慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

Title	RNase DiCE : Ribonuclease Discriminative Conserved Element for identifying novel Ribonuclease
Sub Title	RNase DiCE : 新規リボヌクレアーゼ予測のための進化的に保存された識別配列
Author	今井, 淳之介
Publisher	慶應義塾大学湘南藤沢学会
Publication year	2014
Jtitle	生命と情報 No.21 (2014.) ,p.73- 80
JaLC DOI	
Abstract	RNA代謝は全ての転写産物のプロセシング, RNA編集, 翻訳, 分解の全ての過程を示し, 全ての生物が持つ普遍的な機構である。このRNA代謝の中でもRNA分子のプロセシング•分解を担う実態はRibonuclease(RNase)であり, 本酵素とRNA分子との相互作用やその分子進化を解明する事は生物の有する遺伝子制御系の中心的な理解につながると考えられる。 これまでRNaseについては様々な生物種で実験的にその同定が行われてきた。しかし, 同一の機能を持ち, 幅広い生物種に保存されているRNaseであっても生物種に応じてRNA結合ドメインや切断活性領域が多様であるため, 既存の情報科学的な配列解析手法では, 多様なRNaseを統一的に理解したり, 新しいRNaseの予測が困難であった。これらの問題に対処するため, 本研究では既知RNaseに対して, 全生物種におけるゲノム上での存在比を勘案したRNaseホモログの識別に有用なペプチド断片(Discriminated Conserved Element ; DiCE)の抽出手法を考案し, 原核生物48種(Bacteria : 34種, Archaea : 14種), 真核生物5種の既知RNaseから実際に139個のDiCEが抽出に成功した。抽出されたDiCEを用いて既知RNase111本(Bacteria : 77本, Archaea : 18本, Eukaryota : 16本)に対して予測精度検証を行った結果, 97%(108/111)の精度で予測に成功し, その有用性を確認することができた。そこで, 将来の分子生物学的な検証実験も視野に入れて, 超好熱性のアーキアであり, 所有するタンパク質が耐熱性を呈することにより生化学的な解析に適したPyrococcus furiosus(P. furiosus)の全タンパク質2,045本を対象として新規RNase候補タンパク質の予測を行った。その結果, 40本の新規RNase候補タンパク質と14本の既知RNaseが予測された。これらの結果に加えて本卒業論文ではDiCEの既知RNaseにおける保存部位や系統分布を明らかにし, 既知RNaseの配列進化過程についても議論を行いたい。
Notes	慶應義塾大学湘南藤沢キャンパス先端生命科学研究会 2014年度学生論文集 卒業論文ダイジェスト
Genre	Technical Report
URL	https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=KO92001004-00000021-0073

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

RNase DiCE: Ribonuclease Discriminative Conserved Element for identifying novel Ribonuclease

(RNase DiCE: 新規リボヌクレアーゼ予測のための進化的に保存された識別配列)

慶應義塾大学 環境情報学部
今井淳之介

論文要旨

RNA代謝は全ての転写産物のプロセシング，RNA編集，翻訳，分解の全ての過程を示し，全ての生物が持つ普遍的な機構である．このRNA代謝の中でもRNA分子のプロセシング・分解を担う実態はRibonuclease (RNase) であり，本酵素とRNA分子との相互作用やその分子進化を解明する事は生物の有する遺伝子制御系の中心的な理解につながると考えられる．

これまでRNaseについては様々な生物種で実験的にその同定が行われてきた．しかし，同一の機能を持ち，幅広い生物種に保存されているRNaseであっても生物種に応じてRNA結合ドメインや切断活性領域が多様であるため，既存の情報科学的な配列解析手法では，多様なRNaseを統一的に理解したり，新しいRNaseの予測が困難であった．これらの問題に対処するため，本研究では既知RNaseに対して，全生物種におけるゲノム上での存在比を勘案したRNaseホモログの識別に有用なペプチド断片 (Discriminated Conserved Element; DiCE) の抽出手法を考案し，原核生物48種 (Bacteria : 34種，Archaea : 14種)，真核生物5種の既知RNaseから実際に139個のDiCEが抽出に成功した．抽出されたDiCEを用いて既知RNase 111本 (Bacteria : 77本，Archaea : 18本，Eukaryota : 16本) に対して予測精度検証を行った結果，97% (108/111) の精度で予測に成功し，その有用性を確認することができた．そこで，将来の分子生物学的な検証実験も視野に入れて，超好熱性のアーキアであり，所有するタンパク質が耐熱性を呈することにより生化学的な解析に適した*Pyrococcus furiosus* (*P.furiosus*) の全タンパク質2,045本を対象として新規RNase候補タンパク質の予測を行った．その結果，40本の新規RNase候補タンパク質と14本の既知RNaseが予測された．これらの結果に加えて本卒業論文ではDiCEの既知RNaseにおける保存部位や系統分布を明らかにし，既知RNaseの配列進化過程についても議論を行いたい．

