

## 科学研究費助成事業 研究成果報告書

平成 28 年 6 月 20 日現在

機関番号：32665

研究種目：基盤研究(C) (一般)

研究期間：2013～2015

課題番号：25463230

研究課題名(和文) 重度歯周病原細菌のnon-coding RNA検索とその動態

研究課題名(英文) A bacterial non-coding RNAs search and the dynamics in serious periodontal disease

## 研究代表者

平塚 浩一 (HIRATSUKA, Koichi)

日本大学・松戸歯学部・教授

研究者番号：80246917

交付決定額(研究期間全体)：(直接経費) 3,800,000円

研究成果の概要(和文)：本研究は、重篤な歯周病と関連する*P. gingivalis*, *T. forsythia*, *T. denticola*に加え、侵襲性歯周炎の原因である*A. actinomycetemcomitans*のnon-coding RNAsを次世代シーケンサーを用いて網羅的に解析・同定を行った。その結果、*in silico*で予測されていた各種細菌のnon-coding RNAの中で、実際に発現が認められたものは約50～60%であった。また、予測されていない新規のRNAを発見した。本研究は、細菌の新たな病原性遺伝子転写調節メカニズムの解明と、将来的に歯周病診断に有用な診断マーカーとなる。

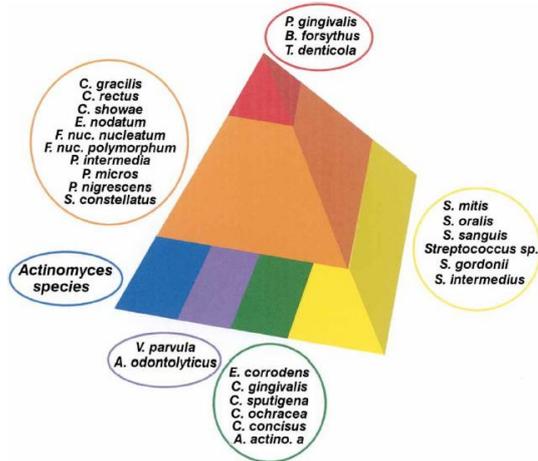
研究成果の概要(英文)：In this study, non-coding RNAs in *P. gingivalis*, *T. forsythia*, *T. denticola* related to severe periodontitis and in *A. actinomycetemcomitans* related to aggressive periodontitis were analyzed and identified comprehensively using a next generation sequencer. As a result, approximately 50-60% of ncRNA in various bacterial ncRNAs predicted *in silico* had their expressions actually. Also, new ncRNAs were discovered. This study is useful for the elucidation of new transcriptional regulations in pathogenic bacterial genes and for the diagnosis marker in a periodontal disease in the future.

研究分野：生化学・分子生物学

キーワード：遺伝子 細菌 RNA ncRNA 遺伝子発現 歯周病

1 . 研究開始当初の背景

( 1 ) 現在 , 口腔内に 800 種を超える細菌種が生息すると言われ , その大部分は嫌気性菌であると考えられている。歯肉縁下プラーク細菌種は他の細菌種との共生集 ( microbial complex ) を形成し , 集合体としてのプロークを構成している。この共生集団は 6 つに色分けされ , かつその病原性により 3 階層に区分されている。その最上層は Red Complex と称され , *Porphyromonas gingivalis* (Pg), *Tannerella forsythia* (Tf), *Treponema denticola* (Td) が属する。これらは概ね重篤な歯周病と最もよく関連し , 口腔内より高頻度で検出され , しばしば歯周病細菌検査のターゲットになっている。



Red Complex に属する 3 菌種は互いの病原性を高め合う性格をもち , 互いの存在が個々の病原遺伝子の発現を促進し , 相乗効果を発揮していると考えられているが , そのメカニズムはいまだ良く分かっていない。

また , 歯肉の発赤 , 腫脹等の症状は軽度であるものの , 急速な歯周組織の破壊を認める侵襲性歯周炎では , Red Complex 3 菌種に加え *Aggregatibacter actinomycetemcomitans* (Aa) の対総菌数比率が健常者よりも高く , 長期間の静止期から短期間で急激な組織破壊を惹起する活動期への周期的な反復により重篤化することが知られている。

