

Bioinformatics (1) 分子進化

津本 周作
島根大学医学部医療情報学
E-mail: tsumoto@computer.org
平成16年7月14日

分子進化の背景

- 進化のしくみ(系統)を、遺伝子レベルで探る.
- これまでの進化の研究は生物および「化石」の形態学的特徴から調べられてきた。
 - 形態学的特徴の類似性
 - 炭素の同位元素による年代推定
 - しかし、化石が見つからなければ、進化のしくみはわからない。
 - 現在の生物と絶滅した生物との間の関係からの推測となってしまう。

分子進化の背景(2)

- DNAは、現存生物から抽出でき、少なくとも、現存する生物の間関係は推測可能。
- クローニング技術の発達により、化石からDNAを抽出、複製できる場合がある。
 - 人骨
 - 琥珀内の昆虫
- 化石に残ったDNAから、進化のしくみを探る技術が発展してきた。

例: 類人猿とヒト

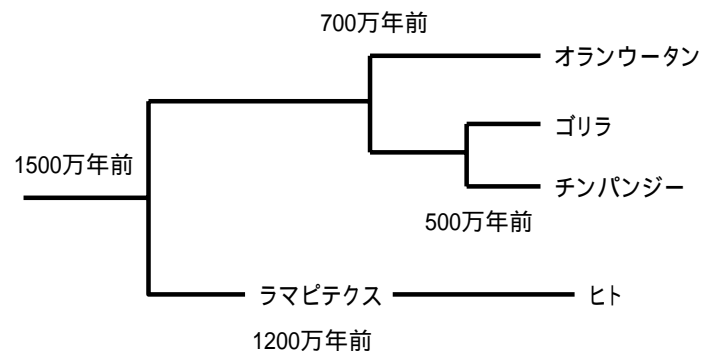
- 類人猿: チンパンジー、オランウータン、ゴリラ
- どれが一番人間に近いか?
 - 以前は形態学的特徴から「類似性」が、議論された
 - 骨(特に、頭蓋骨)の形
 - 形態学的な特徴点を計測し、その類似性を検討する (Biometrics, Morphology)

例: 類人猿とヒト (2)

- 類人猿: チンパンジー、オランウータン、ゴリラそれぞれの形態学的特徴を抽出
- 絶滅した原人の化石等と現存する類人猿との比較。
- 原人の化石の年代測定を考慮に入れ、進化的に分岐した時期を推測する。



例: 類人猿とヒトの系統樹(化石)



1960年まではこれが正しいと考えられてきた

ところが

- Nutall(1901) & Goodman(1961)の免疫拡散法
 - ウサギに比較したい動物の血清を注射
 - 抗体反応によって生じる沈殿の程度で動物の類縁関係が調べられるのでは?
 - どうも,化石学者の推定した系統関係はおかしいのでは?
 - チンパンジーとゴリラはヒトに近いと推定されるが,オランウータンはもう少し遠い.
- 大論争となる.

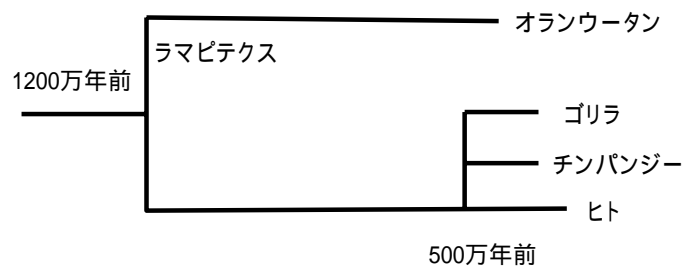
その後

- ZuckerkandlとPauling(1962)による分子進化速度の一定性の発見.
 - DNAやタンパク質といった情報分子は,時間の経過とともにほぼ一定の割合で塩基やアミノ酸の置換を蓄積する
 - これらの塩基の置換の度合いを見ることで,進化的な分岐時期を推測できる
 - 分子系統進化学の誕生
- SarichとWilson(1967)
 - ヒトにもっとも近い類人猿はチンパンジーで,分岐は500万年前と推測.

