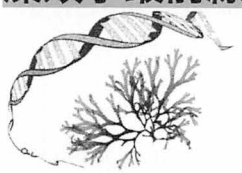
藻類 Jpn. J. Phycol. (Sôrui) 55: 111–116, July 10, 2007



稲垣祐司：ロングブランチの誘惑—分子系統解析のダークサイド

塩基・アミノ酸配列情報に基づく分子系統解析の主要目的は，配列情報の起源となる生物の進化的類縁関係を推定することであるが，その手法は現代生物学のすべての分野で何らかの形で利用されているといっても過言ではない。しかし，大多数のユーザーは，解析プログラムの基本設定に従ってコンピューター画面をクリックすれば，インスタントに正しい系統樹を得ることが出来ると考えているようである。実際には，①計算結果はあくまで推定であり，②その推定は解析条件に大きく依存し，③僅かな設定の違いが劇的な推定の違いに結びつきうる。つまり，解析条件・配列データの特性によって，推定結果には著しい偏り（アーティファクト）が生じうるのだが，このような認識はあまり広まっていないようである。

これまでの分子系統解析の方法論的研究では，系統的に近縁でない2つの配列の進化速度が極めて速い場合，これらの「ロングブランチ」配列が誤って近縁であると推定される「ロングブランチアトラクション（long-branch attraction or LBA）アーティファクト」がよく研究されている。分子系統解析において，これまでの知見と著しく異なる複数のロングブランチ配列のグルーピングが復元された場合，その解析結果はLBAアーティファクトの影響を受けていると解釈できる。本稿では，分子系統解析でよく用いられる最大節約法・距離法・最尤法のLBAアーティファクトに対する感受性を，シミュレーションデータをもちいた解析と現存配列データの解析により議論する。特に，最尤法解析における条件設定（進化モデル選択）の重要性について解説していきたい。最後にKolaczkowski & Thornton (2004) が行ったシミュレーション解析およびそれに対する反論論文における，最大節約法と最尤法のパフォーマンス比較についてコメントを加えた。

4-taxon tree をもちいたシミュレーション解析

我々は，現存する配列データ解析から系統関係を推定することは可能であるが，真の系統関係は知り得ない。一方，系統樹の形・枝長と置換モデルを予め設定し，その系統樹を元に配列データをシミュレーションすることが可能である。予め真の系統樹が判っているシミュレーションデータを解析することにより，①解析対象データの特性により推定結果にどんな傾向の偏りが生じるか，②もし偏りがあるとするならその度合いはどの程度か，を調べることが出来る。本項では「4-taxon tree」に基づき塩基配列データをシミュレートし，最大節約法・距離法・最尤法を用いてシミュレーションデータを解析した。

「4-taxon tree」はその名の通り4つのタクサからなる樹形である（図1A上）。この場合，タクソンAとタクソンBが姉妹群となっている（残りのタクサC & Dは自動的に姉妹群となる）。実験の都合上，タクソンAからタクサA & Bの共通祖先に至る枝の長さ，タクソンDからタクサC & Dの共通祖先に至る枝の長さをともにxとし，残りの枝の長さをすべてyとする。これら枝長xとyを，5通り（0.01, 0.15, 0.30, 0.45, 0.60）に変化させた樹形を用意し，各データポイント（合計$5 \times 5 = 25$ポイント）で100個の核酸シミュレーションデータを作成した。置換モデルには，トランジション・トランスバージョン比を2.0としたKimura 2 parameter（K2P）モデルを使用した。また，データサイズは50，300，1000，5000 nucleotides（nt）と変化させた4セットのシミュレーションデータを調製した。すべてのシミュレーションにはSEQ-GEN v.1.3.2 (Rambaut & Grassly 1997) を使用した。

枝長xがyよりも大きい場合，真の系統関係は2本のロングブランチは短い中央枝で隔てられている（図1A中央）。このようなロングブランチtreeから生成したシミュレーションデータを解析すると，タクサA & Dが「互いを引きつけ合う」典型的なLBAアーティファクトが起こりやすくなる（図1A下）。当然枝長xとyとの比率や，使用する解析法によってLBAの強さは変わると予想される。各データポイントでのLBAの強さは，100個のシミュレーションデータ解析がどの程度「正しい推定」，即ちタクサA & Bを姉妹群として復元できたかによりモニターできる。

