキ *Prunus laurocerasus* L. の *rbcL* の塩基配列は GenBank から引用した。またバクチノキ *Prunus zippeliana* Miq. では *matK* の配列のみを決定し、残りの18種では *rbcL*, *matK* と同じ個体を用いて配列を決定した。

プライマーの配列と位置

matK 遺伝子を増幅するプライマー primer の塩基配列は Johnson & Soltis (1994) や Steele & Vilgalys (1994) にも示されている。しかしそのプライマーではバラ科で *matK* の断片をうまく増幅させることができなかったため、タバコ *Nicotiana tabacum* L. (Shinozaki, *et al.*, 1986) とイネ *Oryza sativa* L. (Hiratsuka, *et al.*, 1989) の *matK* の塩基配列を比較し、タバコの塩基配列の保存的な部分に基づいて新たに PCR 用のプライマーを設計した (表1)。*matK* はコード領域が長いので、その増幅には中間部の配列を持ったインターナルプライマーを必要とする。インターナルプライマーとしての MR と MF は、BF と R のプライマーで増幅された約900塩基対の断片 (図1, c) を低い電圧で緩やかに泳動してバラ科の数種において塩基配列を調べ、保存的な部分を用いてプライマーを設計した。AF, BF, R, 8R の4種類のプライマーを用いて AF-R, AF-8R, BF-R, BF-8R の4通りの組み合わせで *matK* の断片の増幅を試みたところ、バラ科, キク科 Asteraceae, ゴマノハグサ科 Scrophulariaceae, イネ, イチジク属 *Ficus*, ネコノメソウ属 *Chryso-*

表1. *matK* 遺伝子の増幅に用いたプライマーの塩基配列(Ooi *et al.* 1995 ; 遠藤 未発表).

| プライマー | 5'—3' | 設計 |
|-----------------|----------------------------|-------------------|
| <i>matK</i> -AF | CTATATCCACTTATCTTTTCAGGAGT | Ooi <i>et al.</i> |
| <i>matK</i> -BF | TCAGAGGGATTTGCGTTTATTGTGG | Ooi <i>et al.</i> |
| <i>matK</i> -MF | ACGATTATTCTTGCTCCTATATAA | 遠藤 |
| <i>matK</i> -MR | CCAGAAGAGGTTAATCGTAAATGA | 遠藤 |
| <i>matK</i> -R | CTGCATATACGCCCAAATCGGTCAA | Ooi <i>et al.</i> |
| <i>matK</i> -8R | AAAGTTCTAGCACAAGAAAGTCGA | Ooi <i>et al.</i> |

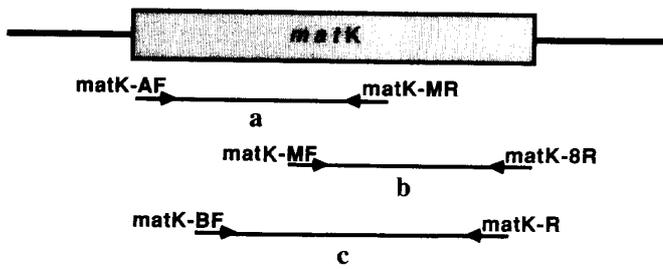


図1. *matK* 遺伝子とプライマーの位置関係 (遠藤 未発表).
a, b, cについては本文参照

plenium などで少なくとも1種類の組み合わせで *matK* の断片が増幅されることが確認された。また、AF-MR (図1, a), MF-8R (図1, b) の組み合わせではゴマノハグサ科, バラ科では *matK* の断片が問題なく増幅された。今回の解析ではこの AF-MR, MF-8R の組み合わせを用いて全自動塩基配列決定装置によって直接塩基配列を決定した。新たに *matK* で系統を解析しようとするときのプライマーの適用上の問題

点は, *matK* 遺伝子の両端のプライマー (AF, BF, R, 8R) では比較的多くの分類群で良好な増幅が行えるが, インターナルプライマー (MF, MR) の適用範囲が狭いということである。このため対象となる分類群によってはインターナルプライマーを設計し直さなければならない場合がある。

