SS-1

プロローグ：次世代系統解析の問題と課題

田村浩一郎

首都大・生命情報研究センター、首都大・院理工

次世代 DNA シーケンサーの出現により、今までとは比較にならない量の配列データが容易に手に入るようになった。分子系統解析も次世代データを活用する段階に突入しつつある。しかし、同時に次世代データによる新たな問題も顕わになり、それらを解決するための系統解析技術の開発も喫緊の課題となっている。
これまでの系統解析では、あらかじめターゲットとなる遺伝子は決まっていて、PCR によって増幅、配列決定されていた。次世代シーケンサーを使うと、基本的にあらゆる遺伝子の配列が無作為に決定される。そこで、OTU 間で共通する相同部位を効率的に得る工夫や、得られたデータの中から有用な部分を抽出する作業が必要になる。前者には RNA-seq、RAD-seq が含まれ、後者では得られる全てのデータの中からオーソログのみを識別することが必須である。得られた配列データから系統関係を推定するためには、通常、多重配列アライメントを行う。多重配列アライメントは、むしろ樹形推定よりも難しい問題で、特にデータ量が多いと困難を極める。本夏の学校では、これらの問題に挑む最先端の研究を紹介し、これからの系統解析技術について議論したい。

SS-2

RAD-seq 法を用いた超高解像度系統解析

奥山雄大

科博・植物園

サンガー法による DNA 塩基配列決定法が普及し、標準化された手法で簡便に分子系統樹が得られるようになったことで、生物多様性の記述やその進化史の可視化は劇的に進んだ。一方で従来の分子系統学的手法の限界も明らかになって来た。特に、急速に適応放散し、高い種多様性を誇るような生物群はよりその進化プロセスが興味深いにもかかわらず、その系統関係の解明には困難を極めることが多い。しかしこのような従来困難であった系統解析の問題も、大規模並列 DNA シーケンサーの利用によって解決が期待できる。特に、RAD-seq 法はゲノム情報が未知の生物からでも大量の相同な塩基配列を取得できる点で、分子系統学との親和性が高い手法である。本発表では、すでに種の系統関係が既知であるチャルメルソウ属と、日本列島での著しい適応放散の結果としてこれまで種間の系統関係を解明することが困難であったカンアオイ属の2つの植物群で本手法を適用し、劇的に高解像度の系統樹が得られた事例を具体的な解析手法とともに紹介する。今後本手法は標準化し、さらなる技術の進歩も予想されるため、本手法が与える自然史研究へのインパクトと課題についても議論したい。

SS-3

全ゲノム重複と種の系統樹を考慮した系統解析マーカ遺伝子の選定

井上 潤

OIST

ゲノムワイドなデータを用いた系統解析が，未解決の分岐解明に期待されている．しかし，重複遺伝子を多く含むゲノムデータから，種の歴史を反映する系統解析マーカ遺伝子を選択するのは容易ではない．特に脊椎動物の系統解析では，2～3 回生じた太古の全ゲノム重複に由来するペア遺伝子の欠失パターンが遺伝子や種で異なるため，系統的枠組みを考慮したマーカ遺伝子の選定が必須である．そこで我々は，脊椎動物太古のゲノム重複のうち最も若い真骨類根幹で生じたゲノム重複に注目して，遺伝子系統樹を用いたマーカ遺伝子の抽出を試みた．真骨類を中心とした条鰭類 10 種と四足類 4 種の全タンパク質遺伝子配列を種の系統樹と比較解析したところ，約6000遺伝子について信頼性の高い遺伝子系統樹が得られた．このうち約1000遺伝子は，ゲノム重複直後にペア遺伝子の片方を失っただけでなく，全種で共有されるうえに系統特異的な重複遺伝子も持たなかった．これらの遺伝子を条鰭類の系統解析マーカとして報告する．我々の手法はゲノム重複を経験した他の生物の解析にも適用できる．この発表は佐藤，シンクレア，塚本，西田との共同研究に基づく．

SS-4

MAFFT による巨大アラインメントの計算、配列の対話的な選択などについて

○加藤和貴[1]、山田和範[2]、富井健太郎[3]、工樂樹洋[4]

[1]阪大・IFReC、[2]東北大・院情報、[3]産総研・創薬基盤、[4]理研・CLST

配列決定手法の進歩にともなって大量の配列情報が利用可能になってきており、多数の配列からなるアラインメントを計算する必要が生じている。ただ、多重配列アラインメントは必ず相同な配列からなるので、配列の数が増えると配列の間の類似性はたいてい高まる。この性質を利用すると、比較的簡単に大きなアラインメントを計算できる。(1) その計算のための、配列アラインメントプログラム MAFFT の新しいオプションを紹介する。(2) 信頼性の高い既存アラインメントに、個々の新しい配列をマップすることによって大量の配列の比較解析が可能なことが多い。そのためのオプションも紹介する。既存のアラインメントが、相同でかつ充分に多様な配列を含んでいる場合に有用である。(3) 大量の配列データを実際の系統解析に利用するためには、類似性の高い多くの配列から、意味のある配列セットを効率よく選ぶことが必要である。その作業を支援するウェブサービス（http://aleaves.cdb.riken.jp/aleaves/）を紹介する。