

機関番号：12601

研究種目：若手研究（B）

研究期間：2009～2010

課題番号：21700330

研究課題名（和文）ヒトゲノム上の Alu 配列の二次構造進化

研究課題名（英文）Secondary structure evolution of Alu sequences in the human genome

研究代表者

木立 尚孝（KIRYU HISANORI）

東京大学・大学院新領域創成科学研究科・准教授

研究者番号：80415778

研究成果の概要（和文）：RNA 二次構造アルゴリズム使い、RNA 配列から多種多様な構造情報を抽出するための基礎技術の開発を行った。また、進化系統樹の理論を応用して、これまでに計測されることがない進化的保存度を計算するアルゴリズムの開発を行なった。

研究成果の概要（英文）：We have developed technologies to extract various structural information in RNA sequences by using the algorithms of RNA secondary structure prediction. We have also developed algorithms for computing novel kinds of evolutionary conservation measures by applying the theory of phylogeny.

交付決定額

（金額単位：円）

	直接経費	間接経費	合計
2009年度	1,400,000	420,000	1,820,000
2010年度	600,000	180,000	780,000
総計	2,000,000	600,000	2,600,000

研究分野：バイオインフォマティクス

科研費の分科・細目：情報学・生体生命情報学

キーワード：進化系統樹、Alu配列、RNA二次構造

1. 研究開始当初の背景

ヒトゲノム中でタンパク質をコードしない領域が占める割合は98%にもものぼるが、この領域の生物学的機能についての我々の理解は未だ限られたものでしかない。これらタンパク質をコードしない領域のおよそ半分はLINEやSINE、LTRなどの反復配列からなっており、なかでも100万を超える反復数をもつAlu配列ファミリーはヒトゲノムの10%もの領域を占めている。Alu配列は、一度RNAとして転写されてから外部のTransposaseなどのタンパク質によりゲノムに挿入されることで転移（Transposition）をおこす、非LTR型、SINE型のレトロトランスポゾン的一种である。その起源はSignal Recognition P

articleの構成要素である7SL RNA遺伝子であり、Primateのゲノム進化の過程で爆発的な反復数を獲得した。転写されたAlu RNAは7SL RNA同様強い二次構造を形成することが知られている。

Alu配列は1970年代に発見されて以来、長いこと生理活性を持たない「ジャンク」なゲノム要素だと考えられてきたが、ここ数年の間に、Alu RNAの生理機能の一端が明らかになってきた。例えば、細胞に様々なストレスがかけられたときに20-100倍に発現量が増加すること、翻訳開始点でmRNAの翻訳の活性化や不活性化を制御すること、PolyIと複合体をつくり、転写制御を行なうことなどが示されている。これらの結果は、Alu配列がれっきとした機能性・構造RNA遺伝子的一种であり、

tRNAやrRNAなどに劣らない重要な生理機能をもっていることを意味している。Alu配列の解析の歴史は長いが、これまで構造RNAとして働くAluRNA遺伝子が、Alu配列全体の中でどのような分布をとっているかを調べた研究はなかった。

2. 研究の目的

研究の目的は、比較ゲノム的手法とRNAの二次構造解析の手法を用いてAlu配列の構造RNAとしての配列進化の形態を明らかにすることである。また、その目的を達成する過程で、一般の反復配列や構造RNAの進化・機能解析に役立つ実用的な配列解析アルゴリズムとツールを開発することも目的としている。

(1) Alu配列群の網羅的サブファミリー分類：反復配列は、一般のタンパク質遺伝子の進化とは大きく異なる進化形態をとる。例えば、タンパク質遺伝子では、遺伝子ファミリー内のどの遺伝子もほぼ均等に遺伝子重複を起こす可能性があるが、反復配列の場合は、少数のマスター遺伝子が、短期間の間に何百回も転移を起こしたのち転移の不活性化を起こす、という過程を繰り返す。ここで、同じマスター遺伝子を起源とするAlu配列群（Aluサブファミリー）を網羅的に同定することが研究の最初の目標である。このために、100万本の配列に対しても適用可能な、配列相同性によるクラスタリングツールの開発を行う。このツールとコンセンサス配列に基づく従来からの分類法を合わせて用いることで、Alu配列の網羅的なサブファミリー分類を行なう。

(2) 構造RNAとしての共塩基置換速度の解析：ゲノムに100万のAlu配列が存在するといっても、その全てが構造RNAとしての機能をもつとは考え難く、ほんの少数のAlu配列のみが、機能性RNAとして働いていると考えるのが自然である。従って、構造RNAを保存するような進化的圧力は少数のAlu配列にのみ働き、その他のAlu配列には、二次構造を破壊するような置換が蓄積していくと思われる。構造RNAの場合、塩基対を組む2塩基は、置換後も対を組めるように共置換を起こすことが知られている。このような構造RNAの進化的性質を利用して、構造RNAとして機能を持つAlu配列の検出を行なう。