結 果

図2に *matK* と *rbcL* で構築したサクラ亜科とその周辺属の分岐図を示す。 *rbcL* の塩基配列には挿入・欠失はなく, *matK* の塩基配列には挿入・欠失があったが, 挿入・欠失は今回解析した19種では8箇所みられ, アライメントを行う上でとくに問題とはならなかった。また系統解析は挿入・欠失の箇所を除いてから PAUP (Swofford, 1993) を用いて行った。分岐図の数字はブートストラップ bootstrap 法 (Felsenstein, 1985) を用いて100回反復する事によって得られた再現確率を示している。 *matK* の分岐図は CI (一致指数 consistency index) = 0.731, RC (修正一致指数

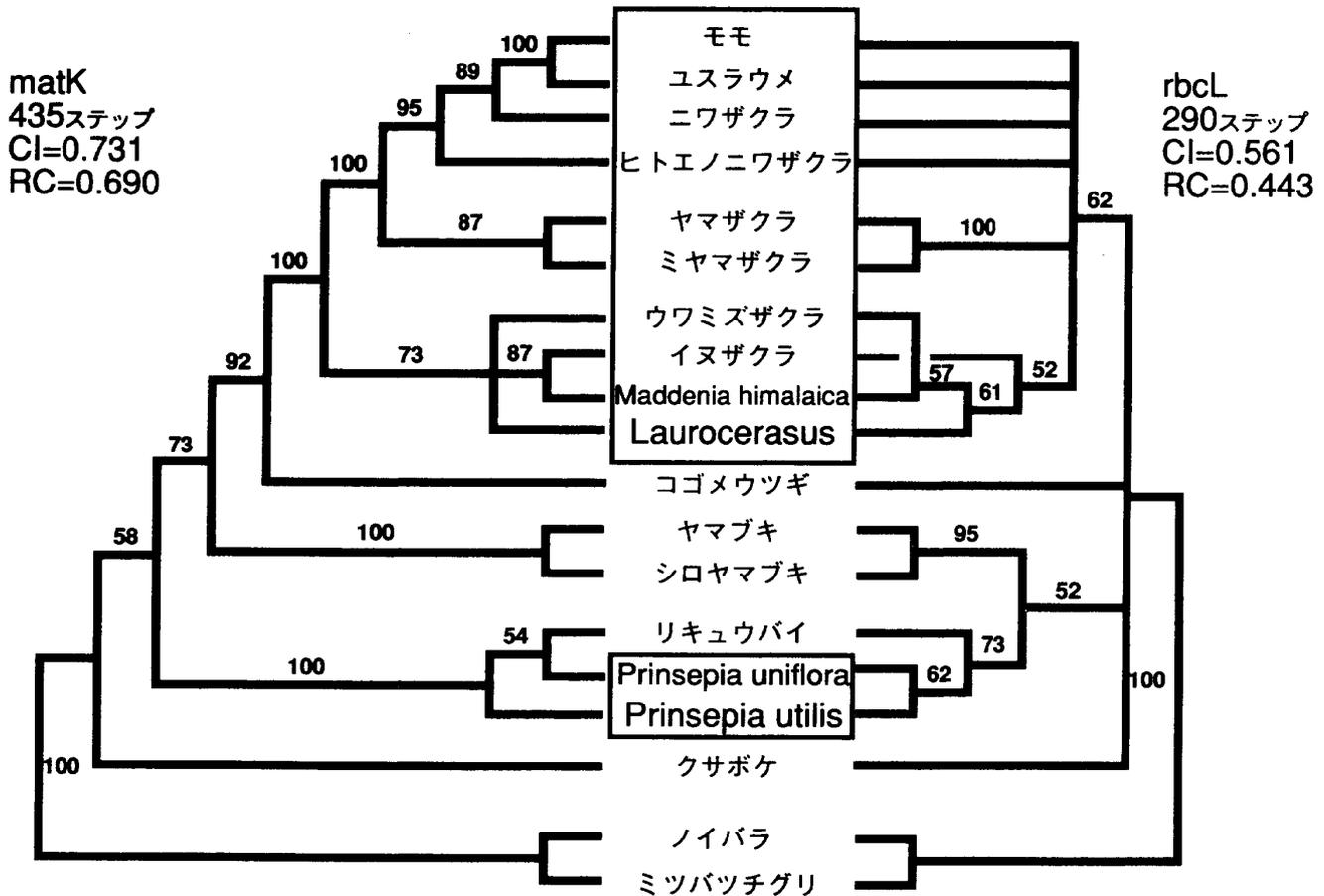


図2. *matK* と *rbcL* により構築したサクラ亜科とその周辺属の分岐図 (遠藤 未発表).
数字はブートストラップ法 (100回反復) によって得られた再現確率を示す。
四角の枠内: サクラ亜科