最大節約法： 図1Bは，各種データサイズのミュレーションデータを最大節約法により解析した結果である。解析にはPHYLIP v.3.65パッケージ (Felsenstein 1993) のDNAPARSを用いた。ここでは，各データポイントにおけるタクサA & Bの姉妹群関係を復元した「正解率」0~100%までを，白から黒で表している。枝長xにかかわらず枝長yが0.30以上の場合，正解率100%だったため図から省略した。最大節約法の解析では，$x>>y$の場合，タクサA & Bの姉妹群関係を復元することが困難であり，この傾向はデータサイズを大きくしても変化せず，むしろLBAの影響が強くなっていくことが分かる(図1B)。$x = 0.60/y = 0.01$および$x = 0.45/y = 0.01$のデータポイントにおける正解率はまったく同じで，データサイズを50 nt，300 nt，1000 nt，5000 ntのとき7%，0%，0%，0%となった（図1C）。$x = 0.30/y = 0.01$のデータポイントでようやく若干の正解率の改善が見られたが，データサイズを50 nt，300 nt，1000 nt，5000 ntのとき15%，4%，0%，0%であった（図1C）。

では，タクサA & Bの姉妹群関係を復元することができなかった場合，どのような（誤った）系統関係が推定されていたのであろうか。4-taxon treeの場合，不正解には2通りが考えられる。即ち，LBAアーティファクトであるタクサA &

Dの姉妹群関係を復元する場合と，LBAとは関係なくタクサA & Cの姉妹群関係を復元する場合である。ここには示さないが，最大節約法の解析結果を精査したところ，不正解の多くはタクサA & Dが姉妹群となるLBAアーティファクトであった。特に，データサイズが1000，5000 ntの場合，すべての不正解はLBAアーティファクトであった。以上の実験結果により，最大節約法推定は，LBAに敏感であり，かつデータサイズが大きい場合LBAアーティファクトを選択的に，しかもきわめて強く支持することが分かる。

4-taxon treeによるシミュレーション解析において，LBAによる影響が大きい場所を一般的に「Felsensteinゾーン」と呼ぶ。今回の最大節約法による解析では，$x > 0.30/y = 0.01$となる3つのデータポイントが著しいFelsensteinゾーンとなる。このシミュレーション解析の結果から，現存配列データ解析における最大節約法の推定の振舞を演繹すると，データサイズが十分に大きく，しかも不幸にしてFelsensteinゾーンに陥った場合，最大節約法は著しくLBAの影響を受けると考えられる。しかし現存配列データの解析では，我々には真の系統関係がわからないため，推定結果がアーティファクトであることを認識できない危険性が高い。

距離法： Neighbor-joining法による距離法の解析では，PHYLIP v.3.65 (Felsenstein 1993)を使用した。DNADISTにより距離マトリックスを計算し，そのマトリックスをもとにNEIGHBORを使って系統樹を推定した。DNADISTによる距離計算では，シミュレーション時と同一モデルであるK2Pモデル（Ts/Tv = 2.0）を指定した。各データポイントにおける正解率は濃淡で表示してある（図1C）。距離法による解析でも，枝長yがきわめて短い場合（$y = 0.01$），タクサA & Bの姉妹群関係を復元するのは難しいが，最大節約法（図1B）との比較をすると，全体的な正解率は高いことが分かる。例えば$x = 0.60/y = 0.01$のデータポイントでデータサイズを50 nt,