その後 (2)

- 木村資生(1968)による分子進化の中立説
 - Darwinの自然淘汰説との競合
 - 現在のDNA配列の解析結果は,中立説を支持している
- 系統樹の推測は推測にすぎないという化石学者の反論による激論
- 最終的には,1982年にパキスタンで発見されたシバピテクスの顔面の化石で決着
 - ラマピテクスの仲間
 - 顔面はオランウータンと共通の特徴

例: 類人猿とヒトの系統樹(分子)



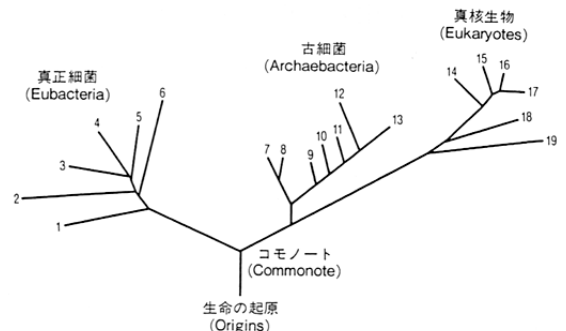
人類はアフリカから誕生したという説を支持

例2: 古細菌

- 高温,高塩,強酸といった異常環境で棲息可能な細菌
- リボゾームRNAに関する分子系統進化学的分類から,細菌とは異なる系統
- 上記の環境は,地球の原子環境と酷似しているため,「古細菌」と名付けた.
- ところが:

例2: 古細菌 (2)

- その後,詳細な検討によって,古細菌は真核生物に近縁であることが確認された (1989)



分子進化系統学

- 生物に比較的共通な酵素等の配列を比較することで、進化の様子を推測する
 - Fitch & Margoliash (1967)
 - 酵母からヒトに至るまでのチトクロームcを比較 それまでの系統樹と類似の系統樹が得られた。
 - Adachi & Hasegawa (1990)
 - ミトコンドリアDNAによる系統樹
 - Gojoboriら(1996): HTLV, HIVの系統樹
 - HIVの素のウィルスは、アフリカ奥地産

分子進化系統学の仮定: 中立説

- ダーウィン: 突然変異で、生存に有利なものだけが生き残る。
- 中立説
 - 突然変異: 不利な変異は自然淘汰で除去され、進化に寄与しない
 - 残りの変異
 - 集団に有利な変異
 - 有利でもない変異と不利でもない変異
 - “中立”な変異
 - DNA, タンパク質の置換をから、中立な変異が大部分を占める。
 - 中立な変異は「機会的不動」、すなわち偶然に集団に固定する。

分子進化系統学の仮定: 中立説 (2)

- DNA, タンパク質の置換をから、中立な変異が大部分を占める。
- 中立な変異は「機会的不動」、すなわち偶然に集団に固定する。
- 不利, 中立, 有位変異な変異は「確率過程」と考えられ、この中で中立な変異と有位な変異の蓄積が、「進化」と考えられる。
 - 中立な変異は、種内の変異をよく説明できる
 - 例えば、白色, 黄色, 黒色人種はいつ進化的に分かれたか?
 - 黒人との分岐: 11万年前
 - 白人との分岐: 7万年前 と推定される。

系統樹をどのように推定するか? (まずはMorphology)

- 類人猿: チンパンジー, オランウータン, ゴリラの形態学的特徴を抽出 (類似性の把握)
- 絶滅した原人の化石等と現存する類人猿との比較。 (類似性の把握と類型化)
- 原人の化石の年代測定を考慮に入れ、進化的に分岐した時期を推測する。 (進化時間の推測)

系統樹をどのように推定するか? (分子進化)

- 類人猿: チンパンジー, オランウータン, ゴリラ共通のタンパク質のアミノ酸配列, あるいはDNA配列の比較 (類似性の把握)
- 類似性から系統関係を推測 (類似性による類型化・グループ化)
- 置換数から進化的に分岐した時期を推測。 (進化時間の推測)

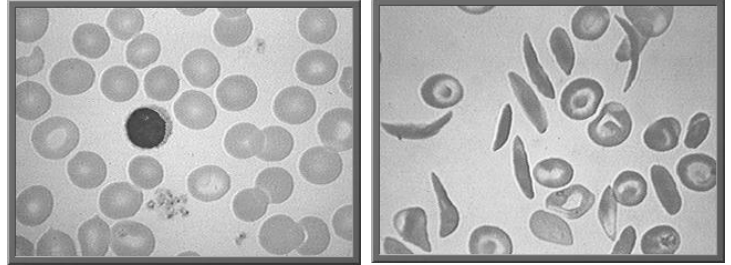
類似性の把握

- 配列解析
 - 例: -ヘモグロビン
 - ヒト VLSPADKTNVKAAWGKVGAAH.....
 - ウマ VLSAADKTNVKAAWSKVGGH.....
 - コイ SLSDKDKAAVKIAWAKISPKA.....
- 種が近いと配列の類似性は高まる
 - 種が遠いということは、進化的に分岐した時期が遠い。ことを予測。

