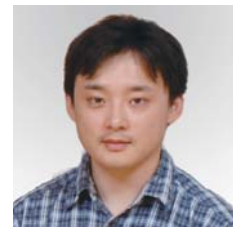


RNAの二次構造を考慮したマルチプルアライメント法の開発

木立 尚孝

RNA 情報工学チーム
産総研特別研究員

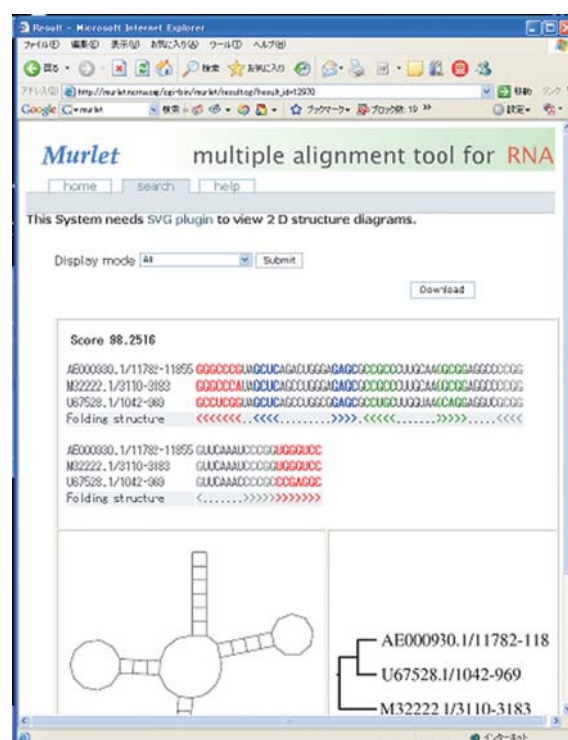


ヒトのゲノム配列が2003年に解読を終えたのは記憶に新しいところですが、その後もサル、ネズミ、イヌなどのゲノムが続々と解読され、現在までに30種類に迫る脊椎動物のゲノムが明らかになっています。一般にゲノム配列は、生命機能への影響が小さい領域ほど、世代を重ねるごとに、塩基置換による変化が蓄積していくと考えられています。このため、複数の動物のゲノムを相互に比較し、類似した領域を探すことで、ゲノムの中で種を通じて重要な役割を果たしている領域の見当をつけることができます。例えば、ヒトを含む哺乳類のゲノムで特に類似度の高い領域は、ヒトゲノムのおよそ5%程度の長さにとわって存在することが知られています。これらの領域は、タンパク質をコードする遺伝子領域など、これまでにその役割が少しでも知られている領域を30%含んでいますが、残りの70%は、従来の研究では機能が明らかでなかった未知領域からなり、生命活動において果たす役割の解明が望まれています。この機能未知領域の役割の候補としては、転写はされずにタンパク質の転写因子などが結合することで転写制御を行うシスエレメントや、RNAとして転写されてタンパク質やメッセンジャーRNAと相互作用する非コード機能性RNA遺伝子などが挙げられます。特に後者は、ゲノムワイドな転写発現解析実験からヒトゲノムの大部分がRNAへ転写されると分かったことや、マイクロRNA、siRNAなどによる新しい遺伝子制御システムが明らかになったことから、近年非常に注目を集めています。これまで良く知られている機能性RNAには、配列の塩基同士が対を組むことで形成される二次構造と呼ばれる特殊な構造をとり、塩基配列にその構造を反映した統計的な特徴があるものが多く存在します。従って、ゲノム内の機能未知領域にも、このような二次構造をとる機能性RNA遺伝子が多数含まれていることが予想されます。私たちは、この二次構造の配列的特徴を利用したゲノム配列解析アルゴリズムの研究およびそれを用いたゲノムの機能解析研究を行っています。多数の生物種のゲノムを比較する際に欠かせない道具のひとつに、生物種間で進化的に関係のある部分配列同士を対応付けるマルチプルアライメントツールがあります。構造RNA遺伝子領域では、ゲノム進化の中でその構造を保つような特殊な塩基置換が起こるため、そのような置換を適切に取り扱えるモデルを用いてアライメントを行う

ことで、従来のモデルによるものより正確なマルチプルアライメントを構成できることが知られています。このような二次構造を考慮したアライメント法は1980年代にSankoffにより考案されましたが¹、その膨大な計算時間のために実用的なものとは見なされてきませんでした。私たちは、最近その有効性が確かめられつつあるMaximal Expected Accuracy 原理に基づいた、非常に効果的なアライメント評価法を考案し、精度を落とすことなく実用的な速度で二次構造の進化を考慮したアライメントの計算を行うことを可能にしました^{2,3}。現在はこの方法を用いてヒトゲノムから新しい機能性RNA領域を発見する研究を行っています。

Reference

- 1 Sankoff, D.: "Simultaneous solution of the RNA folding, alignment and protosequence problems", *SIAM J. Appl. Math.*, **45**, pp.810–25 (1985).
- 2 Kiryu, H., Tabei, Y., Kin, T., Asai, K.: "Murlet: A practical multiple alignment tool for structural RNA sequences", *Bioinformatics*, (2007) (in press). doi: 10.1093/bioinformatics/btm146
- 3 Murlet ウェブサイト: <http://murlet.ncrna.org/>



Murlet (<http://murlet.ncrna.org>)