

科学研究費助成事業（科学研究費補助金）研究成果報告書

平成25年 5月31日現在

機関番号：82401

研究種目：若手研究(B)

研究期間：2010～2012

課題番号：22710199

研究課題名（和文） 抗生物質生成菌と光合成細菌の転写関連データベースの構築とその応用

研究課題名（英文） Database construction for transcriptional regulation in *Streptomyces* and *Synechocystis* and these applications

研究代表者

蔭田 由布子 (MAKITA YUKO)

独立行政法人理化学研究所・生命情報基盤研究部門・研究員

研究者番号：80443026

研究成果の概要（和文）：

転写制御因子の進化速度は他のタンパク質に比べ速く、近縁種においてもその制御関係が保存されていない。そのため、一部のモデル微生物にのみ注目するのでは、バクテリアの多様性を理解するのに不十分である。そのため本研究では、特出すべき特性を持つ、抗生物質生産菌の放線菌と光合成細菌の藍藻に注目し、約 500 文献から転写制御情報を抽出し、データベースを構築した。またそのデータを用いて、比較ゲノム解析やネットワークモチーフ解析を行ない、生物種特有の機能（放線菌の抗生物質合成、藍藻の光合成）に関する転写ネットワークの特異性を示した。

研究成果の概要（英文）：

Recent works in bacterial transcriptional network analysis reveal that transcriptional regulations are flexible and evolve rapidly. It is therefore important to construct reference databases on transcriptional regulation of multiple model organisms. I constructed referenced databases in *Streptomyces* and *Synechocystis* with manually collected data from nearly 500 literatures. By using the collected data, I performed comparative genome analysis and network motif analysis and suggested specificity of transcriptional networks in species-specific functions, such as antibiotic synthesis of *Streptomyces* and photosynthesis of *Synechocystis*.

交付決定額

(金額単位：円)

	直接経費	間接経費	合計
2010年度	1,600,000	480,000	2,080,000
2011年度	222,659	66,797	289,456
2012年度	1,277,341	383,202	1,660,543
年度			
年度			
総計	3,100,000	929,999	4,029,999

研究分野：複合新領域、バイオインフォマティクス、データベース

科研費の分科・細目：ゲノム科学・システムゲノム科学

キーワード：リファレンスデータベース、転写制御、文献抽出、比較ゲノム、プロモータ解析、ネットワークモチーフ解析

1. 研究開始当初の背景

バクテリアの転写ネットワーク解析により、転写制御因子は、通常のタンパク質配列に比べ進化速度が非常に速く、近縁種においてもあまり保存されていないことが明らかにされた (Price, M., *et al.* 2007. *PLoS Computational Biology*, 3(9):1739-50.)。これは既にある、グラム陰性菌のモデル微生物、大腸菌の転写関連データベース (RegulonDB) やグラム陽性菌のモデル微生物、枯草菌の転写関連データベース (DBTBS) のデータをもとに、全バクテリアに普遍的な転写ネットワークの性質を理解しようという研究には無理があり、もっと多くのモデル微生物種において、同様のデータが集められ、比較解析される必要性が示された。

2. 研究の目的

本研究では抗生物質生産菌の放線菌と光合成細菌の藍藻に着目し、それらの転写制御に関するリファレンスデータベースの構築を行うことを目的とした。その際、文献からマニュアルキュレーション作業でデータ収集を行ない、データの質を高め、より多くの研究者に使われるデータベースを構築する。具体的には、先行研究を簡単にまとめて確認し、研究の推進に貢献するだけでなく、バイオインフォマティクス研究における学習 (正解) データセットともなり、様々な予測方法の確立に貢献すること、さらに比較ゲノム解析の元としても使われるものを目指す。次に、収集したデータを用いた解析においては、放線菌のゲノム GC 含量の高さ (72.1%) に注目し、ゲノム GC 含量の転写因子の結合配列へ与える影響を調べる。それにより、転写因子がどのようにゲノム中からモチーフをみつけているのかを考察する。さらに、ネットワークモチーフ解析を行い、既に知られている大腸菌の傾向 (Shen-Orr, S., *et al.* 2002. *Nat Genet* 31, 64-68.) と比較・検討する。

3. 研究の方法

転写制御データの収集

本研究では、データの質を重要視したため、文献からマニュアルキュレーションにより転写関連情報 (転写因子、被制御遺伝子、転写因子の制御機能、転写因子の結合配列、転写開始点、オペロン情報、ターミネータの位置とヘアピン構造、実験方法、実験条件など)

を収集した。

藍藻では *Synechocystis* sp. PCC 6803 のデータを、放線菌では *Streptomyces* 属の 15 種を対象にデータ収集を行った。

データベース構築

収集したデータは xml フォーマットで管理し、そこから自動的に HTML へ変換し、データベースとして公開・更新できるようにして構築した。

検索機能は、遺伝子名からのデータ検索機能、配列をクエリーに、その配列中に転写因子結合配列モチーフを検索する機能、転写因子欠乏配列の重み行列やマルチプルアライメントをクエリー二、重複してでてきたモチーフが既存かどうかを確認する検索機能を付け、転写因子の結合配列に着目した検索を充実させた。

ユーザに配布するデータフォーマットは、xml だけではなくタブ区切りのテキストファイルや、セマンティックウェブに準拠した RDF ファイルなどを用意した。

転写因子の結合配列における比較解析

今回集めた放線菌、藍藻の転写関連データに加え、次にあげる既存の 4 生物種の転写関連データベースのデータを利用した (大腸菌 (RegulonDB)、枯草菌 (DBTBS)、緑膿菌 (PRODORIC)、結核菌 (MtbRegList))。これらの多様なゲノム GC 含量やプロモータ領域のみ、rRNA、tRNA、シグマ因子の結合モチーフの GC 含量と、転写因子の結合配列の GC 含量を比較した。

