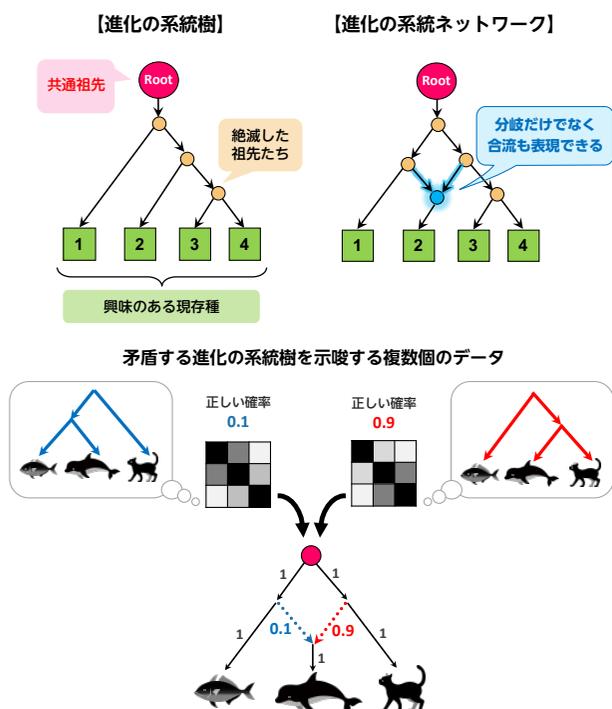




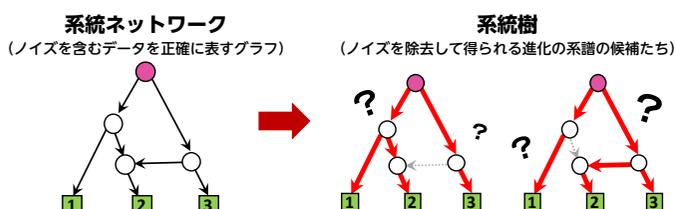
系統樹から系統ネットワークへ

「系統樹」は生物の進化を描出する手段として古くから使われているグラフ構造です。しかし、系統樹はあまりにも単純な進化のモデルであり、実際には生物の進化的な繋がりや網目のように複雑になることもあります。そのため、系統樹を拡張した「系統ネットワーク」というグラフ構造を活用したデータ解析手法を多くの生物学者が切望しています。系統ネットワークは系統樹より表現力に優れ、例えばノイズを含むデータや複数の系統樹の確率的重ね合わせなど、多様な現実の情報を表せます。



系統ネットワークの中に潜む系統樹

系統ネットワークで現実のデータを正確に描出したら、ノイズを除去して系統ネットワークの中に潜む真の系統樹を知りたくなります。しかし、真の系統樹は一通りに決まるとは限りません。



「系統ネットワークの構造定理」と系統樹推定に関する幅広い応用

系統ネットワークに潜む真の系統樹は一つに決まらないので、それらが全部で何個あるか(数え上げ問題)、リストアップせよ(列挙問題)、一番良いものを求めよ(最適化問題)、一番からk番まで順に求めよ(上位ランキング問題)という色々な問題が自然に思い浮かびます。

これらの問題の解き方(アルゴリズム)を個別に研究しても悪くはないのですが、異なる問題を統一的な視点で解く枠組みとなる理論基盤を作って、そこから色々なアルゴリズムを続々と生み出せたら物凄くインパクトがありますよね。

そこで私の研究では一般に系統ネットワークがどんな部品から構成されているかを解明する「系統ネットワークの構造定理」を示し、前述の全ての問題に対して、理論的に存在しうる中で最も高速なアルゴリズムを導出しました。なお、この定理の根底には代数学で学ぶ「有限アーベル群の構造定理」に通じる精神があります。どの自然数も素数の積で一意に表せるように(素因数分解)、どんな系統ネットワークもジグザグ型の単純なパーツに一意に分解され、その系統ネットワークに潜む系統樹全部の集合は各パーツ内における有向辺の選び方の集合の直積で表せるというのが要点です。詳しくは下記のYouTube動画や解説記事をご覧ください。

参考文献

- 【YouTube】2020年度応用数理概論 第26回動画: 「根付き二分系統ネットワークの構造定理と全域系統樹に関する色々な問題への応用」(※上のQRコード)
- 【雑誌連載記事】進化の系統樹とデータ解析(1)&(2), 『数学セミナー』2020年1月号&2月号

