

# 理学のススメ

## 生命の進化・発生の過程を計算機で解く



今野 直輝  
Naoki Konno

(生物科学専攻 博士課程 1年生)

Profile

出身地 神奈川県  
出身高校 筑波大学附属駒場高校  
出身学部 東京大学理学部生物情報科学科

私たち生命は、沢山の細胞や遺伝子の機能から成るとも複雑なシステムだ。私は生き物が好きで中高の頃から生物学に興味を持ってきたが、生物学を学ぶにつれて、どうすればこんなによく出来たシステムが出来上がるのか？という問いに思いを馳せずにはいられなかった。この問いに答えるには第一に、生命が出来上がる過程、すなわち進化（1つの共通祖先から多様な生命が生まれる過程）や発生（1つの受精卵から多細胞の成体が出来上がる過程）の系譜を網羅的に明らかにすることが必要である。そこで私は、本学の理学部生物情報科学科で学びながら、進化や発生の系譜を調べる情報解析技術の研究を始めた。

生命の進化の系譜は進化系統樹、すなわち二分岐（種分化）を繰り返す木構造として表現できる（図1）。また発生の系譜は細胞系譜、こちらもやはり二分岐（細胞分裂）を繰り返す木構造として表現できる。これらの進化や発生の系譜情報は、実はどちらもDNA配列の情報から計算機を用いて推定することができる。例えば進化であれば、どの種とどの種のDNA配列がよく似ているのかを調べることで、種間の近縁度合いをもとに系譜の情報を推定できる。発生の過程も同様に、成体を構成する各

細胞のDNA配列を比較することで系譜を推定できるようになってきた。

しかし、この系譜を推定する計算には非常に長い時間がかかる。なぜなら、あり得る木構造の個数は、DNA配列数が多くなるにつれて爆発的に増大するからである。そのため従来の計算技術では約100万本を超える数のDNA配列に対しては系譜の推定が困難であり、進化・発生の過程の全体像を明らかにするための大きな障壁となっていた。

そこで私は、巨大な系統樹・細胞系譜情報を高速に推定できる計算技術、FRACTALを開発した。FRACTALは図2に示す計算サイクルを繰り返し、大量のDNA配列を次第に細かい近縁グループへと束ねていきながら、木構造を根本から末端に向かって決定していく（計算過程が「フラクタル」のようになる）。その際に、異なる配列グループたちを複数の計算機で同時並列に処理することで計算の大幅な高速化を達成した。その結果、実際に2億3500万配列という、従来の限界の200倍以上の配列数からでも二日足らずで正確な系譜を推定することに成功し、主著論文として報告した（2022年1月7日プレスリリース：<https://www.s.u-tokyo.ac.jp/ja/press/2022/7702/>）。

でもどうだろう。もし系統樹・細胞系譜が網羅的に明らかになったとして、それだけで私たちは生命というシステムが出来上がる過程を理解した、という気持ちになれるだろうか？きっとならない。なぜなら、その系譜は人間の頭で理解するには複雑すぎるからである。次に必要なのは、網羅的に明らかにした現象の背後に何か単純なルールを見つけ出すことだろう。そこで現在私は、明らかにされた進化の過程の中から、機械学習という計算技術を用いてパターンを見つけ出すという研究を行っている。特に原核生物と呼ばれる微生物の進化に注目し、どんな遺伝子をどんな順番で獲得・欠失する傾向があるのか？を調べている。進化のルールがわかれば、未来の生命進化（例：薬剤耐性菌の出現など）の予測もできるかもしれない。その実現に向けて、日々研究を進めている。

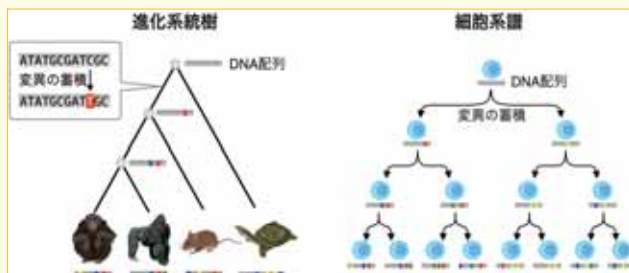


図1（左）進化系統樹と細胞系譜。生物のイラスト：© 2016DBCLS TogoTV / CC-BY-4.0

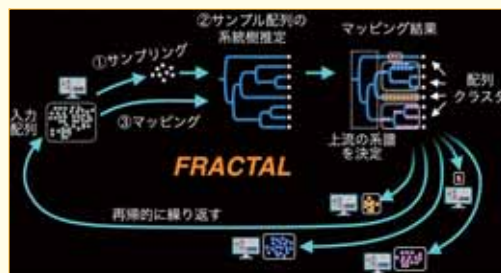


図2（右）FRACTALのワークフロー。図に示す計算サイクルを繰り返しながら、配列を互いに近縁なグループへと徐々に細かく分けていき、同時に系譜が上流から決定される