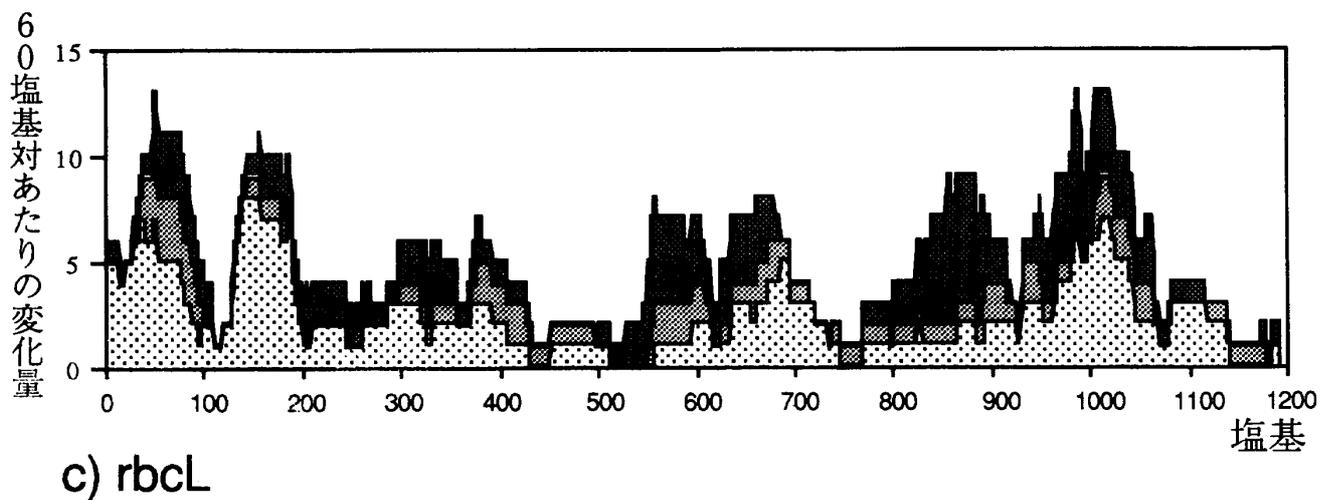
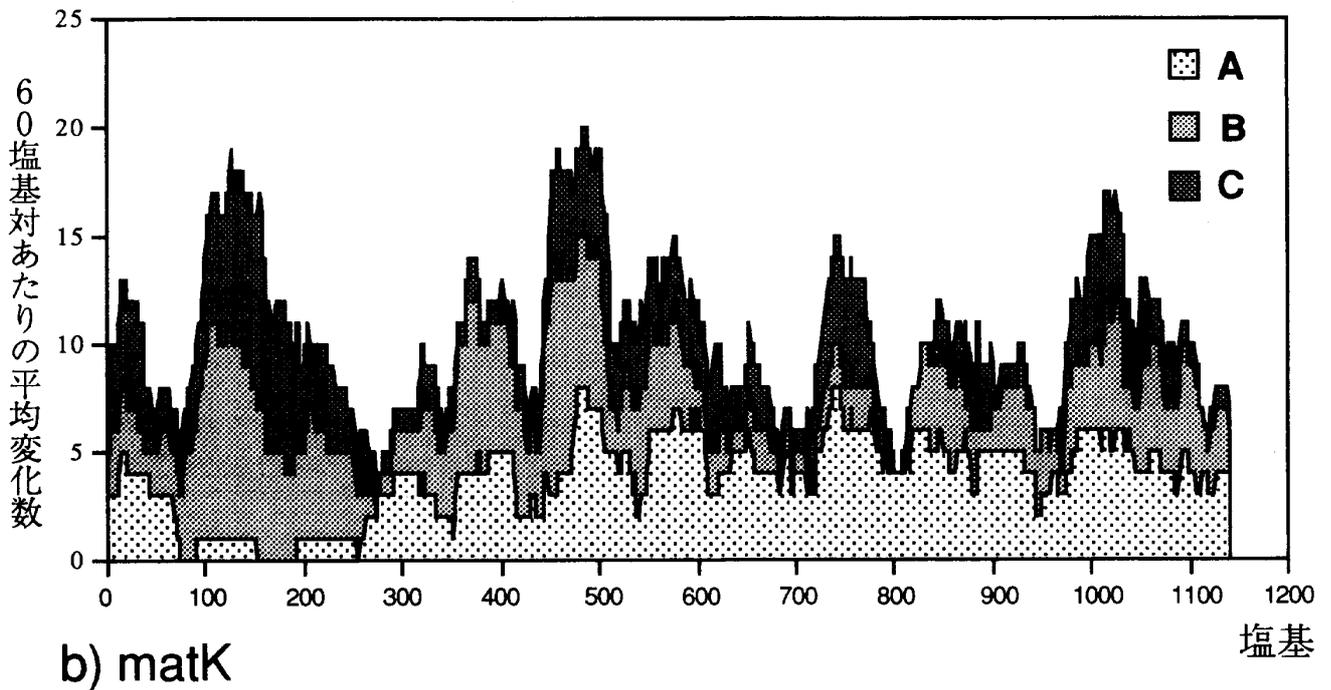
