

Title	Evolutionary and Information Theoretical Analysis of RNA Splicing Mechanisms and its Application to Improvement of a Spliced Alignment Program(Abstract_要旨)
Author(s)	Iwata, Hiroaki
Citation	Kyoto University (京都大学)
Issue Date	2011-03-23
URL	http://hdl.handle.net/2433/142117
Right	
Type	Thesis or Dissertation
Textversion	none

(続紙 1)

京都大学	博士 (情報 学)	氏名	岩田 浩明
論文題目	Evolutionary and Information Theoretical Analysis of RNA Splicing Mechanisms and its Application to Improvement of a Spliced Alignment Program (RNAスプライシング機構の進化的・情報論的解析ならびにそのスプライスアラインメントプログラム改良への応用)		
(論文内容の要旨)			
<p>本論文は、互いに関連するふたつの課題について論じており、全7章により構成される。前半は、RNAスプライシング機構に関わるRNA塩基配列上の特徴についての進化的・情報理論的な解析に当てられている。一方、後半では、その知見に基づくスプライスアラインメントプログラムの精度向上について述べている。</p> <p>第1章は序章であり、研究の背景と目的、目的達成のための方針および本論文の概要を示している。</p> <p>第2章は、本論文に記述される研究内容を理解するために必要な生命情報学に関する最小限の知識をまとめている。</p> <p>第3章および第4章が、冒頭で述べた本論文主要部分の前半に相当する。真核生物においてRNAスプライシングの基本的なプロセスは保存されており、イントロンの5' および3' 末端 (5' ssと3' ss) ならびにブランチポイント (BP) の3者がスプライシングに直接関与するシグナルとしてはたらく。また、イントロンの長さやイントロン内部の塩基組成などもイントロンの認識に関与していると考えられている。そこで、申請者は、動物、植物、菌類、原生生物にまたがる様々な真核生物61種のゲノム配列とそれらの転写産物 (cDNA, EST) のDNA配列を用いて5' ss、3' ssおよび、BPを同定し、進化と情報理論の観点からスプライシング機構の解析を行なった。第3章はそのための詳細な方法、手続きを記述している。解析の結果は第4章にまとめられており、以下のように要約できる。5つの特徴量 (5' ss, 3' ss, BP, イントロン長、イントロン内塩基組成) ごとに、生物種間のシグナルの類似度を反映する距離を求めた。それを基に作成したデンドログラムと、18S rRNA, あるいはU2 RNA配列に基づく標準的な系統樹とを比較したところ、5' ss > 3' ss ≧ イントロン長 > BP > 塩基組成の順に保存性が高いことが分かった。相対エントロピーより求めた情報量は、どの特徴も少なからずイントロン認識に貢献していることを示した。特に5' ssと3' ssのシグナルが最も重要な役割を果たしていることが確かめられた。また、イントロン長やイントロンの塩基組成の貢献度は種や近縁グループごとに大きな違いがあることが分かった。例えば、哺乳動物ではこれらの貢献度が低いのに対し、菌類や原生生物ではかなり重要な役割を果たすことが判明した。また、どの生物種でも、5つの特徴だけでは正確なイントロン認識にとって不十分であることも判明した。5つの特徴がイントロン認識に及ぼす相対的貢献度に情報の不足分を加えた6変数を用いてクラスタ</p>			