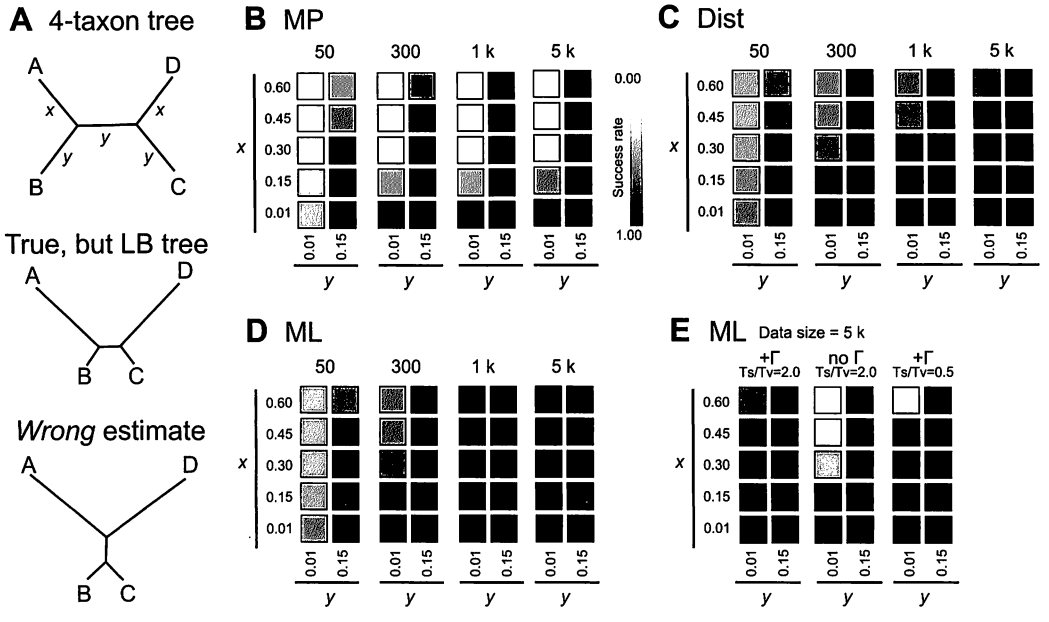


図1　4-taxon treeを用いたシミュレーション解析　(A) 4-taxon tree：データシミュレーションに用いた4-taxon treeでは，AとDに至る枝の長さをx, その他の枝の長さをyとした（上）。xおよびyの長さは0.01, 0.15, 0.30, 0.45, 0.60と変化させ，シミュレーションを行った（合計25データポイント）。各データポイントに100個のシミュレーションデータを作成した。$x>>y$となる典型的なロングブランチtree(中央)をもとにシミュレーションし，そのデータを解析した場合，真の系統関係であるタクサA & Bの姉妹群が復元されず，タクサA & Dが姉妹群と推定される(下)。 (B) 最大節約法（MP）による解析：シミュレーションにはK2Pモデル（トランジション・トランスバージョン比：Ts/Tv = 2.0）を用いた。最大節約法による解析で復元されたタクサA & Bが姉妹群となる樹形を正解とし，正解率0~100%を白から黒へのグラデーションで示した。y>0.30のすべてのデータポイントで正解率が100%だったので省略した。シミュレーションデータのサイズを50，300，1000（1 k），5000（5 k）塩基と変化させて解析を繰り返した。$x \geq 0.30/y$=0.01データポイントで300塩基以上のシミュレーションデータを解析すると，正解率は0%となった。(C)距離法(Dist)による解析:詳細は(B)と同じ。$x>>y$の条件でも，シミュレーションデータのサイズを上昇させると正解率は上昇した。(D) 最尤法（ML）による解析:詳細は（B）と同じ。距離法による解析と同じく，シミュレーションデータのサイズを上昇させると正解率は上昇した。ただし，距離法よりも全体的な正解率は高かった。例えばx=0.60/y=0.01データポイントでも，1000塩基データを解析すれば正解率は100%となった。(E) 最尤法（ML）による解析：この解析では5000（5 k）塩基データを，アライメント座位間での進化速度差を考慮したK 2Pモデル（+ Γ, Ts/Tv = 2.0）のもとシミュレーションした。シミュレーションと同一のモデルをもちいて最尤法で解析した場合（モデル不整合なし: + Γ, Ts/Tv = 2.0），テストしたすべてのデータポイントで正解率は100%となった。解析に，アライメント座位間での進化速度差を無視することでモデル不整合を発生させた場合（no Γ, Ts/Tv = 2.0），$x \geq 0.45/y$=0.01データポイントで正解率が0%に低下した。アライメント座位間での進化速度差は考慮するがTs/Tv = 0.5と設定することでモデル不整合を発生させた場合（+ Γ, Ts/Tv = 0.5），$x \geq 0.45/y$=0.01データポイントで正解率の低下が観察された。

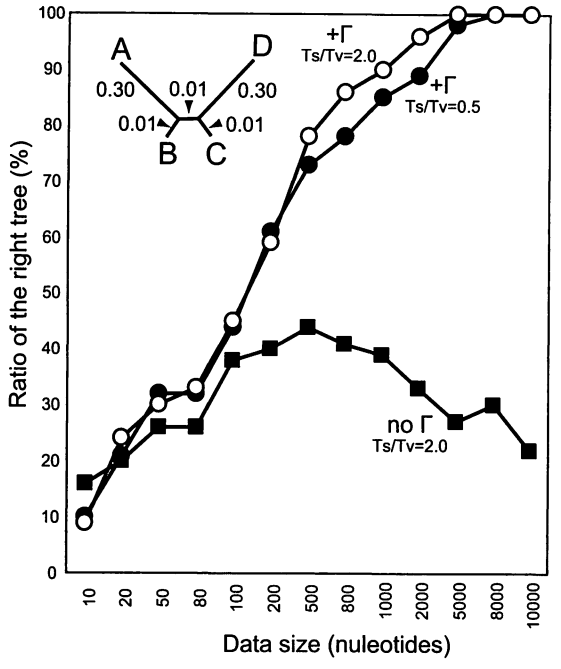


図2　モデル不整合条件下の最尤法　これらの解析では，2種類のモデル不整合条件下における，ロングブランチアトラクションアーティファクトとデータサイズとの関係を調べた。x=0.30, y=0.01 に設定した 4-taxon tree を元に 10~10000 塩基のシミュレーションデータを，アライメント座位間での進化速度差を考慮した K2P モデル（+ Γ, Ts/Tv = 2.0）をもちいてシミュレーションした。モデル不整合がない最尤法による解析での正解率は白丸でプロットした。アライメント座位間での進化速度差は考慮するが Ts/Tv = 0.5 と設定することでモデル不整合を発生させた場合（+ Γ, Ts/Tv = 0.5），解析でアライメント座位間での進化速度差を無視することでモデル不整合を発生させた場合（no Γ, Ts/Tv = 2.0），それぞれの正解率を黒丸，黒四角でプロットした。アライメント座位間での進化速度差を無視した場合，データサイズが 500 塩基より大きくなると正解率は低下した。