類似性の把握 (2)

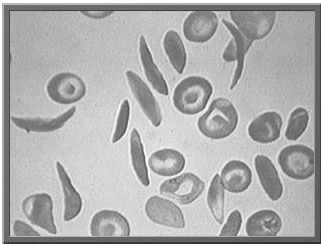
- 単純に、配列の一致数を「類似度」としてもよいが....
 - 配列の置換の意義には差がある:
 - ヒト: P, ウマ:A, コイ: D
 - P:フェニルアラニン, A: アラニン (疎水性)
 - D: アスパラギン酸 (親水性)
 - 親水性のアミノ酸が、疎水性のアミノ酸 になると、タンパク質の構造がはっきり変わることがある。
 - 例: 鎌状赤血球症.

アミノ酸置換の極端例: 鎌状赤血球症



- ある部位のアミノ酸がGlu(親水性) => Val (疎水性)に変化しているだけ.
- 常染色体劣性遺伝: アフリカ 400人に一人

アミノ酸置換の極端例: 鎌状赤血球症 (2)



- 溶血しやすい
- 貧血症状が著名
- 普通の環境であれば、自然淘汰される対象.
- 唯一の利点: マラリアにかかりにくい.
- マラリアに罹患しにくいという利点から、アフリカで淘汰されなかったと考えられている.

類似性の把握

- 配列解析
 - 例: -ヘモグロビン
 - ヒト VLSPADKTNVKAAWGKVGGAH.....
 - ウマ VLSAADKTNVKAAWSKVGGH.....
 - コイ SLSDKDKAAVKIAWAKISPKA.....
- 種が近いと配列の類似性は高まる
 - 種が遠いということは、進化的に分岐した時期が遠い. ことを予測.

類似性の把握 (3)

- アミノ酸配列の置換によって、同じタンパク質でも大きく構造が変わりうる。
 - 同義座位: アミノ酸の変化を伴わない座位
 - 非同義座位: アミノ酸の変化を伴う座位
- どちらの座位での置換かで、類似性の意味合いが異なる.
- 同義置換と非同義置換で類似度の計算法を変える必要がある。(類似度の補正)

類似性の把握 (4)

- コドン間座位数
 - 同義座位 P (UUU) A(GCU) 78
P (UUU) D(AAU) 49
 - 非同義座位 P (UUU) A(GCU) 222
P (UUU) D(AAU) 260
- コドン間置換数
 - 同義座位 P (UUU) A(GCU) 0
P (UUU) D(AAU) 47
 - 非同義座位 P (UUU) A(GCU) 200
P (UUU) D(AAU) 200

類似性の把握 (5)

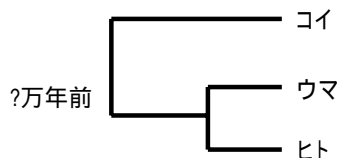
- 総同義/非同義座位数の算出: 差異のあるものに関して総加算
 - 総同義座位 P(UUU) A(GCU) 78
P(UUU) D(AAU) 49
 - 非同義座位 P(UUU) A(GCU) 222
P(UUU) D(AAU) 260
- 総コドン間置換数の算出: 差異のあるものに関して総加算
 - 同義座位 P(UUU) A(GCU) 0
P(UUU) D(AAU) 47
 - 非同義座位 P(UUU) A(GCU) 200
P(UUU) D(AAU) 200
- 同義座位相違度 $K_s(\text{ヒト, ウマ}) = 0/78$ $K_s(\text{ヒト, コイ}) = 47/49$
- 非同義座位相違度 $K_s(\text{ヒト, ウマ}) = 200/222$
- $K_s(\text{ヒト, コイ}) = 200/260$

類似性の把握 (6)

- 置換数の推定値
- ここでは, 一つしか置換座位がなかったと仮定すれば,
 - 同義座位相違度 $K_s(\text{ヒト, ウマ}) = 0/78$ $K_s(\text{ヒト, コイ}) = 47/49$
 - 非同義座位相違度 $K_s(\text{ヒト, ウマ}) = 200/222$
 - $K_s(\text{ヒト, コイ}) = 200/260$
 - 置換数(同義座位): $k(\text{ヒト, ウマ}) = 0$, $k(\text{ヒト, コイ}) = 2.31$
 - 置換数(非同義座位): $k(\text{ヒト, ウマ}) = 1.73$, $k(\text{ヒト, コイ}) = 0.76$
- この置換数が「進化距離」: 非類似度と見なせる.