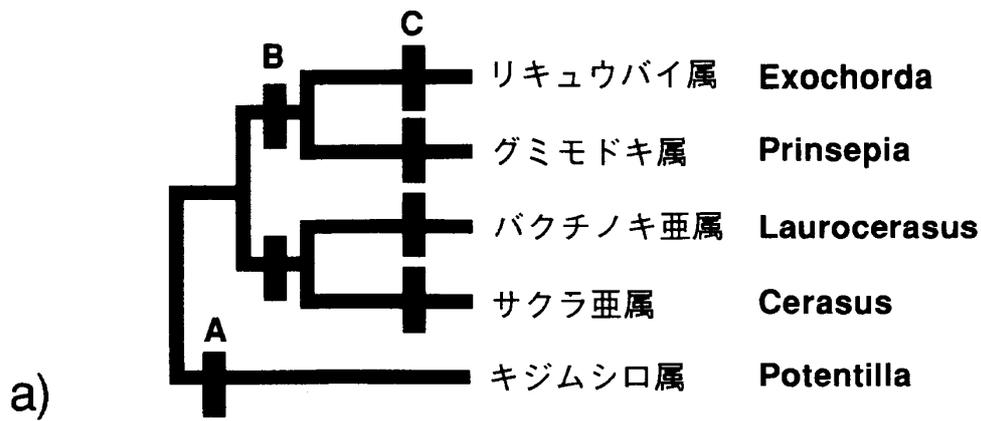


図3. *matK* と *rbcL* の塩基配列の変異の比較 (遠藤 未発表).

