

論文の内容の要旨

論文題目 Theoretical Studies on Molecular Phylogenetic Analysis

Based on a Large Number of OTUs and Loci

(多数のOTUと遺伝子座を用いた分子系統解析に関する
理論的研究)

三沢計治

現在、多数の生物・多数の遺伝子の配列データが蓄積されている。遺伝子配列の情報は、遺伝学に関する研究はもちろんのこと、系統分類学から疫学にいたるまで、幅広く利用されている。これからもDNA配列データは増え続け、その重要性は、今後ますます大きくなって行くであろう。そこで、これらの多量のデータを有効に解析する方法が必要となる。

私は、2種類のDNAデータ解析方法の研究を行なった。一つは、遺伝子が多数ある時に、それを分類する方法の研究であり、もう一つは、多数の遺伝子座を用いて正確に系統推定を行なう方法の研究である。

1. 多数の遺伝子座を用いた際の系統解析

最近、多数の遺伝子座の遺伝子配列から、分子系統樹を推定する事が多い。しかし、遺伝子によって、得られる系統情報の正確さに違いがある。多数のデータを使う際には、不正確なデータが系統推定の精度を下げることもあり得る。そこで、本研究では、距離法

を用い、多数の遺伝子座から、できるだけ正確に分子系統樹を推定する方法を開発した。

この方法は、次のような手順からなる。まず、それぞれの遺伝子座から距離行列を求める。その上で、それぞれの距離行列の正確さを評価し、その正確さに合わせてウエイトを与える。正確な距離行列には、大きなウエイトを、不正確な距離行列には小さなウエイトをかける。その上で、距離行列を足し合わせ、プールされた距離行列を一つ作り、そこから系統樹を作成する。

この方法の効果を調べるため、コンピュータシミュレーションを行った。図1に示したモデル樹にそって遺伝子配列を生成した。遺伝子座の数は10で、そのうち5つはDNA配列、残りの5つはアミノ酸配列であるとした。DNA配列の置換はKimura (1980)のモデルに従うとした。transition と transversion の比を4.0とした。アミノ酸配列においてはすべてのアミノ酸間で同じ確率で置換が起こるものと仮定した。それぞれの遺伝子の長さは100塩基対、または100アミノ酸残基であるとした。遺伝子座ごとの進化速度のばらつきは、ガンマ分布に従うものと仮定した。ガンマ分布の形はパラメータ α に依存し、 α が小さいほど、ばらつきが大きくなる。遺伝子座内のサイトごとの置換率のばらつきはないと仮定した。全体の分子進化速度の平均値を u とする。

この条件の下で生成された遺伝子配列から、DNAについてはKimura's(1980) two parameter 距離、アミノ酸についてはPoisson 距離を用いて、距離行列を求めた。そして、それらの距離行列の正確さを評価し、それに応じたウエイトをかけた上で距離行列をプールし、系統樹を再構築した。この試行を1000回繰り返して、モデル樹と同じトポロジーの系統樹が再構築された回数を調べた。Table 1にその結果を示してある。ウエイトをかけることによって、正しい系統樹を得られる確率が高くなることがわかる。また、遺伝子座間の進化速度のばらつきが大きいほど、ウエイトの効果が大きいことがわかる。

我々のウエイト方法を用い、ミトコンドリアにコードされている全遺伝子座を使って、哺乳類の系統関係を調べた。DNAについてはJukes & Cantor 距離、アミノ酸についてはPoisson 距離を使った。遺伝子座内の置換率のばらつきはないと仮定した。

図2がNJ法で書いた系統樹である。枝の上にある数字は、ウエイト法を使った場合、枝の下にある数字は、ウエイト法を使わなかった場合のブートストラップ確率である。4つのクラスターでは、ウエイトをかけることによって、ブートストラップ値が有意に増加

した。ウエイト法により、系統推定の誤差が少なくなることがわかった。

2. 遺伝子の分類方法

遺伝子配列を元にした分類は広く行なわれている。例えば、遺伝子は、遺伝子族に分類され、記載されている。ウイルスの分類は、疫学上必要だと考えられている。ヒトの歴史を推論するために、ヒトから取られたDNAサンプルがタイプ分けされることもある。しかしながら、遺伝子の分類を求める際には、統一的に利用できる方法がなく、混乱の原因となる事もあった。そこで、本研究では、広く使うことができる遺伝子の分類方法を開発した。

Klasterin (1982)は、Dendrogram を用いた病院の分類方法を提案した。今回、私は、Klasterin(1982)の方法を遺伝子の分類に応用した。ただし、病院の分類の場合とは異なり、遺伝子の分類は分子系統樹をもとに決められるものとする。この方法により、遺伝子は、共通祖先で起こった進化的変化が大きい単系統群へと分類される。この方法は、遺伝子の系統樹が得られている限り必ず遺伝子の分類を求める。この方法で分類を求めるためのコンピュータ上のアルゴリズムも開発した。そのため、遺伝子の数が多い時でも高速に分類を求めることができる。また、ブートストラップ(Efron and Tishirani 1991)に基づいて、分類の信頼性をテストする方法を開発した。

実際のデータに応用したのが、図3である。この図は、opsin の系統樹と今回開発された方法で得られた分類である。太線は、グループを分けている枝で、*は、系統推定の際に、高い bootstrap 確率($P>90\%$)で支持された枝である。数字は、それぞれのグループの bootstrap 確率である。この方法で得られた opsin の分類は、先行研究において吸収波長から得られた opsin の分類と一致した。