類似性からの系統樹: 近隣結合法

- 同義座位相違度 $K_s(\text{ヒト, ウマ}) = 0/78$ $K_s(\text{ヒト, コイ}) = 47/49$
- 非同義座位相違度 $K_s(\text{ヒト, ウマ}) = 200/222$
- $K_s(\text{ヒト, コイ}) = 200/260$
- 置換数(同義座位): $k(\text{ヒト, ウマ}) = 0$, $k(\text{ヒト, コイ}) = 2.31$
- 置換数(非同義座位): $k(\text{ヒト, ウマ}) = 1.73$, $k(\text{ヒト, コイ}) = 0.76$
- この置換数が「進化距離」: 非類似度と見なせる.
 - 進化距離の小さいものから, 結べると考える.
- 同義座位の置換数からは:



分岐時期の推定

- 進化速度一定の法則
- ヘモグロビンでは 10^{-9} (/座位/年)であることが推定されている.
- 置換数(同義座位): $k(\text{ヒト, ウマ}) = 0$, $k(\text{ヒト, コイ}) = 2.31$
- ここでは, ウマとヒトとの差は推定できないが, コイとヒトでは 1.15×10^9 年と推定できる.
- 11億年前?
 - ここでは, 一つの座位だけで推定しているので, おおざっぱ

系統樹の推定法 (近隣結合法)

- 類人猿: チンパンジー, オランウータン, ゴリラ
共通のタンパク質のアミノ酸配列, あるいは DNA配列の類似度の把握 (確率モデル)
- 類似度の近い順からグループ化 (類型化)
- 類似度(進化距離)から分岐時期を推定 (進化時間の推測)
 - 進化速度はどう推定するの?
 - 確率モデルはどう与えるの?

系統樹の推定法 (近隣結合法) (2)

- 問題:
 - 進化速度の推定
 - タンパク質によって, 進化速度が異なる
 - できるだけ, 共通性の高いタンパク質を使う必要がある.
 - 例えば, ヘモグロビンは脊椎動物しか使えない.
 - 確率モデル
 - アミノ酸間, 遺伝子間の置換速度の推定
- 系統樹推定の問題
 - データから推測したが, 信頼できるか?

置換速度/進化速度

- すでにこれまでの解析結果の集積から:
 - 性質の類似したものは置換しやすい
 - 疎水性, 親水性
 - 性質の遠いものの置換には時間がかかる.
 - すでに置換に関しては, ある程度確立している.
- 機能的に重要な部分が少ないタンパク質ほど, 進化速度は速い.
- 機能的に重要な部分が多いタンパク質ほど, 進化速度は遅い.
 - フィブリノペプチド < ヘモグロビン < ヒストン

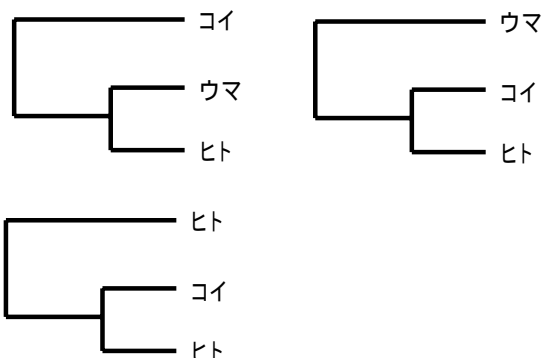
置換の性質から:

- 核では, 突然変異の他, 組み替えがおこる.
 - 組み替えは, DNA配列に著しい変異を与える.
 - 組み替えの効果を考慮した進化速度推定が必要
- ミトコンドリアのDNA
 - 組み替えは起こらない.
 - 配列の比較という観点からは, ミトコンドリアのDNAの方が望ましい.