300 nt, 1000 nt, 5000 nt へ変化させると，距離法の正解率は 45%, 56%, 66%, 78% と上昇した（図 1C）。以上の解析から，距離法による推定は LBA アーティファクトの影響を受けるが，最大節約法に比べその程度は軽微であると考えられる。また，データサイズを大きくすることにより，LBA アーティファクトの軽減できる。この実験で試したデータサイズよりもさらに大きなデータ（5000 nt 以上）を解析すれば，ある時点で LBA アーティファクトを排除することができると考えられる。

（モデル整合下における）最尤法：　シミュレーションと解析に同一モデル（K2P ; Ts/Tv = 2.0）を指定し，PHYLIP v.3.65 (Felsenstein 1993) の DNAML により最尤法の解析を行った。結果は図 1D に示したが，その詳細は最大節約法・距離法の解析結果（図 1B & C）と同じである。データサイズに関わりなく，最尤法からの正解率は他の解析法からの正解率よりも高かった。x = 0.60/y = 0.01 データポイントにおいて，データサイズを 50 nt, 300 nt, 1000 nt, 5000 nt と変化させると，最尤法の正解率は 56%, 63%, 83%, 100% と上昇した（図 1D）。つまり，十分に大きなサイズのデータを与えれば，最尤法による推定から LBA アーティファクトを排除することができることを示している。これまでのシミュレーション解析の結果を総合すると，最尤法による推定が，最大節約法・距離法の推定に比べて LBA アーティファクトに対して最も頑健であることが分かる。では，十分に大きなサイズの配列データを最尤法で解析すれば，LBA アーティファクトの影響を受けない「正しい」推定が可能なのであろうか。残念ながら最尤法はオールマイティーではない。その事実を以下の 2 つの実験で確かめる。

モデル不整合下における最尤法（1）：　今回は，アライメント座位間での進化速度差を考慮した K2P モデル（+ Γ, Ts/Tv = 2.0）を使用し，5000 nt の配列データのみをシミュレーションした。このデータを，①アライメント座位間での進化速度差を考慮した K2P モデル（+ Γ, Ts/Tv = 2.0），②アライメント座位間での進化速度差を考慮しない K2P モデル（no Γ, Ts/Tv = 2.0），③アライメント座位間での進化速度差を考慮した K2P モデルだが Ts/Tv = 0.5 と設定した場合（+ Γ, Ts/Tv = 0.5），の 3 パターンで最尤法をもちいて解析した。①ではシミュレーションとデータ解析における置換モデルが一致しているが（モデル整合条件），②と③ではシミュレーションとデータ解析における置換モデルに食い違いが生じている（モデル不整合条件）。

モデル整合条件下での最尤法による解析では，データサイズ 5000 nt と大きいため解析したすべてのデータポイントで正解率が 100% に達した（図 1E 左）。ところが，配列データをシミュレーションの際に考慮したアライメント座位間での進化速度差を，解析する際には無視した「モデル不整合」条件下では Felsenstein ゾーンが出現した（図 1E 中央）。x = 0.60/y = 0.01, x = 0.45/y = 0.01 データポイントでは正解率は 0% で，復元されたすべての樹形でタクサ A & D が姉妹群，即ち LBA アーティファクトとなっていた。x = 0.30/y = 0.01 データポイントでは正解率 28%，不正解だった 72% 全ては LBA アーティファクトであった。

トランジション・トランスバージョン比を，シミュレーションでは Ts/Tv = 2.0，解析では Ts/Tv = 0.5 としてモデル不整合を発生させた場合も，最尤法の推定に偏りが生じた（図 1E 右）。アライメント座位間での進化速度差を無視した場合と比べ，LBA アーティファクトの程度は軽いが x = 0.60/y = 0.01, x = 0.45/y = 0.01 データポイントにおいて推定ミスが生じ，正解率はそれぞれ 14%, 82% となった。

モデル不整合下における最尤法（2）：　モデル不整合下での最尤法の推定の偏りとデータサイズの相関を，細かく解析したのが図 2 である。この解析では，1 種類の 4-taxon tree（x = 0.30/y = 0.01;図 2 参照）をもとにシミュレーションを行い，データサイズを 10 から 10000 nt まで変化させ，正解率の変化をモニターした。アライメント座位間での進化速度差を考慮した K2P モデル（+ Γ, Ts/Tv = 2.0）をシミュレーションに使用した。モデル整合条件下で解析する場合，データサイズに比例し