図1. モデル樹

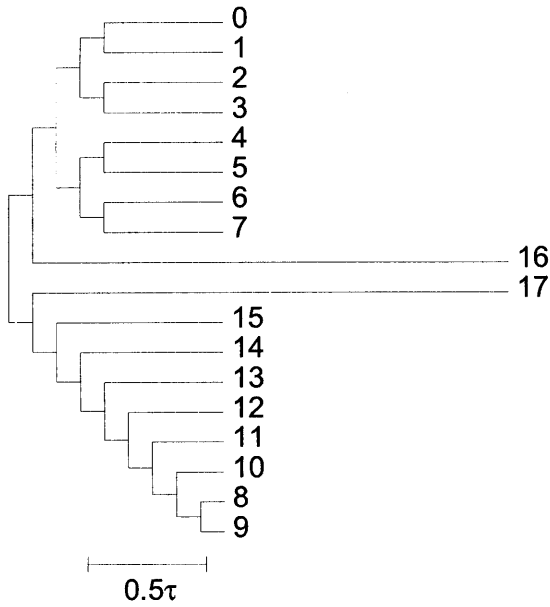
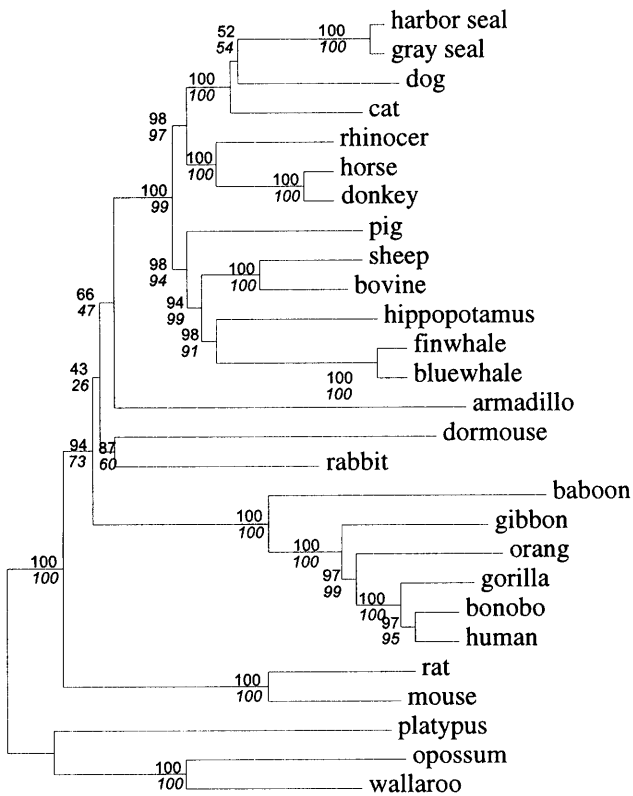


図2. mtDNAから求められた哺乳類の系統関係



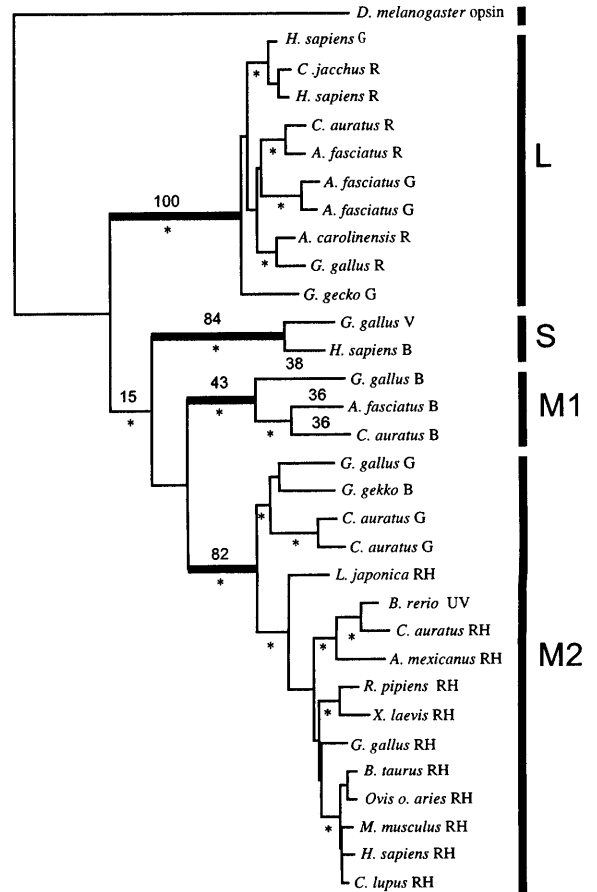
枝の上の数字はウエイトをかけた時のbootstrap確率であり、枝の下に数字はウエイトをかけたなかった時のbootstrap確率である。

Table 1

1000試行のうち正しいtopologyが推定された数

	α without weight	with weight
	$u=0.5$	
0.2	50	331
1.0	57	661
5.0	53	785
	$u=1.0$	
0.2	11	228
1.0	2	274
5.0	1	181

図3. opsinの系統樹と分類



グループを分けている枝は太線で示してある。*は系統推定の際に、高いbootstrap確率(P>90%)で支持された枝を指す。数字はそれぞれのグループのbootstrap確率である。