

情報システム工学科 平成 20 年度「自主課題研究」
テーマ：遺伝子配列の類似性に基づく生物種分類
3 学年 053 番 宮原 大貴

1. はじめに

遺伝子配列の類似性を用いて生物を分類し、実際の生物分類と符号するかどうかを調べた。類似性の計算には BLAST を使用し、分類にはフリーの統計解析ソフトウェアである R を用いて階層的クラスタリングを行った。

2. 実験に用いたデータ

解析にはデータベース”Uniprot”のデータをダウンロードして用いた。解析対象としては、Uniprot に登録されており完全ゲノムが決まっている哺乳類 8 種(ヒト:HUMAN、チンパンジー :PANTR、マカクサル:MACMU、ドブネズミ:RAT、ハツカネズミ:MOUSE、カモノハシ:ORNAN、オポッサム:MONDO)を用いた。解析対象の遺伝子は哺乳類 7 種全員が共通して持っているものを選択した。

3. 解析内容

① BLAST によるアラインメント

クラスタリングを行うための 距離(非類似度)を決定するために、最も普及している BLAST を使用してアラインメントを行った。アラインメントとは整列を意味しており、2 本の配列で対応するアミノ酸同士が並ぶように適宜ギャップを入れる動作である。

② R による階層的クラスタリング

共通している遺伝子は次の 6 個となった。

b2m: β 2 ミクログロブリン

cfr: 囊胞性線維性膜貫通調節因子

bdnf: 脳由来神経栄養因子

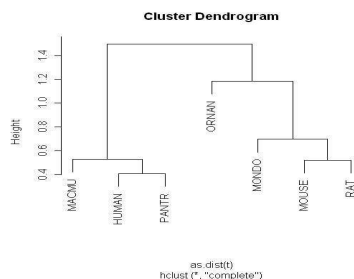


図 1. 全体的なクラスタリング結果

prnp: プリオン蛋白遺伝子

mt-cyb: ミトコンドリア・チクローム b

mt-nd4l: NADH-ユビキノン酸化還元酵素

これらの遺伝子それぞれについて、各生物種が持つ遺伝子配列のアラインメントを網羅的に行い、得られるスコアから距離を計算して距離行列を作り、それぞれの遺伝子に関するクラスタリングを行った。さらに、ユークリッド距離を用いて 6 種類の遺伝子に関する距離を統合した距離行列を作成し、クラスタリングを行った。

4. 考察

図 1 は 6 個の遺伝子における 6 次元ユークリッド距離でのクラスタリング結果である。ヒトを含めた猿系のクラスとネズミ系のクラスに分かれており、カモノハシはこの解析の観点からは哺乳類の中では他生物と類似性が低いことが分かる。また、クラス内でもヒトとチンパンジーは非常に近い関係にあるが、図 2 の prnp においては離れていることが分かりその生物間の違いの原因と考えることもできる。

5. まとめ・感想

1 個体の遺伝子データは大きなものでアラインメントをするのに大変時間がかかった。まず解析を行うには収集したデータを自分が使い易いようにする必要があり、またそれを行うプログラミング能力を身に付けておく必要を感じた。得られたデータに対しきちんと考察を行うことが解析ということを考慮してこれからに生かしたい。

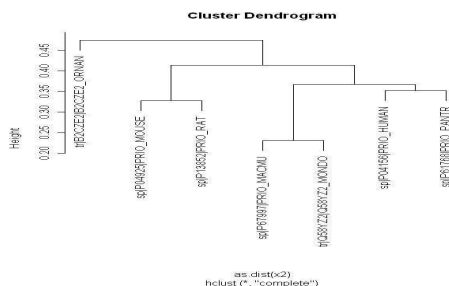


図 2. prnp のクラスタリング結果