縦軸: *matK* と *rbcL* とともに同じ樹形を示す分岐図 a) の A, B, C 各位置での60塩基あたりの平均変化数.

rescaled consistency index) = 0.690で、*rbcl*の分岐図はCI = 0.561, RC = 0.443であった。このCIとRCは、形質状態の分布が得られた分岐図がどの程度一致しているかを示す尺度で、値が1に近いほど平行進化が少ない。*rbcl*で構築した分岐図は図2の右側のように解像度が低いのにに対して、*matK*で構築した分岐図(図2左側)は高い解像度と高いブートストラップ確率を示している。

*matK*の分岐図によって*rbcl*の分岐図ではわからなかった次の興味深い事柄がわかった。

1) サクラ亜科は単系統ではない。

コゴメウツギやヤマブキ、およびシロヤマブキがサクラ亜科の中に入っている。

2) サクラ亜属は単系統ではない。

ユスラウメ節 Sect. *Microcerasus* に属するユスラウメ *Prunus tomentosa* Thunb. ex Murray, ニワウメ *Prunus japonica* Thunb. ex Murray, ヒトエノニワザクラ *Prunus glandulosa* Thunb. ex Murray がモモ *Prunus persica* Batsch と単系統を形成する。

また、*rbcl*の分岐図でも示唆されているがより強く支持された事項として

3) これまでシモツケ亜科 Spiraeoideae に分類されていたリキュウバイ *Exochorda racemosa* (Lindle.) Rehd. はグミモドキ *Prinsepia uniflora* Batal. や *Prinsepia utilis* Royle などのグミモドキ属 *Prinsepia* に近いことがあげられる。

このように *matK* を用いた系統解析が *rbcl* のそれよりも有効であることがサクラ亜科の具体的な系統解析により明らかになった。

matK と *rbcl* で構築した分岐図の比較

次に、*rbcl*で構築した分岐図よりも *matK*で構築した分岐図の方が解像度が良い理由を調べた。図3のbとcは *matK* と *rbcl* の塩基配列の変化量を比較したエリアグラフである。塩基の長さを横軸に、縦軸には60塩基対あたりの変化数を示している。縦軸に注目するといかに *matK* の方が変化が多いかがわかる。次に分岐図上でどこにどれだけの変異があるのかを調べた。これはたとえ *matK* に変異が多くても分岐に関与する変化が少なければ解像度の高い分岐図は得られないからである。*matK* と *rbcl* で構築した分岐図(図2)から共通のトポロジーを持つ5種を選び、A, B, Cの位置に配置される変化を測定した(図3a)。この3つの位置に配置される変異形質のうちサクラ亜科の分岐図を作成する上で意味のある形質はBの位置の変異(共有派生

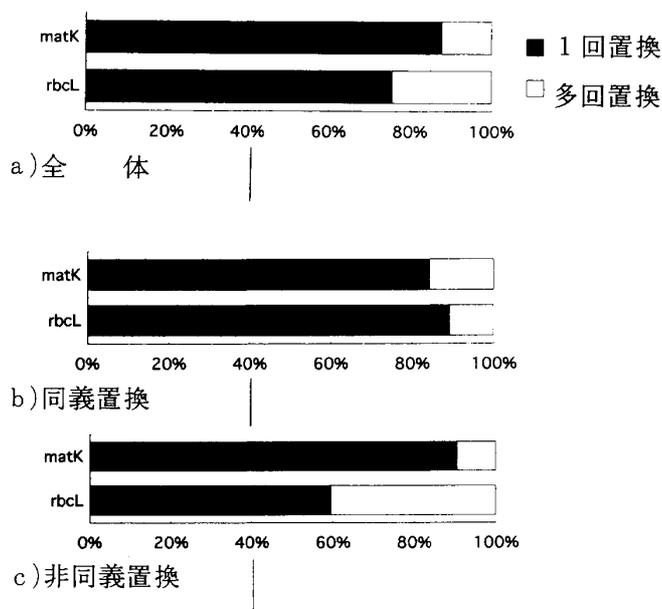


図4. *matK* と *rbcl* の塩基配列における1回置換と多回置換の割合(遠藤 未発表)。

形質)であることに注意してA, B, Cにおける変化量をみると *matK* では *rbcl* に比べてBの位置に配置される変異が多いことがわかる。このことから *matK* ではただ単に変異が多いだけではなく系統解析上で意味を持つ変化が多いことがわかった。これは今回解析したバラ科サクラ亜科の解析において、より高い解像度を得られることを意味する。