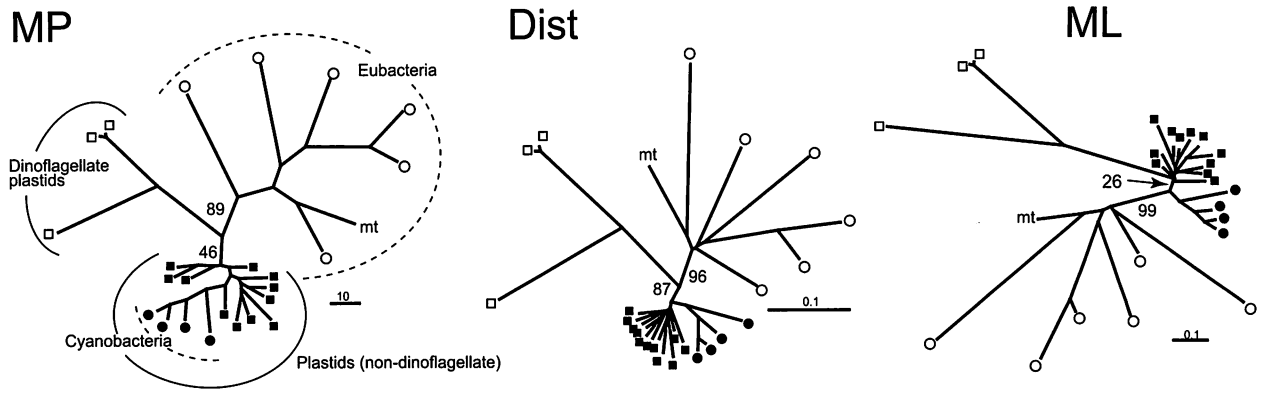


図3　葉緑体 ATP synthase β subunit 遺伝子（*atpB*）の系統解析　左から右へ最大節約法（MP），距離法（Dist），最尤法（ML）による推定結果を示した（ブートストラップ 100 回）。詳しい配列名はすべて省略したが，渦鞭毛藻類配列は白四角，その他の葉緑体配列は黒四角，シアノバクテリア配列は黒丸，シアノバクテリア以外の真正細菌類配列は白丸で示した。出芽酵母ミトコンドリア配列は「mt」と表記した。最尤法（右）ではすべての葉緑体の単系統性を復元したが，最大節約法（左）と距離法（中央）による推定では，渦鞭毛藻類配列と外群配列間のロングブランチアトラクションの影響のため葉緑体配列が単系統とならなかった（図 4 参照）。

て正解率が上昇した（図 2；白丸）。データサイズ 2000 nt で正解率 90% 以上となり，5000 nt で正解率は 100% に到達した。一方アライメント座位間の進化速度差を無視した場合，データサイズ 500 nt 程度までは正解率が上昇するが，その後正解率は低下した（図 2;黒四角）。実験はしていないが，さらにデータサイズを大きくすれば正解率は 0% になると予想できる。つまりアライメント座位間の進化速度差を無視して生じるモデル不整合，それに関連する LBA アーティファクトは，きわめて深刻でありデータサイズを大きくしても解消できないと考えられる。それとは反対に，アライメント座位間の進化速度差は考慮するが，トランジション・トランスバージョン比を Ts/Tv = 0.5 に設定することで生じるモデル不整合条件下では，データサイズに比例して正解率も上昇し，8000 nt 以降で正解率 100% に到達した（図 2；黒丸）。この解析結果は，トランジション・トランスバージョン比に関するモデル不整合を原因とする LBA アーティファクトは，データサイズにより解消することができることを示している。

最尤法によるシミュレーション解析のまとめ：　最尤法をもちいたシミュレーション解析から，2 つの結論を導くことが可能である。第 1 は，最尤法はモデル不整合条件下では LBA アーティファクトの影響を強く受けるということである。第 2 は，配列生成時に考慮したどのパラメーターを無視するかにより，LBA アーティファクトの度合いが変わりうることである。

これまで見てきたように最尤法による系統推定には置換モデルの選択がきわめて大きな影響を与えるため，モデル選択を慎重に行う必要がある。モデル不整合条件下のシミュレーション解析を鑑みると（図 1E & 2），少なくともアライメント座位間の進化速度差をモデル化することは必須であろう。ただ，現在使用できるもっとも複雑な置換モデル（例えば核酸解析に使用する，アライメント座位間の進化速度差を考慮した General-Time-Reversible モデル）でも，現存配列データがどのように進化してきたかを完全に記述することは不可能である。つまり，いかなる解析もモデル不整合から逃れ得ず，系統推定には必ず何らかのアーティファクトが含まれている可能性がある。