Keywords: Bioinformatics, Ribonuclease, Phylogenetics, Molecular evolution

1. 背景

RNA代謝はすべての転写産物のプロセシング，編集，発現抑制，分解のすべての過程を示し，すべての生物が持つ普遍的な機構である．このRNA代謝の中でもRNA分子のプロセシング・分解はRibonuclease (RNase) によって行なわれている．RNA分子の分解を担うRNaseとRNA分子との相互作用や起源や進化を解明する事は複雑な生命システムの包括的に理解につながると考える．

これまで，RNA代謝については主にRNaseによる機能性non-coding RNAのプロセシングを中心に議論がなされてきた**[1-3]**．しかし近年，機能性non-coding RNAのプロセシングだけでなく，RNA分子の分解による生体内で様々な役割が明らかにされてきた．例えば，真核生物では生体内におけるウィルスや酸化ストレスに対するに応答反応としてtRNAのアンチコドンループやrRNAを切断することでmRNAの翻訳を制御する事が報告されている**[4]**．またtRNAの分解産物がsmall RNAとして遺伝子発現制御に関与する事が報告されている**[5]**．さらに環境サンプル中のメタトランスクリプトーム解析により，tRNAの新規切断部位が同定された事により新規のtRNA分解機構が示唆された**[6]**．

上述のようにRNA分解の生体内における役割は様々な広がりを見せている．しかしながら，同一の機能を持ち，幅広い生物種に保存されているRNaseであっても生物種に応じてRNA結合ドメインや切断活性領域が多様であり，既存の配列解析手法ではRNaseの予測が困難であった**[7]**．そのため，既存のRNaseに関しては，その分子メカニズムについては精力的に研究されている一方で，配列解析による多様なRNaseの統一的な理解や配列進化過程に関する報告が少ない．

そこで我々は，これらの問題に対処するため，既知RNaseに対して，全生物種におけるゲノム上での存在比を勘案したRNaseホモログの識別に有用なペプチド断片 (Discriminated Conserved Element; DiCE) の抽出手法を考案し，原核生物48種 (Bacteria : 34種，Archaea : 14種)，真核生物5種の既知RNaseから実際に139個のDiCEが抽出に成功した．

本卒業論文ダイジェストでは卒業論文で言及した，以下の2つの成果について述べる．まず，この抽出されたDiCEを用いて既知RNase 111本 (Bacteria : 77本，Archaea : 18本，Eukaryota : 16本) に対して予測精度検証を行った結果，97% (108/111) の精度で予測に成功し，その有用性を確認することができた．さらに，既存のモチーフ抽出ソフトウェアであるHMMER，及びMEME/MASTから抽出されたモチーフとの検出精度比較を行った結果，本手法により抽出されたRNase DiCEは正確性の高い新規RNase候補の予測が可能である事が示唆された．

次に，生物種間での既知RNaseとRNase DiCEとの系統分布に着目し，既知RNaseとRNase DiCEを用いて系統プロファイル解析を行った．その結果，種特異的に保存されているRNaseであっても，RNase DiCEレベルでの保存性をみると幅広い生物種に渡ってRNaseホモログが存在する事が示唆された．