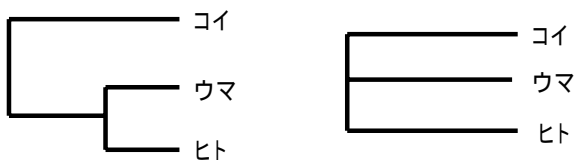
最尤法

- 類型化と進化距離推定とは, ある程度分離している.
- もう少し, 確率モデルから直接的に系統樹を推定できないか?
- 最尤推定法 (maximum likelihood)
 - Felsenstein (1981)
 - 尤度 (likelihood)
 - あるモデルから推定されるデータの出現確率
 - $P(\text{Observation} | \text{Model})$
 - モデル毎に確率(尤度)を計算し, どのModelが最もデータを説明できるかを評価する.
 - 通常は対数をとって, 対数尤度を計算

モデルの例 (同トポロジ)

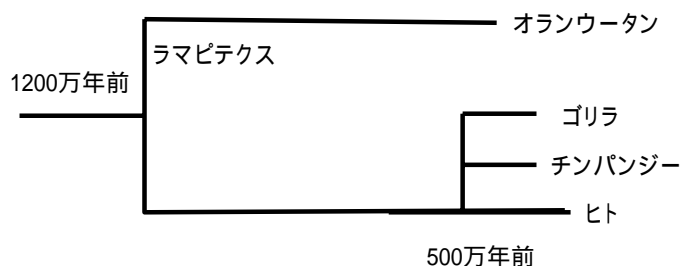


モデルの例 (異なるトポロジ)



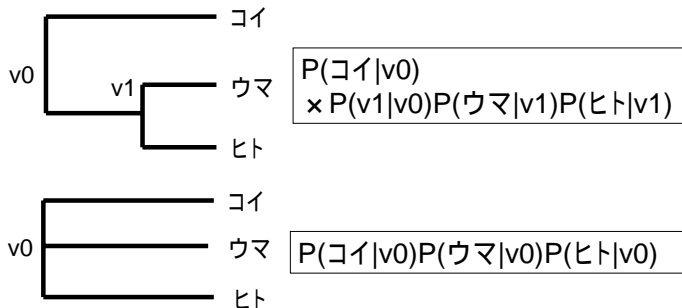
動物数が増えると, モデルの数は爆発的に増加
組み合わせ爆発: 最適解を見つけるのは困難.

例: 類人猿とヒトの系統樹(分子)



4つの場合の異なるトポロジはいくつあるか?

モデルから尤度の計算



分岐: それぞれの節で独立
Pは置換に関する確率モデルとDNAデータから計算

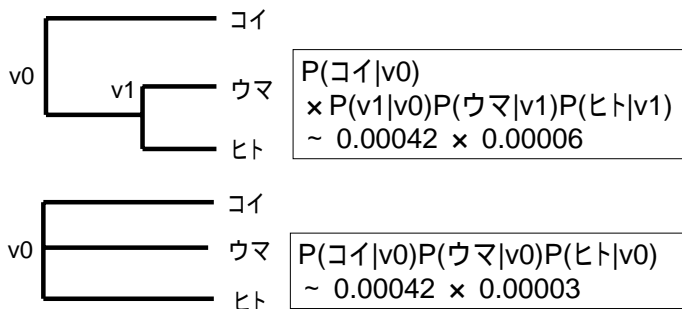
遷移確率

• JTTモデルのアミノ酸置換確率行列

- P A 0.00011
- P D 0.00003
- D A 0.00042
- D P 0.00003
- A P 0.00006

• 順番から言えば, D A, A Pかな?

モデルから尤度の計算 (2)



上の方が, 尤度の値は高い.
(実際には, v_0, v_1 の効果を補正する必要がある)

系統樹の推定法 (最尤法)

- 系統樹のモデルを作成 (確率モデル)
- 類人猿: チンパンジー, オランウータン, ゴリラ
共通のタンパク質のアミノ酸配列, あるいは
DNA配列の遷移確率の計算 (確率モデル)
- モデルに基づく尤度の計算 (モデルの評価)
- 最もらしいモデルを選択
 - 最適モデルから遷移確率の選択
 - 遷移確率から進化時間を推測 (進化時間の推測)

まとめ

- ゲノムプロジェクトによる, 遺伝子データの蓄積
- PCR等のクローニング技術の発達
- 配列の解析によって, 進化の系統を調べる
- 配列の類似性が, 種の類似性を反映
- 確率モデルを背景とした類似度に基づく類型化
- 類似度から, 確率モデルを逆に利用して, 分岐時期の推定
- まだ, 「完璧な」推定法がなく, 推定法は開発がすすめられている.