( 2 ) microRNA や small 干渉 RNA ( siRNA ) として知られている small non-coding RNA は , 真核生物において , 成熟化・細胞死・染色体サイレンシングの新しい制御因子として作用し , 様々な疾患にその発現が関与することが知られている。近年 , 原核生物においても , 多くの細菌種で相当数の non-coding RNA が発見され , その多くが病原性や環境ストレスに関与する mRNA の翻訳や安定性を調整する鍵となるエフェクターとして作用することが次第に明らかになった [Cell Host Microbe. 2010,2:116-127, Handb Exp Pharmacol. 2006, 173:9-43, Curr Opin Microbiol. 2006, 9:229-236.] が , 新規 non -coding RNA を同定する手法・戦略

は , いまだ手探り状態である。

生命科学の進歩は , 技術の進歩に依存するところが大きい。米国を中心に 50 機関 ( チーム ) 以上のベンチャー企業や研究組織がしのぎを削り次世代シーケンシング ( 塩基配列解読 ) 技術の開発を行った結果 , 新しい原理に基づく次世代シーケンサーが次々と開発され , その性能の伸びは , まさに革命的なインパクトを生命科学に与えると言っても過言ではない。ヒトゲノム計画は高速シーケンサーを用いて , 13 年間 , 約 3 千億の費用を費やし 2003 年に終了したが , 近年のシーケンサーを用いると , わずか 1 ヶ月 , 約百万円以内の費用であり , 1 塩基あたりの費用は数千分の 1 と大幅にコストダウンしている。このように , 次世代シーケンサー 1 台で , 従来型のシーケンサーの 200-1000 台分の能力があるといえる。しかしながら , 機器が高額であるため , 次世代シーケンサーを設置している所は日本では 100 台程度 ( 2011 年 8 月段階 ) である。

2 . 研究の目的

( 1 ) 細菌 non-coding RNA 研究は , その病原性を探る上で重要な研究であるにも関わらず , 効率の良く見出し , 発現解析する実験手法が確立されていないため , 全ゲノムシーケンスが決定されている 417 細菌 ( 申請時現在 ) であっても , non-coding RNA 発現領域の解析を行っているのは極めて稀であった。

Los Alamos National Laboratory は , そのような現状から , 全ゲノム配列の解析が終了している口腔内細菌を対象として , 遺伝子間配列 ( IGS ) 内での潜在的な non-coding RNAs を *in silico* で予測公表している ( [http://www.oralgen.lanl.gov/\\_index.html](http://www.oralgen.lanl.gov/_index.html) ) 。申請者は LANL のデータベースから Red Complex 細菌の登録されてる non-coding RNA の数と平均サイズを計算したところ Table に示すように比較的大きなサイズであった。

Table 1 Predicted Non-coding RNA in Red Complex (Los Alamos)

Bacteria	Number	Mini (bp)	Max (bp)	Average (bp)
<i>Porphyromonas gingivalis</i>	649	100	3,882	435
<i>Tannerella forsythia</i>	795	100	4,899	267
<i>Treponema denticola</i>	195	101	2,445	238

本研究を進めて行く上で , *in silico* での予測と本研究で行われる *in vitro* で発現しうるデータとをつき合わせることで , より精度の高い解析と詳細なアノテーションを知ることができるであろう。また , 加えて , *in silico* で予測できなかった新規の non-coding RNAs の発見にもつなげる可能性が高いと考えられる。また , 理想的な歯周病治療は , 歯周病原細菌やその他関与するすべてのリスクファクターを除去することで , 疾患の進行を停止させ , 口腔機能および審美性の回復を歯周組織を再生することで獲得できると考えられる。従って , 将来的には , 診断マーカーとなり得る non-coding RNA を含む細