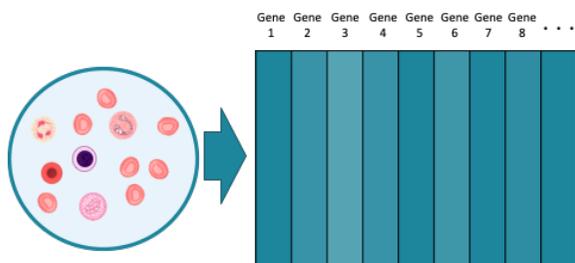
生物学的な新知見の宝庫とされる 1細胞の遺伝子発現データとは

早水研究室では生物の進化だけでなく、他にも色々なテーマに取り組んでいます。特に重点的に進めているのは生命科学の研究者から非常に大きなニーズのある「1細胞(シングルセル)」のデータ解析手法についての研究です。これは、植物学から再生医療まで、とても幅広い応用がある内容です。

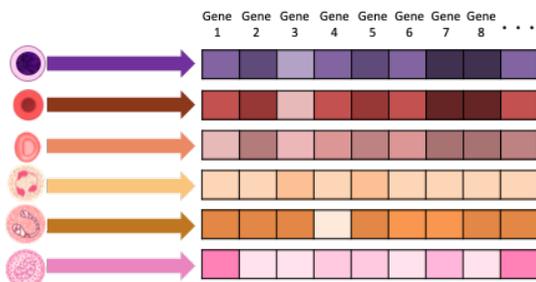
私たちの体を構成する細胞には外見も機能も異なる様々な細胞がありますが、それらの細胞の違いは、細胞内でどの遺伝子がどのくらい使われているかという遺伝子発現状態の違いによるものです。「遺伝子発現データ」というと、ごく最近までは全て「バルク」のデータ(色々な細胞が混在した細胞集団における平均的な遺伝子発現量のデータ)でしたが、この十年程で「シングルセル」の革新的な実験技術が普及し、今では細胞一つ一つの遺伝子発現量を測定し、ある時点での一つの細胞の遺伝子発現状態を一つの高次元ベクトルの形でキャプチャーできるようになりました。

このようなシングルセルの遺伝子発現データは新たな生物学的発見の宝庫として期待されていますが、既存のデータ解析手法では生物学的に意味のある情報を抽出できないことが多く、新たな数理的手法が強く必要とされています。

従来のバルクの遺伝子発現データ

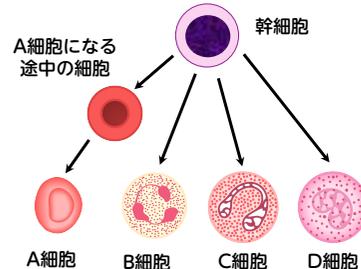


シングルセルの遺伝子発現データ



1細胞の遺伝子発現データを用いた 「細胞分化の軌跡」の推定

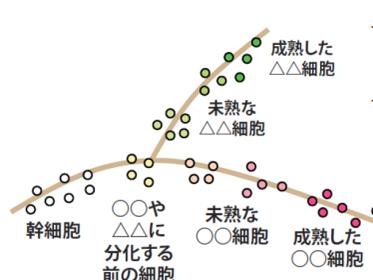
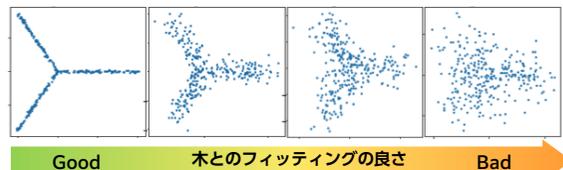
「細胞分化の軌跡」とは、細胞がどんな細胞を経て別の細胞に分化・成熟するかという道筋を表すグラフで、ヒトをはじめとする動物の細胞の分化の軌跡は木構造で表せるとされています。



細胞分化の軌跡を1細胞の遺伝子発現データから推定できれば多様な細胞同士の関係性を描出できるようになるので、細胞の分化という現象を制御する遺伝子を調べやすくなります。細胞の運命をコントロールする仕組みの解明は、例えば、ヒト幹細胞を所望の細胞に分化させて移植する再生医療などへの応用があるため、医学的・社会的にも重要性が高いといえます。

細胞分化の軌跡を推定・評価する 世界初の統計解析ソフトウェア

1細胞の遺伝子発現データから分化の軌跡を推定するツールはバイオインフォマティクスの研究分野で多数開発されてきましたが、数学的・統計学的に未開拓の事柄が沢山あるため、科学的に信頼できる統計解析手法といえるものはまだ確立されていません。早水研究室では、その先駆けとなる数理的手法を作って世界中の生物学者が使えるオープンソースのデータ解析ソフトウェアTreefitをリリースし、グラフ理論、幾何学、統計学、深層学習など、幅広い分野の知見を融合させ、さらなる改良や一般化などの新展開に向けた研究・開発を行っています。



詳しくは...
Treefitのwebサイト