(3) 最後に、これらの結果を総合して、Alu配列の転移と構造RNAとしての進化の間の相関関係を明らかにする。

3. 研究の方法

AluRNAだけではなく、一般的な構造RNAの進化を解析する手法は未発達の状態にある。このため、まずRNA配列から様々な構造的特徴を抽出するアルゴリズムの開発を行う。

またそれと同時に、進化系統樹の理論を応用して、RNA配列に働く進化的制約を精密に見積もる事ができるアルゴリズムの開発を行う。

人には100万コピーを超えるAlu配列要素が存在する。これら大量の配列から二次構造的、進化的情報を抽出するための効率の良いクラスタリングアルゴリズムを開発する。

4. 研究成果

(1) RNA二次構造アルゴリズムを使い、与えられたRNA配列からできる限り多様な情報を抽出するための基礎技術の開発を行った。まず、前年度に作成した、Raccessを用いた、miRNAとsiRNAのターゲットサイト近傍のアクセシビリティと抑制効果との関係を調べる研究を論文誌に出版した。

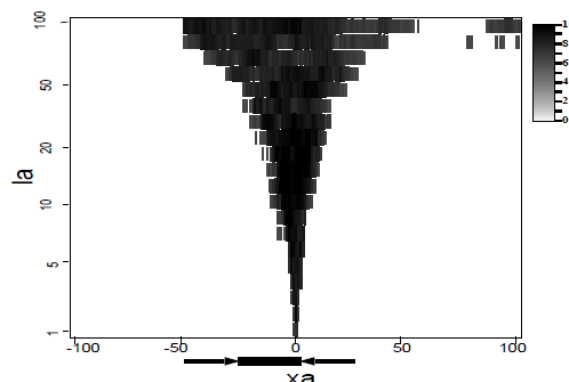
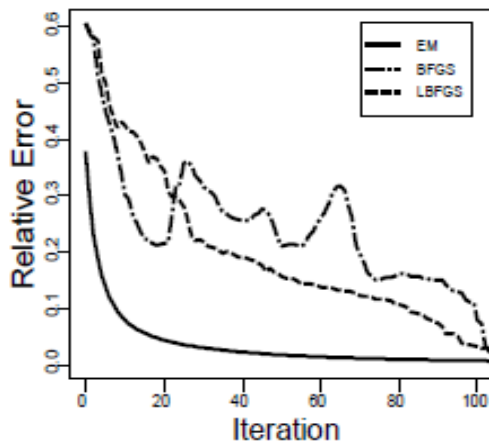


図1：siRNAのターゲット領域のアクセシビリティと、siRNAの発現抑制効果の相関、横軸は、siRNAのターゲット領域（矢印で挟まれた領域）を基準としたmRNA上の位置を表す。縦軸は、開かれているRNA領域の長さを表す。色の濃さは、開かれたRNA領域のパターンごとに、その領域のアクセシビリティとsiRNAの抑制効果の強さとの間にどの程度相関があるかを示している。

このRaccessを使うことで、siRNAのターゲット領域の3'端が開かれた状態にあることで、siRNAの抑制効果が高くなることなどが示された（図1）。構造RNAとしての機能をもつAluRNAは、強い二次構造をとると考えられるため、アクセシ

ビリティの強さを調べることで Alu RNA の RNA としての機能の欠失の分析に使えると予想している。

(2) 塩基配列の保存性の計算に使われる連続時間マルコフモデルについての研究を行ない、モデルの完全な十分統計量と期待値最大化 (Expectation Maximization) アルゴリズムを完全な形で世界で初めて導出した。この期待値最大化アルゴリズムを用いて系統樹のパラメータを最適化する方法は、従来までの勾配法を用



いる方法よりも収束が2倍程度速く、また最適解へ向かって後戻りなしに収束するよい性質をもつことが示された。(図2)

図2: 我々のEMアルゴリズムとその他のアルゴリズムとの最適化計算の収束の速さを示した。横軸は、期待値計算の回数を表し、縦軸は真の解と各繰り返し回数におけるパラメータ値との距離を表す。一番下の線が我々のアルゴリズムを表し、従来の手法より、効率良く最適解に近づいていくことがわかる。

この完全な十分統計量には、従来の期待塩基置換数 (Number of Expected Substitutions) の他に、期待塩基滞在時間率 (Expected Fractional Duration) が含まれることを明らかにした。

この期待滞在時間率をゲノムスケールで計算したところ、保存性が低いにも関わらず、期待滞在時間率が低い遺伝子上流領域をもつヒト遺伝子には、体の形成に関わる遺伝子が多く含まれることなどが明らかになった。この研究についても論文誌に掲載された。