今回の解析で明らかなように，最大節約法・距離法・最尤法いずれの解析でも Felsenstein ゾーンは発生し，この領域ではどんな解析を行ってもアーティファクトしかアウトプットとして出てこない。この事実に関連して，もう 1 つコメントを述べたい。学会等で分子系統解析の発表を聞いていると，「最大節約法・距離法・最尤法からの結果は互いに一致し，かつ BP 値 100% でその関係をサポートした」というコメントをよく耳にする。「異なる 3 種類の解析が一致したので，その結果は尤もらしい」と主張したいのであろうが，3 種類の解析が共通してアーティファクトをアウトプットする Felsenstein ゾーンのような条件が存在することは明白である。今後この事実を認識し，自らの分子系統解析の結果を慎重に評価すべきである。

葉緑体 ATP synthase β subunit 配列解析：LBA アーティファクトのケーススタディ

前項では 4-taxon tree に基づくシミュレーションデータを用いて，LBA アーティファクトに対する最大節約法，距離法，最尤法の頑健性を評価した。本項では，シミュレーション実験の結果と現存配列データの解析結果との間に整合性があるかを検討する。ケーススタディとして葉緑体遺伝子 ATP synthase β subunit（*atpB*）の系統解析を紹介する。

渦鞭毛藻類の葉緑体遺伝子配列は，他の葉緑体のホモロガスな配列に比べて著しく置換速度が上昇している (Iida *et al.* 2007)。従って，渦鞭毛藻類葉緑体遺伝子配列は LBA アーティファクトの原因となる可能性が高く，これらの配列を含む系統解析とその推定結果の解釈は慎重に行うべきである。今回の解析では，渦鞭毛藻類から 3 種，他の真核藻類から 11 種，シアノバクテリアから 4 種，外群としてシアノバクテリア以外の真正細菌類から 7 種（パン酵母ミトコンドリア配列を含む）の *atpB* アミノ酸配列アライメントを作成した（合計 25

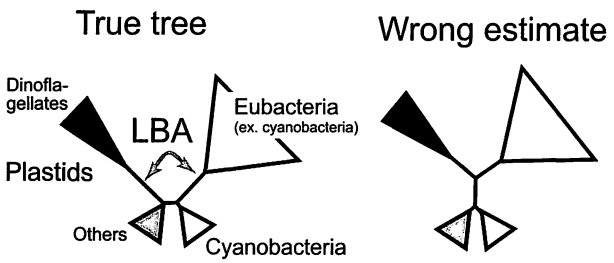


図4　葉緑体 ATP synthase β subunit 遺伝子（*atpB*）解析におけるロングブランチアトラクションアーティファクト　真の系統関係では，渦鞭毛藻類配列を含む全ての葉緑体は単系統となる（左）。しかし，最大節約法と距離法による推定では，枝長の長い渦鞭毛藻類配列と真正細菌類配列間にロングブランチアトラクションが起こり，葉緑体配列の単系統性が復元できない（右）。

タクサ，242 アミノ酸座位）。このデータを，PAUP* v.4.0b10 (Swofford 1998) を使用した最大節約法，PROTDIST（距離計算は JTT モデルに基づく）と NEIGHBOR を使用した距離法，PHYML v.2.4.4 (Guindon & Gascuel 2003) を使用しアライメント座位間の置換速度差をモデル化した JTT モデルに基づく最尤法により解析した（図 3）。これまでの知見では，シアノバクテリアー真核生物間で起こった 1 回の細胞内共生イベントがすべての葉緑体の起源となったと考えられている。従って全ての葉緑体配列は *atpB* 系統樹中で単系統となるはずだが，進化速度の速い渦鞭毛藻類配列を含む系統解析は，葉緑体の単系統性を LBA アーティファクトの影響を受けずに復元できるであろうか。

最大節約法と距離法による推定では，渦鞭毛藻類以外の葉緑体配列とシアノバクテリア配列がクレードを形成したが，渦鞭毛藻類葉緑体配列は，葉緑体＋シアノバクテリア配列クレードと外群をつなぐ枝から分岐した（図 3 左・中央）。内群配列（すべての葉緑体配列とシアノバクテリア配列）のグルーピングは，最大節約法・距離法により高い BP 値 89%・96% で支持された。最大節約法では，内群クレード内の分岐に 50% 以上の BP 値が付かないため，渦鞭毛藻類葉緑体とその他の葉緑体との関係に結論を出すことはできない（図 3 左）。一方，距離法による推定では，渦鞭毛藻類葉緑体配列と，他の葉緑体配列＋シアノバクテリア配列から構成されるクレードとは高い BP 値 87% で隔てられた（図 3 中央）。この推定結果は，渦鞭毛藻類葉緑体の起源とその他の葉緑体の起源は異なることを示唆する。しかし，渦鞭毛藻類葉緑体と外群配列（おもに真正細菌類配列）の枝長は，渦鞭毛藻類以外の葉緑体・シアノバクテリア配列の枝長よりもかなり長い。従って，本来なら渦鞭毛藻類葉緑体配列とその他の葉緑体配列は進化的に近縁関係にあるはずだが，最大節約法・距離法による推定では進化速度が速い（枝長が長い）渦鞭毛藻類葉緑体と外群配列の間に LBA が働き，推定結果に偏りが生じたと考えることができる（図 4）。