2. 対象と手法

2.1. データセット

本研究ではタイプ種と認定されている真核生物 5種，原核生物 34種，古細菌 14種を対象とし，タンパク質データベースであるUniprotKB/Swiss-Protより計601本の既知RNaseのアミノ酸配列情報を取得した(**表1**)．また，ゲノム中のRNaseの部分領域の抽出にはNational Center for Biotechnology Information Reference sequence database (NCBI Refseq) に登録されている原核生物 1,951種 (3,819本)，古細菌 125種 (201本)，計4,020本のゲノム配列を取得した．

2.2. 配列類似度に基づいた既知RNaseの分類

RNaseは広範な生物種にわたって多様な機能カテゴリが存在する．まず，これらのRNaseを機能カテゴリ毎に分類を行うため配列類似性に基づいてクラスタリングを行った．クラスタリングを行う際に，既知RNaseのアミノ酸配列に対して総当たりでBLAST (Version 2.2.25, -p blastp, -e 1e-10, -F F, -M BLOUSUM 62)を用いて相同性検索を行った**[7]**．次に相同性検索により得られた*Bits score*に基づいて配列類似スコア*Sim(x,y)*を算出した (**式1**).

$$Sim(x,y)=\frac{max(S'_{bits}(x,y),S'_{bits}(y,x))}{max(S'_{bits}(x,x),S'_{bits}(y,y))} \qquad \textbf{(式1)}$$

ここで*Sim(x,y)*は配列x,y間の*Bits score*の最大値を比較した値であり，これにより配列x,y間の相同性が$0 \leq Sim(x,y) \leq 1$の範囲の距離として算出される．601本のRNaseのアミノ酸配列の*Sim(x,y)*算出後，Spectral Clustering of Protein Sequence (SCPS) (Version 0.9.5, Clustering algorithm : Spectral clustering, Number of clusters : Automatic) を用いてスペクトラルクラスタリングを行った**[8]**．スペクトラルクラスタリングはネットワークの構造に基づいてネットワークを分割する手法であり，多様なファミリータンパク質をもつタンパク質の機能分類を行うための有効な手法であると考える．

2.3. RNase DiCEの抽出

真正細菌・古細菌のゲノム配列中から進化的に高度に保存されたRNaseの領域を抽出するため，BLAST (Version 2.2.25, -p tblastn, -e 1e-2, -F F, -M BLOUSUM 62) を用いて601本の既知RNaseを問い合わせ配列として，4,020本のゲノム配列に対して相同性検索を行った．

次に，RNaseのアミノ酸配列上より共通した保存領域を取得するためスペクトラルクラスタリングにより分類後の既知RNaseのアミノ酸配列それぞれに対してマルチプルアライメントを行った．本解析では類似性の高い配列に対してアライメント精度の信頼性が高いT-COFFEE (Version 9.01.r11213, -gapopen = 0, gapext = 0) を用いてアライメントを行った**[9]**.

RNaseのアミノ酸配列上の保存領域に対してゲノム上の保存度合いを見るため，T-COFFEEによってアライメントされた配列上の各残基に対して，BLAST解析により得られたRNaseの領域をBLOUSUM 62のスコアに基づくスコアを積算して行った．また各残基における保存性を定量化するため，1残基ごとに以下のように *(O-E)/E* 値 を定義した (**式2**) .

$$\frac{(O_i - E_i)}{E_i} = O_i\left(\frac{\sum_{j=1}^{l} O_j}{\sum_{k=1}^{q} L_k} N_i\right)^{-1} - 1 i = 1, ..., n \qquad \textbf{(式2)}$$

(O-E)/E値によるアライメント後の各残基における保存性の定量化後，RNaseの進化的に保存された領域を定義するため，マイクロアレイ発現データなどの正規化に用いられるノンパラメトリック局所回帰手法の一つであるLocally Scatterprot Smoothing (LOWESS) 法を用いて(O-E)/E値の平滑化を行った．LOWESS法はデータに対して局所的に回帰曲線を当てはめる手法の一つであり，RNase DiCE抽出の際に生じる外れ値の影響の除外を意図して使用した．最終的にLOWESSによる平滑化後の領域でかつ*(O-E)/E*値が0以上の領域を本研究ではDiCEと定義した．

