

論文内容の要旨

論文題目 Multiple Alignment for Structural RNA Sequences
(構造 RNA のマルチプルアラインメントに関する研究)

氏名 木立尚孝

(論文内容の要旨)

ゲノムワイドな転写発現解析により、高等真核生物の細胞内では大量のたんぱく質をコードしない RNA が発現していることが明らかになってきた。これらの転写物の機能的な役割を明らかにするために、ゲノムスケールでマルチプルアラインメントを収集し、その中から保存した二次構造をもつものを発見する研究は重要である。本論文では、構造 RNA 配列のマルチプルアラインメントの問題についての研究を行った。

本論文の前半部では、構造 RNA 配列のマルチプルアラインメントを行う新しいアルゴリズムについて述べた。構造 RNA は、その二次構造を保存するような進化のパターンを示すため、構造 RNA の精度の高いマルチプルアラインメントを生成するためには、この独特な進化パターンをアラインメントモデルに取り込むことが必要である。そのために塩基対の共置換を考慮に入れてアラインメントを行う新しいアルゴリズムを考案した。このアルゴリズムは、アラインメントの精度を落とすことなく、計算量を大幅に削減することを可能にする、効果的なスコア関数を備えている。このスコア関数の主要部分は、アラインメントされる二本の配列で、塩基同士のアラインメントされやすさを表すマッチ確率と、個々の配列の異なる位置の塩基対の組みやすさを表す塩基対確率行列の積として表される。従来のアラインメントプログラムに比べて、このアルゴリズムは、アラインメント精度の点においても、アラインメント結果から予測される二次構造の精度においても優れていることが、いくつかの計算機実験の結果から示された。さらにこのアルゴリズムにより、従来の二次構造を考慮するアラインメントプログラムでは困難だった、比較的長い RNA 配列、例えばおよそ 300 塩基の長さを持つ eukaryotic type signal recognition particle RNA などのマルチプルアラインメントを行うことも可能になった。

本論文の後半部分では、構造 RNA のマルチプルアライメントから RNA の共通二次構造を予測する、新しいアルゴリズムについての研究を行った。従来の同様のプログラムが、入力されたマルチプルアライメントが正確であることを仮定して、共通二次構造を予測するのに対し、この新しいアルゴリズムでは、アライメントが完全に正確であることを仮定せず、より多くの配列に支持される二次構造を予測することで、アライメントのエラーに対してロバストに共通二次構造の予測を行う。具体的には、この方法では、まず個々の配列について、離れた塩基位置の塩基対の組みやすさを表す塩基対確率行列をもとめ、その行列の値をアライメントの情報に基づき平均化を行い、その後平均化された塩基対確率の和を最大とするような動的計画法を行って共通二次構造予測が行われる。計算機実験を行うことにより、アライメントの質が低いときや配列本数が多い場合に、この方法は、従来の方法よりも精度良く共通二次構造の予測を行うことが示された。

| | |
|----|------|
| 氏名 | 木立尚孝 |
|----|------|

(論文審査結果の要旨)

1月11日に開催した公聴会の結果を参考に、2月13日に本博士論文の審査を行った。以下のとおり、本博士論文は、提案者が独立した研究者として、研究活動を続けていくための十分な素養を備えていることを示すものと認める。

木立尚孝は、本博士論文において、構造 RNA 配列のマルチプルアライメントを効率よく行う新しいアルゴリズムを提案し、実装して、その優れた実行効率を評価した。

本論文では、まず塩基対の共置換を考慮に入れてアライメントを行う新しい構造化 RNA 配列アルゴリズムを提案している。このアルゴリズムは、アライメントの精度を落とすことなく、計算量を大幅に削減することを可能にする、効果的なスコア関数を備えている。従来のアライメントプログラムに比べて、このアルゴリズムは、精度の点においても、結果から予測される二次構造の精度においても優れていることを、いくつもの計算機実験の結果から示した。さらにこのアルゴリズムにより、比較的長い RNA 配列、例えばおよそ 300 塩基の長さを持つ eukaryotic type signal recognition particle RNA などのマルチプルアライメントを行うことも可能になった。

本論文ではさらに、構造 RNA のマルチプルアライメントから RNA の共通二次構造を予測する新しいアルゴリズムについての研究を行っている。このアルゴリズムでは、アライメントが完全に正確であることを仮定せず、より多くの配列に支持される二次構造を予測することで、アライメントのエラーに対してロバストに共通二次構造の予測を行う。計算機実験により、アライメントの質が低いときや配列本数が多い場合に、この方法は、従来の方法よりも精度良く共通二次構造の予測を行うことが示されている。

構造 RNA 配列のマルチプルアライメントを効率よく行う新しいアルゴリズムとその評価に関する本研究は、独創性が高く、しかも実用的であり、バイオ情報学の分野において高い貢献があると評価する。

よって、本論文は、博士（理学）の学位論文として価値あるものと認める。