(3) 二次構造エネルギーのボルツマン分布に基づき、RNA配列の二次構造の熱力学的エントロピー及び平均自由エネルギーを計算するツール (Rentropy) や、RNA配列の塩基が一つまたは二つ変化したときの内部エネルギー及びエントロピーの変

化を求めるツール (Rchange) を開発した。

このソフトウェアを用いて人のミトコンドリア上の tRNA におこる塩基置換と、それにより起こる二次構造のエントロピーの変化を網羅的に調べた結果、疾患が報告されていない多型については、エントロピーの変化が有意に小さいことがわかった。(図3)

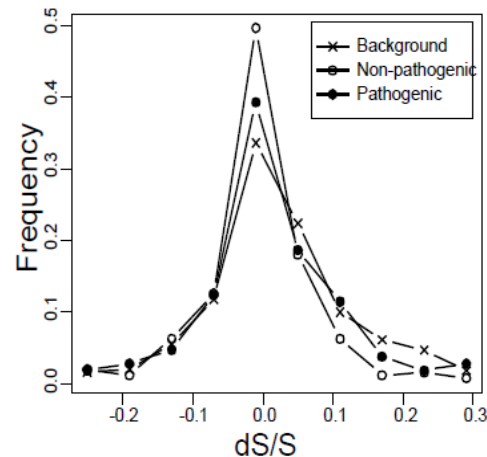


図3: ヒトのミトコンドリア tRNA にみられる置換に付随する二次構造のエントロピー変化の頻度分布、横軸はエントロピーの相対変化を表し、縦軸が頻度を表す。×印はバックグラウンド分布、●は疾患関連の塩基置換、白丸は疾患が報告されていない多型を表す。この図より、疾患に関連しない多型は、エントロピーの変化が小さく、エントロピーが増えるものが特に少ない事がわかる(統計的検定をすると、p値が 10^{-7} 乗程度の有意性がある)

(4) 上記のように、構造 RNA の様々な特徴を抽出する技術、及び、進化系統樹の理論を進展させて、より詳細な進化情報を比較ゲノムの手法で引き出す技術については、大きな進歩が得られた。しかし、当初の目的であった、大量の Alu 配列のクラスタリング手法の開発及び、上記で開発した手法を Alu 配列に適用し、Alu 配列の進化的観点での分類を行う研究は、研究期間内には終了することができなかつた。これらについては、ひきつづき研究を行なっている。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

[雑誌論文] (計5件)

① Hisanori Kiryu and Kiyoshi Asai.

"Rchange: Algorithms for computing the

energy changes of RNA secondary structures in response to base mutations. "

Bioinformatics

doi:10.1093/bioinformatics/bts097、査読有り、(2012)

② Hisanori Kiryu

" Sufficient statistics and expectation maximization algorithms in phylogenetic tree models "

Bioinformatics 27(17), 2346-2353 査読有り (2011)

③ Hisanori Kiryu, Goro Terai, Osamu

Imamura, Hiroyuki Yoneyama, Kenji Suzuki, Kiyoshi Asai .

" A detailed investigation of accessibilities around target sites of siRNAs and miRNAs "

Bioinformatics 27(13), 1788-1797 査読有り (2011)

④ Michiaki Hamada, Hisanori Kiryu, Wataru

Iwasaki, Kiyoshi Asai,

" Generalized Centroid Estimators in Bioinformatics",

PLoS ONE 6(2):e16450,

doi:10.1371/journal.pone.0016450. 査読有り (2011)

⑤ Hamada M, Sato K, Kiryu H, Mituyama T, Asai K.

"CentroidAlign: Fast and Accurate Aligner for Structured RNAs by Maximizing Expected Sum-of-pairs Score ."

Bioinformatics 25(24): 3236-3243 ;doi: 10.1093/bioinformatics/btp580 査読有り (2009)

[学会発表] (計1件)

① Hisanori Kiryu and Kiyoshi Asai,

" On algorithms for computing thermodynamic entropy of RNA secondary structures "

第34回日本分子生物学会年会、2011年12月14日 パシフィコ横浜

[その他]

ホームページ等

①Rchange

<http://www.ncrna.org/software/rchange/>
塩基置換にともなう RNA の二次構造的なエネルギー変化を計算するプログラム (2011)

②Rentropy

<http://www.ncrna.org/software/rentropy/>
RNA の二次構造的なエントロピーを計算するプログラム (2011)

③Fdur

<http://www.ncrna.org/software/fdur/>
系統樹の十分統計量を計算するプログラム (2011)

④Raccess

<http://www.ncrna.org/software/Raccess/>
RNA の二次構造的アクセシビリティを計算するプログラム (2010)

6. 研究組織

(1) 研究代表者

木立 尚孝(KIRYU HISANORI)

東京大学・大学院新領域創成科学研究科・准教授

研究者番号：80415778

(2) 研究分担者

なし

(3) 連携研究者

なし