リング解析を行った結果、一般に近縁の生物種は類似のパターンを示すものの、相対的貢献度は進化の過程で比較的变化しやすい性質を持つことが分かった。

第5章および第6章が本論文主要部分の後半を占める。近年、DNA配列決定技術の発達に伴い大量の転写産物配列やゲノム配列が決定されている。これらの文字列情報から生物学的に有用な知見を得る第一歩はゲノム配列上の遺伝子を同定することである。真核生物の遺伝子同定のためにはいくつかの手法が用いられるが、そのうちスプライスアラインメント法は遺伝子構造を最も精度良く認識できる手法として知られている。この方法を大量データに適用するためには高速で高精度のアルゴリズムの開発が必要となる。この方法では、エキソン・イントロン構造を考慮しながら目的スコアを最大化するアラインメントを求める。そのために、単に配列間の塩基同士の一致だけではなく、エキソン・イントロン境界のシグナルやエキソンらしさ、イントロンらしさなどの統計量も考慮しなければならない。申請者は上述のRNAスプライシング機構の解析から得られた知見を基に、スプライスアラインメントプログラム「Spaln」の性能向上を図った。具体的には、従来のSpalnでは用いられていなかったBPシグナルとイントロン内の塩基組成を新たに目的スコアの一部として導入した拡張版を開発した。この拡張版Spalnと他の多くの既存プログラムとの性能を様々な条件下で比較した。そのため、まずヒト、シロイヌナズナ、アカパンカビそれぞれにつき三種類の高信頼性ベンチマークデータセットを作成し、詳細な精度評価に用いた。その結果、この改良Spalnはほとんどの現実的状況下で全ての既存プログラムより高い精度を示した。特に植物や真菌類において、原典版と比べても有意な性能の向上が認められた。

最終章である第7章は議論と結論であり、これまでの章では尽くされなかったやや例外的な話題について触れるとともに、本研究で達成できたことおよび将来の課題として残されたことが簡潔に議論されている。

(論文審査の結果の要旨)

本論文は、真核生物のRNAスプライシング機構に関わるバイオインフォマティクス研究について述べており、その主要な成果は次の2点である。

(1) 多様な真核生物のRNAスプライシング機構に関する進化的、情報理論的解析。

動物、植物、菌類、原生生物にまたがる多様な真核生物において、RNAスプライシングの基本的なプロセスは保存されているが、ヒトのRNAを酵母が正しく処理できないなど、進化的な不適合性が存在することも知られている。しかし、スプライシングに関わるRNA配列上のシグナルに関し、様々な生物種に対して定量的な解析を行う試みは従来ほとんどなされてこなかった。本論文では、61種類の多様な真核生物の遺伝子からイントロンを抽出し、イントロン認識に関わるシグナルの強度および異なるシグナル間の相対的貢献度を情報理論に基づき定量化し、その結果を進化的に分析した一連の試みについて記述している。これは当該分野における世界初の大規模な定量的解析といえる。5' , 3' スプライスサイトの重要性など、小規模の解析から得られていた従来の知見を確認するとともに、ブランチポイントと3' スプライスサイトとの補完的關係、イントロンサブタイプ間の進化的保存性、ゲノム配列の塩基組成とサブタイプ間相対頻度との関連性などを見出すなど、これまで知られていなかった生物学上の発見も行っている。RNAスプライシングという、極めて複雑な細胞内過程について、新たな視点を与えたことは評価に値する。

(2) スプライスアラインメントプログラムの改良。

スプライスアラインメント法は真核生物の遺伝子構造を最も精度良く認識できる手法として知られ、ゲノム注釈のために広く用いられている。従って、スプライスアラインメントプログラムの性能向上は重要なバイオインフォマティクスの課題の一つである。上述したように、RNA配列上のスプライシングシグナルの強度や相対的貢献度は、生物種により大きく変動し、種によっては5' , 3' スプライスサイト以外のシグナルも相当の貢献をしていることが判明した。これをふまえて、本論文では従来のスプライスアラインメントプログラムで考慮されていなかったシグナルを取り入れた改良法の開発に取り組んでいる。その結果、植物や菌類において、大きくはないが有意の精度向上が達成できた。

以上、本論文は、RNAスプライシング機構に関する大規模な解析を行い、結果として、スプライシングの分子機構に関する理解に寄与するとともに、スプライスアラインメントプログラムの改良を通じて情報学的にも相応の貢献を行っている。よって、本論文は博士(情報学)の学位論文として価値あるものと認

める。

また、平成23年2月7日に実施した論文内容とそれに関連した試問の結果合格と認めた。