2.4. RNase DiCEを抽出する際の最尤パラメーター探索と既存のモチーフ抽出手法との精度比較

本手法では，データセットの構成およびLOWESS法による平滑幅といった2つのパラメーターが存在し，抽出されるDiCEの候補領域はこの2つのパラメーターに依存して変化する．そこで，DiCEのを抽出する際の最尤パラメーターを探索するため，全ての既知RNaseを統合したデータセット，また，真核生物，原核生物，古細菌の3つにデータセット分割したの2パターンのデータセットを用いて，それぞれLOWESS法による3条件の平滑幅を用いてDiCEを抽出を行った．また本研究では，既知モチーフ抽出手法であるMEME/MAST，及びHMMERでの予測精度比較を行った．予測精度検証の際に適合率 (Precision)，再現率 (Recall)，F値 (F-measure) の3つの指標を使用した．

2.5. RNase DiCEと既知RNaseの保存性解析

本研究で抽出されたRNase DiCEと既知RNaseとの系統分布に違いが見られるかを検証するため，既知RNaseとRNase DiCEについて保存性解析を行った．まず，RNase DiCEの抽出の際に用いた601本の既知RNase，また抽出されたRNase DiCEをそれぞれクエリーとして，本研究で対象とした真正細菌 34種，古細菌 14種のゲノム配列に対してBLAST（Version 2.2.25，-p tblastn，-e 1e-2，-F F -M BLOSUM62）を用いて相同性検索を行い，最も高い*Bit score*が算出された領域を抽出した．次に既知RNaseおよびRNase DiCE同士でヒットした*Bits score*の最大値を同様に算出し，その後，配列類似スコア*Sim(x,y)*を算出し，配列の保存度合いを0-1にスケーリングを行った**(式1)**.

3.結果と議論

3.1. RNase DiCEの抽出における最尤パラメーター探索

本手法ではクラスタリング条件，LOWESS法による平滑幅の2つのパラメーター依存して抽出されるRNase DiCEの領域が変化する．RNase DiCEを用いた新規RNase候補予測の予測性能が最大となるRNase DiCEの抽出パラメーターの探索を行うため，クラスタリング条件を2条件，平滑幅の条件を3条件，計6パターンの条件下でRNase DiCEを抽出し，HMMER (ver 3.0b3)を用いて，今回テストデータとした14種の生物種 (真核生物 1種，原核生物 10種，古細菌 3種)の全タンパク質のアミノ酸配列情報 13,223本 (正解セットとして既知RNase 111本を含む) に対して予測精度検証を行った．再現率，適合率による各条件での予測精度を示す **(表1)**.

まず，真核生物，原核生物，古細菌の全ての既知RNaseを統合してクラスタリングを行ったデータセットから抽出されたRNase DiCEの予測精度について述べる．全体として，再現率は0.75 ~ 0.8の間の値になり，その中でも最も高い再現率を示した条件は平滑幅0.01の条件下で抽出されたRNase DiCEの0.792であった．適合率については，0.1また0.01の条件下で抽出されたRNase DiCEが共に0.081であった．F値についても，平滑幅0.01，0.1ともに，適合率，再現率に大きな差がみられなかったため，0.148を示した．

次に真核生物，原核生物，古細菌の門毎にデータセットを分割してクラスタリングを行ったデータセットから抽出されたRNase DiCEの予測精度について述べる．再現率，適合率，F値は共に，真核生物，原核生物，古細菌の全ての既知RNaseを統合してクラスタリングを行ったデータセットから抽出されたRNase DiCEと比較すると，高い予測精度が算出された．予測精度を算出した平滑幅の条件下の中で，再現率，適合率，F値の最大値を示した条件は平滑幅0.1の条件で抽出されたRNase DiCEで，それぞれ再現率0.972，適合率0.096，F値0.175であった．また，0.1での条件下以外にも着目すると，平滑幅0.01，0.001共に，再現率0.981，適合率0.88であった．