シミュレーション解析では，最尤法が最大節約法・距離法よりも LBA アーティファクトに頑健であると考えられる（図 1B-D）。そこで，全く同じ *atpB* 配列データを最尤法により解析したところ，ブートストラップ（BP）値は 26% と低いものの全ての葉緑体配列の単系統性が復元された（図 3 右）。さらに，葉緑体クレードとシアノバクテリアクレードとが姉妹群となり，極めて高い BP 値 99% で支持された（図 3 右）。従って，最尤法による推定は（渦鞭毛藻類葉緑体を含めた）すべての葉緑体の単系統性を支持しているといえる。それに対し，最大節約法・距離法による推定結果で葉緑体の単系統性を復元できなかったのは，LBA アーティファクトが原因であると考えられる（図 4）。

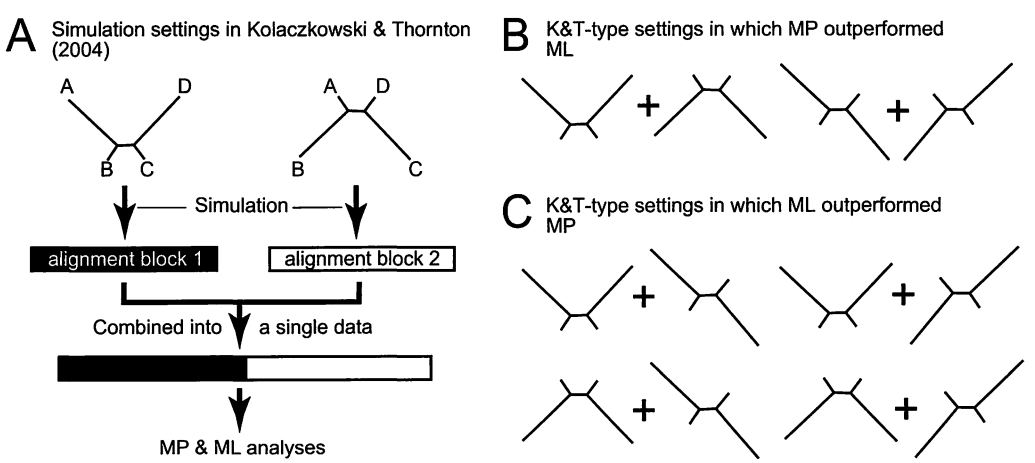


図5　Kolaczkowski & Thornton（K&T）タイプシミュレーション解析　（A）K & T によるオリジナル解析（Kolaczkowski & Thornton 2004）同一のトポロジーだが枝長の異なる 2 つの 4-taxon tree をもとにシミュレーションデータをそれぞれ作成した。これら 2 つのシミュレーションデータを結合して，最大節約法と最尤法による推定を行った。この条件では最尤法よりも最大節約法が優れたパフォーマンスを示した。（B & C）Spencer *et al.* (2005) による K & T タイプシミュレーション解析　（B）に模式的に示した 2 組の樹形ペアから作成したシミュレーションデータの解析では，最尤法よりも最大節約法が優れたパフォーマンスを示した。一方，（C）に模式的に示した 4 組の樹形ペアから作成したシミュレーションデータの解析では，最大節約法よりも最尤法が優れたパフォーマンスを示した。詳しくは Spencer *et al.* (2005) を参照のこと。

一般的に，現存配列の系統解析においてLBAアーティファクトを疑うことは比較的容易である。なぜなら，解析した配列中に極端なロングブランチ配列があれば，何らかの形でLBAアーティファクトが生じていると推測できるからである。特に注意すべきなのは，アーティファクトと思しき解析結果がこれまで蓄積された知見とは異なり，目新しい魅力的な仮説を提供する場合である。系統解析結果は，あくまでも特定の解析（実験）条件下での推定であることを自覚し，推定結果を徹底的に検討する前に安易な（しかし魅力的な）結論に飛びつくことは避けるべきである。

最大節約法 vs. 最尤法：Kolaczkowski & Thorntonの解析について

4-taxon treeを用いたシミュレーション解析では，LBAアーティファクトに対して最大節約法よりも最尤法の推定が頑健であることを示した（図1 & 2）。しかし，2004年にNature誌に発表されたKolaczkowski & Thornton（K&T）のシミュレーション解析はその正反対を示していた(Kolaczkowski & Thornton 2004)。この解析では，樹形は同一だが枝長が異なる2種類の系統樹を用意・シミュレーションに使用し，2つのシミュレーションデータを連結した（図5A）。この連結データを最大節約法と最尤法により解析したところ，驚くべきことに最大節約法が最尤法よりもよいパフォーマンスを示したのである。この結果をもとに，著者らは少なくとも最大節約法と最尤法との推定を同等に評価すべきであると主張した(Kolaczkowski & Thornton 2004)。