図4は、*rbcl*と *matK* で構築した系統樹上で塩基置換が1回だけ観察された(1回置換)サイトの数および同じサイトで複数回の塩基置換が観察された(多回置換)サイトの数の割合を調べたものである。検定の結果、全体についての多回置換数の割合は *rbcl* で有意に多いことが示唆された(図4a)。次に多回置換の割合が同義置換と非同義置換のどちらに多いのかを調べた。その結果、同義置換では *matK* と *rbcl* の間に有意な差は認められなかったが(図4b)非同義置換では *rbcl* で多回置換が有意に高い割合で起こっていることがわかった(図4c)。多回置換の割合が多いということは分子形質が平行進化をしていると考えることができる。*rbcl* ではこうした多回置換が起こりやすいホットスポットが非同義置換サイトに多く分布しており、系統解析をする上で解像度が悪くなる原因となることがわかった。このことはふつうホットスポットは同義置換サイトに多くなるという一般的な予想と異なるため、今後、多くの分類群間にわたる解析が望まれる。

要 約

バラ科において *matK* の塩基配列は *rbcl* の塩基配列

よりも情報のある変異が多く、平行進化が少ないので、ブートストラップなどの検定にも有意となる解像度の高い分岐図を得ることができた。変異が大きく一部のプライマーを分類群に応じて手直ししなくてはならない場合はあるが、*matK* の塩基配列を用いた系統解析は *rbcL* では解像度の低い分岐図しか得られなかった比較的近縁な分類群間の系統解析に有効であると考えられる。

謝 辞

大井和之（九州大学理学部）、大場秀章（東京大学総合研究資料館）、村上哲明（東京大学理学部附属植物園）、邑田仁（東京大学理学部附属植物園）、秋山忍（国立科学博物館）、横山潤（東北大学理学部）、鈴木三男（東北大学理学部）、能城修一（農林水産省森林総合研究所）、岡武男（甲府市立塩山南小学校）、中村正一（株式会社関電工）、西山智明（東京大学大学院理学系研究科）、遠藤純夫（立川市立第一中学校）をはじめデータ解釈や資料の分与、実験施設の提供などでたくさんの方々にお世話になりました。とくに大井和之氏にはデータ解析などで大変お世話になりました。