真核生物，原核生物，古細菌のデータセットを統合した場合，門毎に分割した場合の2条件で，平滑幅を3条件に変化させて，抽出したRNase DiCEを用いて既知RNaseの予測精度が最大となるパラメーター探索を行った結果，門毎に分割した場合のデータセットでRNase DiCEの抽出を行った場合の方が全体として高い再現率，適合率であった．真核生物，原核生物，古細菌のデータセットを統合してクラスタリングを行ったデータセットに着目すると，複数のRNaseがファミリータンパク質として同じクラスターに属している事がわかる (卒業論文を参照) ．これらのクラスターのアライメントに対して，ゲノムから抽出したRNaseの保存領域をマッピングし，(O-E)/E値によるスクリーニングを行う際，種特異的なRNaseが保存しているはずの領域の保存度合いが低く算出されてしまい，RNase DiCEとして抽出されてこない可能性が挙げられる．一方で，門毎にデータセットを分割してクラスタリングを行ったものに関しては，上述したような種特異的に保存されているRNaseの領域をスクリーニングしてしまう可能性を最小限にとどめる事ができたため，高い再現率，適合率を可能にしたと考える．以上の結果から，真核生物，原核生物，古細菌を門毎に分割したデータセットを用いる事で広く保存されているRNaseの保存領域を反映する一方で，種特異的に保存されているRNaseの領域の抽出が可能である事が示唆された．

3.2. 既存のモチーフ抽出手法との精度比較結果

これまで転写因子のDNA結合領域の予測手法をはじめとする，数々のモチーフ抽出手法が考案されてきた．既存のモチーフ抽出手法の代表的なアルゴリズムとして，隠れマルコフモデル，EMアルゴリズム，Gibbs Samplingアルゴリズムの3つが挙げられる[ref-ref]．そこで，本手法により抽出された，RNase DiCEの予測性能の検証するため，アミノ酸配列からモチーフ抽出が可能なソフトウェアであり，隠れマルコフモデルとEMアルゴリズムをそれぞれ実装したモチーフ抽出ソフトウェアであるHMMER，MEME/MASTに対して既知RNaseの検出精度比較を行った(**表1**)．その結果，真核生物，原核生物，古細菌の門毎にデータセットを分割してクラスタリングを行ったデータセット，統合してクラスタリングを行ったデータセット，共にHMMERによりモチーフ抽出を行ったものは0.972 (108/111) と最も高い再現率示した．一方で適合率に着目すると，本手法を用いて抽出されたRNase DiCEは0.096 (108/1120)と最も高い適合率を示した．

表1: データセット及び，モチーフ抽出手法毎の再現率，適合率の比較結果

Clustering	Methods	Smoother span (LOWESS)	Recall (TP/TP+FN)	Precision (TP/TP+FP)	F-measure
All clustering	This method + hmmbuild	0.1	0.756 (84/111)	0.075 (84/1107)	0.137
		0.01	0.792 (88/111)	0.081 (88/1074)	0.148
		0.001	0.783 (87/111)	0.081 (87/1062)	0.148
	hmmbuild	-	0.981 (109/111)	0.090 (109/1201)	0.166
	MEME / MAST	-	0.864 (96/111)	0.078 (96/1223)	0.143
Archaea/Bacteria/Eukaryota	This method + hmmbuild	0.1	0.972 (108/111)	0.096 (108/1120)	0.175
		0.01	0.891 (99/111)	0.088 (99/1120)	0.160
		0.001	0.891 (99/111)	0.088 (99/1116)	0.161
	hmmbuild	-	0.981 (109/111)	0.086 (109/1253)	0.158
	MEME / MAST	-	0.963 (107/111)	0.062 (107/1712)	0.117

データセット及びモチーフ抽出手法毎の再現率，適合率の結果を示す．左からクラスタリングを行ったデータセット，モチーフ抽出手法，RNase DiCE抽出の際の予測パラメーター，再現率，適合率，F値を示す．