このK&Tのシミュレーション解析は多くの反響を呼び，それに反論する論文が発表されている(Gadagkar & Kumar 2005; Gaucher & Miyamoto 2005; Philippe *et al.* 2005; Spencer *et al.* 2005)。これらの反論論文の主張は，K&Tのシミュレーションおよび使用した2種類の枝長セットは，最大節約法が最尤法よりもよいパフォーマンスを示すきわめて限定的な実験条件であることを指摘している。オリジナルのK&Tのシミュレーション解析を拡大し，15条件を検討した場合，2条件では最大節約法が最尤法よりもよいパフォーマンスしたが（図5B），反対に最尤法が最大節約法よりもよいパフォーマンスを示す4条件が発見された（図5C）。残りの9条件では最尤法と最大節約法とで有意な差が検出されなかった。つまり，全体的パフォーマンスでは最大節約法に対して最尤法が優位に立っていると解釈できる。Spencer *et al.* (2005)を初め他の反論論文でも，現存データ解析において，やはり最尤法のパフォーマンスが最大節約法を上回るであろうと結論を下している。詳しくは原論文を参照されたい。

最後にK&Tタイプシミュレーション解析と連結データ解析との関連について述べたい。通常のシミュレーション解析では，最尤法はLBAアーティファクトに対してきわめて頑健であるが，最大節約法は非常に敏感である（図1 & 2）。一方，K&Tタイプシミュレーション解析では最尤法もLBAアーティファクトに対して敏感となる(Kolaczkowski & Thornton 2004)。興味深いことにSpencer *et al.* (2005)は，K&Tタイプシミュレーションが2種類の系統樹から生成した「アライメントブロック」から形成されていることを考慮して最尤法推定を行うと，その結果は非常に正確になることを示している（原論文Fig. 4参照）。一般に，単一遺伝子データにはK&Tタイプの複数アライメントブロックが存在するかどうかさえ不明である。しかし，複数遺伝子配列を連結した場合，その連結データ中の単一遺伝子データが，まさにK&Tタイプアライメントのブロックに対応する。従って，K&Tのオリジナルシミュレーション解析は，配列進化パターンが異なるアライメントブロックの存在を考慮しない，所謂「Concatenate（或はLinked）モデル」に基づく連結データ解析であり，それは極めてミスリードされやすいことを示している。それと対応し，Spencer *et al.* (2005)が行った解析は，単一遺伝子データブロックに特異的な配列進化パターンを考慮する「Separate（或はUnlinked）モデル」による連結データ解析の頑健性を証明したのである。これらK&Tタイプシミュレーション解析結果は，現存配列データでのConcatenateモデルに対するSeparateモデルの優位性ともよく合致している。Concatenate・Separateモデルの詳細，2つのモデルを使用した解析における結果の相違などは，坂口の稿で詳しく述べているので参照してほしい。

引用文献

Felsenstein, J. 1993. PHYLIP: phylogeny inference package version 3.6. University of Washington, Seattle.

Gadagkar, S. R. & Kumar, S. 2005. Maximum likelihood outperforms maximum parsimony even when evolutionary rates are heterotachous. Mol. Biol. Evol. 22: 2139–2141.

Gaucher, E. A. & Miyamoto, M. M. 2005. A call for likelihood phylogenetics even when the process of sequence evolution is heterogeneous. Mol. Phylogenet. Evol. 37: 928–931.

Guindon, S. & Gascuel, O. 2003. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. Syst. Biol. 52: 696–704.

Iida, K., Takishita, K., Ohshima, K. & Inagaki, Y. 2007. Assessing the monophyly of chlorophyl-*c* containing plastids by multigene phylogenies under the unlinked model conditions. Mol. Phylogenet. Evol. (in press).

Kolaczkowski, B. & Thornton, J. W. 2004. Performance of maximum parsimony and likelihood phylogenetics when evolution is heterogeneous. Nature 431: 980–984.

Philippe, H., Zhou, Y., Brinkmann, H., Rodrigue, N. & Delsuc, F. 2005. Heterotachy and long-branch attraction in phylogenetics. BMC Evol. Biol. 5: 50.

Rambaut, A. & Grassly, N. C. 1997. Seq-Gen: an application for the Monte Carlo simulation of DNA sequence evolution along phylogenetic trees. Comput. Appl. Biosci. 13: 235–238.

Spencer, M., Susko, E. & Roger, A. J. 2005. Likelihood, parsimony, and heterogeneous evolution. Mol. Biol. Evol. 22: 1161–1164.

Swofford, D. L. 1998. PAUP*: phylogenetic analysis using parsimony (*and other methods) version 4.

（筑波大学）