引用文献

- Bremer, K. 1988. The limits of amino acid sequence data in angiosperm phylogenetic reconstruction. *Evolution* 42 : 795-803.
- Chase, M.W., D.E. Soltis, R.G. Olmstead, D. Morgan, D.H. Les, R.D. Mishler, M.R. Duvall, R.A. Price, H.G. Hills, Y.-L. Qiu, K.A. Kron, J.H. Rettig, E. Conti, J.D. Palmer, J.R. Manhart, K.J. Sytsma, H.J. Michaels, W.J. Kress, K.G. Karol, W.D. Clark, M. Hedroen, B.S. Gaut, R.K. Jansen, K.-J. Kim, C.F. Wimpee, J.F. Smith, G.R. Furnier, S.H. Strauss, Q.-Y. Xiang, G.M. Plunkett, P.S. Soltis, S. Swensen, S.E. Williams, P.A. Gadek, C.J. Quinn, L.E. Eguiarte, E. Golenberg, G.H. Learn, Jr., S.W. Graham, S.C.H. Barrett, S. Dayanandan and V.A. Albert 1993. Phylogenetics of seed plants: An analysis of nucleotide sequences from plastid gene *rbcL*. *Ann. Mo. Bot. Gard.* 80 : 528-580.
- Cronquist, A. 1981. Rosaceae. *In: An Integrated Classification of Flowering Plants*, p. 573-577. Columbia University Press, New York.
- Felsenstein, J. 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution*. 39 : 783-791.
- Hiratsuka, J., H. Shimada, R. Whittier, T. Ishibashi, M. Sakamoto, M. Mori, C. Kondo, Y. Honji, A. Hirai, K. Shinozaki & M. Sugiura 1989. The complete nucleotide sequence of the rice (*Oryza sativa*) chloroplast genome: Intermolecular recombination between distinct tRNA genes accounts for a major plastid DNA inversion during the evolution of the cereals. *Mol. Gen. Genet.* 217 : 185-194.
- Hutchinson, J. 1964. 24 Rosaceae. *In: The Genera of Flowering Plants*, 1, p. 174-216. Oxford Univ. Press, Oxford.
- Johnson L.A. & D.E. Soltis 1994. *matK* DNA sequence and phylogenetic reconstruction in Saxifragaceae s. str. *Syst. Bot.* 19 : 143-156.
- Kalkman, C. 1988. The phylogeny of the Rosaceae. *Bot. J. Linn. Soc.* 98 : 37-59.
- Les, D.H., E. Landolt & D.J. Crawford 1994. Molecular systematics of the Lemnaceae. *Amer. J. Bot.* 81 : 168 (Abstract).
- Morgan, D.R., D.E. Soltis & K.R. Robertson 1994. Systematic and evolutionary implications of *rbcL* sequence variation in Rosaceae. *American Journal of Botany* 81 : 890-903.
- Neuhaus, H. & Link, G. 1987. The chloroplast tRNA^{Lys} (UUU) gene from mustard. *Curr. Genet.* 11 : 251-257.
- 大橋広好・靱山泰一・大場秀章 1989. バラ科 佐竹義輔ら（編）日本の野生植物 p.179-228. 平凡社
- Ooi, K., Y. Endo, J. Yokoyama, N. Murakami 1995. Useful primer designs to amplify DNA fragments of the plastid gene *matK* from angiosperm plants. *J. Jpn. Bot.* (in press)
- Plunkett, G.M., D.E. Soltis & P.S. Soltis 1994. Phylogenetic relationships within and between Araliaceae and Apiaceae based on *rbcL* and *matK* sequence data. *Amer. J. Bot.* 81 : 180 (Abstract).
- Shinozaki, K., M. Ohme, M. Tanaka, T. Wakasugi, N. Hayashida, T. Matsubayashi, N. Zaita, J. Chunwongse, J. Obokata, K. Yamaguchi-Shinozaki, C. Ohto, K. Torazawa, B.Y. Meng, M. Sugita, H. Deno, T. Kamogashira, K. Yamada, J. Kusuda, F.

- Takaiwa, A. Kato, N. Tohdoh, H. Simada & M. Sugiura 1986 . The complete nucleotide sequence of the tobacco chloroplast genome: its gene organization and expression. *EMBO Jour.* 5 : 2043-2049.
- Soltis, D.E., Q.Y. Xiang & L. Hufford 1995 . Relationships and evolution of Hydrangeaceae based on *rbcL* sequence data. *Amer. J. Bot.* 82 : 504-514.
- Steele, K.P. 1991 . Phylogenetic analyses of angiosperms using DNA sequences of the chloroplast gene *ORFK*. *Amer. J. Bot.* 78 : 220 (Abstract).
- Steele, K.P. & R. Vilgalys 1994 . Phylogenetic analyses Polemoniaceae using nucleotide sequence of the plastid gene *matK*. *Syst. Bot.* 19 : 126-142.
- Sugita, M., K. Shinozaki & M. Sugiura 1985 . Tobacco chloroplast tRNA^{Lys} (UUU) gene contains a 2.5-kilobase-pair intron: An open reading frame and a conserved boundary sequence in the intron. *Proc. Natl. Acad. Sci. U.S.A.* 82 : 3557-3561.
- Swofford, D.L. 1993 . PAUP: Phylogenetic analysis using parsimony, version 3.1.1 . Computer program distributed by the Illinois Natural History Survey, Champaign, Illinois .