菌検査を歯周病治療に組み込み、エビデンスに基づいた確かな抗菌療法を実施する基盤になると考えている。

(2) 本研究では、重篤な歯周病と最も関連性のある red complex と称される Pg, Tf, Td に加え、侵襲性歯周炎の原因細菌である Aa の non-coding RNAs を最新の次世代シーケンサーを用いて網羅的に同定・解析し、*in silico* で予測されている各種細菌の non-coding RNA と照合することで、実際に発現する non-coding RNA を同定し、同時に、*in silico* で予測不可能であった non-coding RNA を新たに発見する。本研究は、細菌の新たな病原性遺伝子転写調節メカニズムの解明と、将来的に歯周病診断に有用な診断マーカーとなるプローブの検索を目的とする。

### 3. 研究の方法

#### (1) ストレス付加と mRNA-rich 試料作成

Red complex 菌種及び Aa に対して、口腔内で起こりえる様々な温度・酸化・酸素ストレス等の負荷をかけ、経時的にサンプルを抽出し、Trizol/column 法 (Invitrogen) で total RNA を抽出した。抽出した RNA は品質管理のため、分光光度計 Nanodrop を用いた 260/280 比の測定と Agilent 2100 Bio-analyzer を用いた RNA 分解度 (RIN 比) を測定した。MEGAClear (Ambion) により、5 S rRNA および 100bp 以下のサイズの RNA (主に 70bp 前後 tRNA) の除去を行った。その後、BICROB Express Bacterial mRNA purification Kit (Ambion) を用いて、主要な 16SrRNA および 23SrRNA の除去を行い、mRNA-rich 試料の精製を行った。

#### (2) Illumina HiSeq による次世代シーケンス (NGS) 解析

Illumina 社 TruSeq RNA Sample Prep Kit を用いて、試料の RNA 断片化、逆転写反応、2nd 鎖 cDNA 合成、及び末端修復 (3'A 付加) を行い、サンプル識別用インデックスタグを含むアダプターライゲーションを行った。PCR 増幅を施した後、AMPureXP ビーズによる精製・低分子除去 (約 150bp) を行い、シーケンシング用のサンプルとした。

HiSeq 2500 (Illumina 社製) を使用し、次の条件でシーケンシング解析を行った。(解析方法: Paired-End, 読み取り塩基長: 100 塩基 / 1 リード, 参考取得リード数: 2,000 万リード, 参考取得データ量: 2Gb)

#### (3) 出力データの解析・カタログ化

次世代シーケンサーで出力されたデータに関して次に挙げる情報処理を実施した。

ベースコール: 出力された解析生データ (画像データ) より、塩基配列のテキストデータを取得、フィルタリング: 所定のフィルタリングによるリードデータの選別、ゲノム配列へのマッピング: のデータを基に

なっている細菌のゲノム上塩基配列へマッピング処理。

#### (4) non-coding RNA 予測

得られたデータを基に下記に挙げるデータベースを用いて本実験結果との比較を行い、実際に得られたデータ基に予測を行った; A. Sanger Institute データベース Rfam (<http://rfam.sanger.ac.uk/>, 様々な細菌を含めた non-coding RNA 予測が可能), B. Los Alamos データベース; Oral pathogens non-coding small RNA prediction, 口腔内細菌に特化して non-coding RNA の予測)

#### (5) マイクロアレイの作製と発現解析

各細菌で得られた non-coding RNA 候補の塩基配列とゲノム情報から得られた全遺伝子の塩基配列情報から、マイクロアレイにスポットするプローブを決定し、Agilent 社にて各細菌用マイクロアレイを作成した。NGS 試料作成に伴い作成した RNA 試料をもとに mRNA-rich 試料から aRNA を増幅し、蛍光標識後、作成したアレイ上で反応させ、専用のソフトで発現解析を行った。

### 4. 研究成果

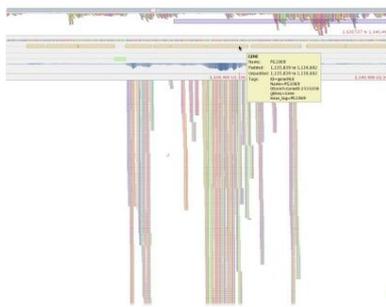
#### (1) 概要

当初、24 年度に準備した次世代シーケンシング用 RNA 試料を業者に委託する予定であったが、RNA に質的な問題 (分解) があることが判明し 25 年度は試料の取り直しを行うこととなった。また、何度か RNA を取り直したが、細菌 RNA のインタクトな状態での精製が極めて困難であったため、専門業者に依頼し、嫌気培養装置の精密検査を実施したところ、装置老朽化のため、ガスホースの劣化により培養装置内に酸素の混入が認められることが原因であることが判明した。30 年ほど同装置を使用しているが、初めてのケースであったため問題の抽出に時間がかかった。次世代シーケンシングにける貴重な試料でもあり、全菌種に対する全てのストレス負荷実験を、再度やり直したことで NGS 解析が 25 年度にずれ込み、全体の計画が遅れた。