ネットワークモチーフ解析

大腸菌や酵母の転写ネットワークで言われている傾向が、放線菌や藍藻においても同じ傾向なのかを確認するため、スケールフリー性の確認や、ネットワークモデルの解析とネットワークモチーフ (feedforward loop (FFL), single input module (SIM), dense overlapping regulons (DOR)) の解析を行った。

4. 研究成果

本研究の成果としては、1) マニュアルキュレーションにより約 500 文献から放線菌・藍藻の転写関連データの収集を行い、2) データベースを構築した。さらに、収集したデータを使って、3) 比較ゲノム解析と 4) ネットワークモチーフ解析行なった。

まず 1) 放線菌のデータ収集においては、研究当初は *Streptomyces coelicolor* をモデルとしてデータ収集するつもりであったが、大

腸菌の K-12 株、枯草菌の 168 株のように、一つのモデル微生物を集中して研究が進むのではなく、抗生物質ごとに放線菌の種が研究されていたため、対象生物種を 15 種と広げた。その結果、302 文献から、転写因子数 152、被制御遺伝子数 459、転写開始点数 325 を収集した。また、転写開始点の収集過程で、翻訳開始点が転写開始点よりも上流に定義されているものがあつた。それらは翻訳開始点の予測の間違いであることが想定されるので、ウェブサイトから公開し、修正を促している。

次に藍藻では、*Synechocystis* sp. PCC 6803 をモデル生物とし、177 文献から、52 転写因子、1093 被制御遺伝子、86 オペロンを抽出した。また藍藻においては全データの約 8 割のデータに実験条件(24時間光やUV-B照射、塩ストレスなど)も付けることができた。放線菌、藍藻ともに、実験方法(マイクロアレイやフットプリント解析、ゲルシフトアッセイなど)や転写因子の結合配列、ターミネータなども整理、収集し、転写因子の結合配列モチーフも、より多くの収集データで再定義している。

2) 以上の収集データを元に、DBSCR(<http://dbscr.hgc.jp>)データベースを構築した。本データベースでは、全データを転写因子ごとと、被制御遺伝子ごととにまとめ、表示している。検索機能としては、収集データから再定義した転写因子結合モチーフ配列(TFBS)に関する検索を充実させた(詳細、研究の方法)。またバイオインフォマティクス研究にも活用してもらうため、全データの配布も行なう。データは、xml フォーマット、tab 区切りのテキストファイル、セマンティックウェブに準拠した RDF ファイルを用意し、より多くの研究者への使い勝手を高めた。

3) さらに、集めたデータを元に、自分自身でも比較ゲノム解析を行なった。データセットは、収集した放線菌と藍藻に加え、大腸菌(RegulonDB)、枯草菌(DBTBS)、緑膿菌(PRODORIC)、結核菌(MtbRegList)を利用した。放線菌ゲノムは、GC含有量が72.1%と非常に高いことが特徴ということもあり、まず、転写因子がゲノム配列中から結合配列を見つける際に、ゲノム全体のGC含量がどのように影響するのかを明らかにすることを目的とした。その結果、リボソームRNAなどでは、そもそも種を超えて配列保存性が非常に高いこともあり、ゲノムGC含量に関係なく、一定(50-60%)のGC含量を示した。一方で転写因子の結合配列(TFBS)は、ゲノムGC含量から一定(15-20%前後)した割合でAT-richとなる傾向を示し、これはGCに比べ水素結合の弱いATを利用することにより、DNA配列に結合・解離の自由度を増している

と考えられ、さらに、その際のATを使う割合は、タンパク質配列に影響されるというよりは、ゲノムGC含量に影響されるということが示唆された。

最後に、4) ネットワーク解析では、少数のハブ転写因子が多く遺伝子の転写を制御しているスケールフリー性は、大腸菌同様の種でも確認できた。しかしその他のネットワークモチーフ解析の結果では、生物種特有の機能に関する転写ネットワーク(枯草菌の孢子形成、放線菌の抗生物質合成、藍藻の光合成)においては、大腸菌とは異なる傾向を示した。これは一部のモデル微生物に注目した解析だけでは不十分で、本研究の重要性が指示された。

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 0 件)

[学会発表] (計 3 件)

① 蒔田由布子、石井崇洋、豊田哲郎、中井謙太

放線菌転写関連データベース(DBSCR)の構築とゲノムGC含量と転写因子認識配列の相関
第35回日本分子生物学会、2012年12月10日~2012年12月14日、福岡国際会議場

② Yuko MAKITA, Takahiro ISHII, Tetsuro TOYODA, Kenta NAKAI

Construction of transcriptional regulatory database in *Streptomyces* and comparative analysis of transcription factor binding sequences
2012年生命医薬情報学連合大会 2012年10月15日~2012年10月17日、東京都江戸川区 タワーホール船堀

③ 蒔田由布子

放線菌転写関連DB: DBSCR
グラム陽性菌ゲノム機能会議、2010年9月3日、長野

[図書] (計 1 件)

① Yuko MAKITA and Kenta NAKAI
Chapter 10: *Bacillus subtilis*

Transcriptional Network
Caister Academic Press
Bacterial Gene Regulation and
Transcriptional Networks
M. Madan Babu Edit

[その他]
ホームページ等

<http://dbscr.hgc.jp>

6. 研究組織

(1) 研究代表者

蒔田 由布子 (MAKITA YUKO)

独立行政法人理化学研究所・生命情報基盤研
究部門・研究員

研究者番号：80443026

(2) 研究分担者

なし

(3) 連携研究者

なし