

|  |                         |    |      |    |  |     |              |
|--|-------------------------|----|------|----|--|-----|--------------|
| 28.  | 樹長分布に基づくDNA分子系統樹の信頼性評価法 |    |      |    |  |     |              |
| 要約 与えられたDNA塩基配列情報のもとで、すべての系統樹の樹長の頻度分布を作成することにより、ある分類群の全樹長に対する平均的貢献度（強度）を測定し、系統樹の信頼性を数値的に評価する方法を開発した。 |                         |    |      |    |  |     |              |
| 農環研 環境管理部 計測情報科 調査計画研究室  |                         |    |      |    |  | 連絡先 | 0298-38-8222 |
| 部会名  | 環境評価・管理                 | 専門 | 情報処理 | 対象 |  | 分類  | 研究           |

#### [背景・ねらい]

農業生物および環境生物の分類・同定にとって、生物の分類体系は重要である。系統分類体系の基礎となる系統樹の統計学的信頼性の評価方法の確立は、複数の対立する系統樹や系統分類体系の相互比較にとっても不可欠であるが、従来用いられてきたブーツストラップ法など標本再抽出に基づく評価法は、形質間の独立性という生物学的に厳しい前提を要求していた。ここでは、DNA塩基配列の情報に基づいて最節約法によって推定される系統樹の枝の信頼性を、塩基間の独立性を仮定せずに評価する新しい手法を開発した。

#### [成果の内容・特徴]

- ① ある系統樹のもとで形質情報（塩基配列データ）を説明するときに必要な形質進化の総数をその系統樹の樹長と呼ぶ。可能な系統樹すべてに対する樹長を計算し、頻度分布を作成した。この頻度分布を無制約樹長分布と呼ぶ。
- ② ある分類群（系統樹のある枝）を樹形上の制約を満足する系統樹について、同じ形質情報のもとでの制約樹長分布を計算する。ここで求めた制約樹長分布の期待値と上で求めた無制約樹長分布の期待値との差をその分類群の強度と定義した。強度は、その分類群が系統樹の全長すなわち形質情報の要約効率に対して平均的にどの程度貢献するかの尺度であると解釈された。強度の値が大きいほど、その分類群はより信頼できる（図1, 2）。
- ③ ここで提唱した樹長分布に基づく評価方法は、形質の無作為再抽出を行わないので、形質間の独立性が仮定できない場合にも適用することができる。したがって、形質間の独立性が必要なブーツストラップ法などの再抽出法よりも適用範囲が広い。
- ④ 黄緑色藻類Prymnesiophytaの光合成暗反応に関するRubisco遺伝子(*rbcL*と*rbcS*)のDNA塩基配列のデータに対してこの計算方法を適用し、得られた最節約系統樹の各枝の強度を計算した（図3）。

#### [成果の活用面・留意点]

- ① ここで提唱する評価方法は、DNA塩基配列だけでなくRNAやアミノ酸の配列データにも適用できる。さらに、形態的な形質にも同様に適用可能である。
- ② 樹長分布の計算は、D.L. Swofford (1993) の系統分析ソフトウェアPAUP, version 3.1を用いた。しかし、計算時間を短縮するためには、系統樹の樹形の完全網羅だけでなく、系統樹集合からの無作為抽出による近似的方法の開発が必要である。

## [具体的データ]

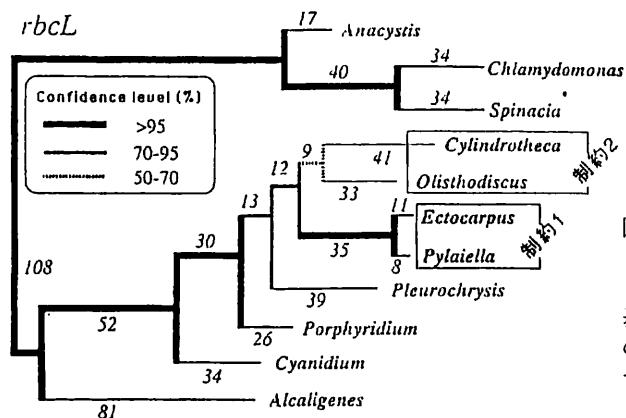


図1：藻類PrymnesiophytaのrbcL遺伝子塩基配列に基づく最節約系統樹。

各枝の太さは1,000回反復のブーツストラップに基づく確率。各枝の数値と相対的長さはその枝での形質進化（アミノ酸置換）の回数。樹形上の二つの制約のもとでの制約樹長分布を計算する。

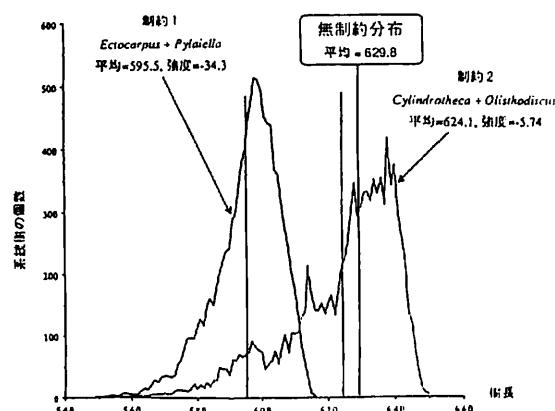


図2：図1の制約1、2のもとでの制約樹長分布とその期待値。

ある分類群の存在を樹形上の制約とする樹長分布の期待値が、無制約樹長分布の期待値よりも小さいならば、その分類群は系統樹全体の最節約性に貢献をしたと判定される。平均的に見て、制約1は制約2に比べて系統樹の樹長をより大きく短縮する。PAUP3.1を用いて樹長分布を計算した。

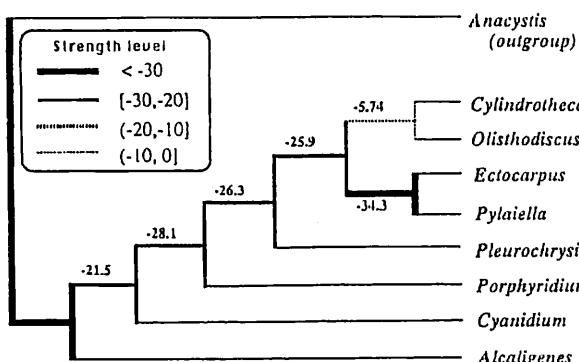


図3：最節約系統樹の枝の強度の計算。

図1のうち、ラン藻（Anacystis）のみを外群とする最節約系統樹を考え、rbcL塩基配列データに基づいて各枝の強度を計算した。枝の数値と太さは強度を表す。各枝の強度はPAUP3.1を用いて計算した。

## [その他]

研究課題名：生物情報の数量的分類理論の開発

予算区分：経常

研究期間：平成5年度（平成2年～平成5年）

研究担当者：三中信宏

発表論文等：Cladograms and reticulated graphs, Bulletin of the Biogeographical Society of Japan, 45卷1号, 1990年。組合せ論視点から見た系統推定, 千葉中央博自然誌研究報告, 2卷2号, 1993年。Structure and cotranscription of the plastid-encoded rbcL and rbcS genes of the marine haptophyte, *Pleurochrysis carterae*. Journal of Phycology, 29卷6号, 1993年。第29回万国地質学会議（1992年），日本植物学会第57回大会（1992年）などでも口頭発表。