3.3. 既知RNaseおよびRNase DiCEの保存性

次に生物種間の既知RNaseとRNase DiCEの保存性の違いを明らかにするため，今回対象とした真正細菌34種，古細菌14種における既知RNaseの全長配列とRNase DiCEの保存性解析を行った (卒業論文参照)．エキソリボヌクレアーゼにおけるRNase全長とRNase DiCEの保存性についてまず，真正細菌のPNPase (Cluster 1)とRNase PH (Cluster 5)に着目したい (**図1**)．まず，これら2つのRNaseは真正細菌27種において幅広く保存されいるが古細菌においては保存性がみられなかった．しかし，RNase DiCEレベルでの保存性に着目すると，PNPase (Cluster 1) 由来のDiCE2において，Euryarcheota門である*Methanopyrus kandleri*, *Pyrococcus furiosus*, *Methanocaldococcus jannaschii*, *Halobacterium salinarum*の4種，Crenarchaeota門である*Pyrobaculum aerophilum*, *Aeropyrum pernix*の２種，Nanoarchaeota門である*Nanoarchaeum equitans*に保存性が見られた．また，RNase PH (Cluster 5) に関しても同様にRNase DiCEレベルの保存性に着目したところ，DiCE 3，DiCE 4において，Euryarcheota門である*Methanopyrus kandleri*, *Methanocaldococcus jannaschii*, Nanoarchaeota門である*Nanoarchaeum equitans*, Korarchaeota門である*Korarchaeum cryptofilum*にいて保存されていた．

同様に古細菌由来のRNase及びRNase DiCEについても同様の傾向がみられ，例えば，Exosome complex (Cluster 1) では，配列全長レベルでは古細菌特異的に保存されていたが，DiCE2，DiCE3では真正細菌全般で強く保存されている．

近年，RNase EのN末端部分またはRNaseドメインの一部分に，相同領域がある新規RNase候補タンパク質が*Pyrococcus furiosus*（*P.furiosus*），*Pyrococcus horikoshii*（*P.horikoshii*）からそれぞれ単離された**[10]**．本研究の結果は，一見，種特異的に保存されているRNase であってもRNase DiCEレベルでの保存性を見ると，上記の先行研究のような部分的な領域にのみ相同領域が存在するRNase候補が幅広く存在する可能性を示唆している．

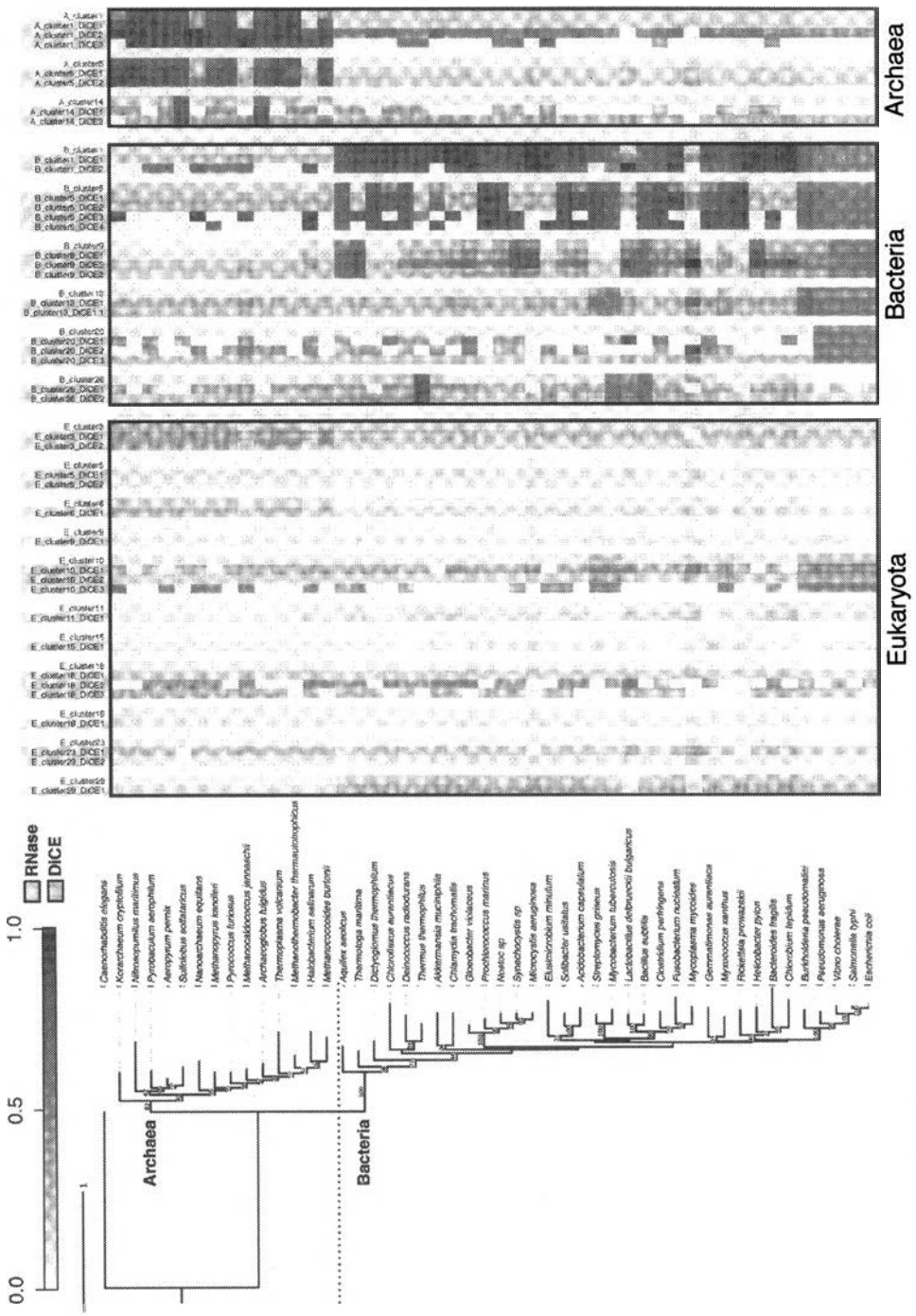

図1: エキソリボヌクレアーゼにおける既知RNaseとRNase DiCEの保存性解析

本研究で対象とした真正細菌34種，古細菌14種のゲノム中におけるRNase全長配列とRNase DiCEの保存性の度合いをヒートマップとして示す．ヒートマップは左から真核生物，真正細菌，古細菌のヒートマップを示している．行の黄色いカラムはRNaseの全長配列，青いカラムはRNase DiCEを示す．ヒートマップの色が赤いほど保存性の度合いが高いことを示す．

謝辞

本研究を行うにあたりアドバイザーである金井昭夫教授には研究計画の段階から解析段階におけるまで多くの助言を頂いた．また，RNAグループの皆様にも多くの助言を頂いたおかげで本研究を進める事ができた．最後に同大学先端生命科学研究所所長の冨田勝教授はこのような環境と成長の機会を与えて下さった．以上の方々に感謝を申し上げます．

参考文献

1. Deutscher MP: **Maturation and degradation of ribosomal RNA in bacteria.** Prog Mol Biol Transl Sci 2009, 85:369-391.
2. Winter J, Jung S, Keller S, Gregory RI, Diederichs S: **Many roads to maturity: microRNA biogenesis pathways and their regulation.** Nat Cell Biol 2009, **11**(3):228-234.
3. Garber RL, Altman S: In vitro processing of B. mori transfer RNA precursor molecules. Cell 1979, 17(2):389-397.
4. Thompson DM, Lu C, Green PJ, Parker R: **tRNA cleavage is a conserved response to oxidative stress in eukaryotes.** RNA 2008, **14**(10):2095-2103.
5. Lee YS, Shibata Y, Malhotra A, Dutta A: **A novel class of small RNAs: tRNA-derived RNA fragments (tRFs).** Genes Dev 2009, **23**(22):2639-2649.
6. Murakami S, Fujishima K, Tomita M, Kanai A: **Metatranscriptomic analysis of microbes in an Oceanfront deep-subsurface hot spring reveals novel small RNAs and type-specific tRNA degradation.** Appl Environ Microbiol 2012, **78**(4):1015-1022.
7. Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ: **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** Nucleic Acids Res 1997, **25(17)**:3389-3402.
8. Nepusz T, Sasidharan R, Paccanaro A: **SCPS: a fast implementation of a spectral method for detecting protein families on a genome-wide scale.** BMC Bioinformatics 2010, 11:120.
9. Notredame C, Higgins DG, Heringa J: **T-Coffee: A novel method for fast and accurate multiple sequence alignment.** J Mol Biol 2000, **302(1)**:205-217.
10. Kanai A, Oida H, Matsuura N, Doi H: **Expression cloning and characterization of a novel gene that encodes the RNA-binding protein FAU-1 from Pyrococcus furiosus.** Biochem J 2003, **372(Pt 1)**:253-261.