#### (2) Pg non-coding RNA 解析

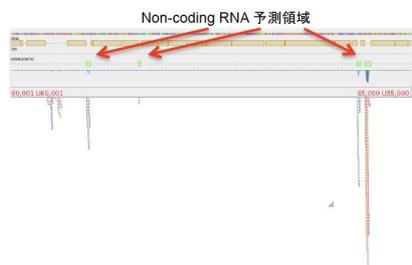
NGS 後ので得られたデータのクオリティチェックで、蛍光純度に基づく Chastity フィルタリングを通過したクラスターの割合は 97.35% であり、取得総塩基数は 4,948 (Mbp) であった。総リード数は 48,989,766 となり、フィルタ済み総塩基数に占める Q30 以上の High Quality Base の割合は 95% (Quality 平均値; 36.24) であった。さらに、アダプター配列除去 (トリミング) 後の配列長が Read1, Read2 共に 50 塩基以上となる read pair 情報のみを抽出すると、対象となる総リード数は 48,834,930 である。本リードを最終リードと見なして、Pg W83 株のレ

ファレンスゲノム ( NC\_002950 ) 上へ Mapping した。遺伝子領域および遺伝子間領域全ての発現領域とリード数を図に示す。



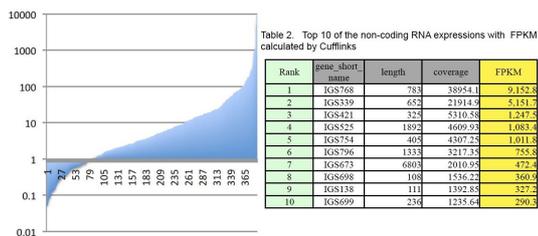
次に遺伝子間領域に限定した non-coding RNA の位置とリード数 ( 発現量 ) の検索を行った。Cufflinks Program を用いて、マッピング結果から、ゲノム上のリードカバー領域を算出し、マッピングデータに含まれる全リード本数と、各領域にマッピングされているリード数、各領域のサイズの情報を用いて、その領域の単位塩基長当たりの平均的な厚みを数値 ( FPKM 値 ) で算出した。

さらに、non-coding RNA 予測データベースとのマッチング検索を Los Alamos National Lab の Oral pathogens non-coding small RNA prediction データベースと、Sanger Institute データベース Rfam 11.0 とを利用し照合した。なお、Sanger Institute データベースでは、W83 株のゲノム配列と BLASTIN 検索を行い、相同性の高い領域を特定し、BLASTIN 検索でのヒットテーブルから、クエリ配列の全長が 80% 以上一致するものの抽出し、non-coding RNA 領域として 85 種を特定した。解析の例を示す。



本解析の結果、LANL で予測された全 649 候補 non-coding RNA 中、59.8% は予測された領域内から発現が認められたが、残りの 261 種 ( 40.2% ) には発現が全く認められなかった。また、LANL で予測されていない領域で intergenic な領域から明らかに発現が認められる領域が 56 カ所存在し、これらは新規の non-coding RNA の領域の可能性が高い。

図は縦軸を FPKM 値、横軸を累計数としている。FPKM 値が 10 以上 100 未満が 106 ( 16.3% )、100~1000 が 22 ( 3.4% )、1000 以上は 5 遺伝子 ( 0.8% ) であった ( ただし、rRNA 遺伝子はすべて除く )。また Table 2 は発現量でのトップ 10 の遺伝子を示した。



結論として、Pg へのストレスで、LANL の non-coding RNA 予測領域の 60% で発現が認められ、これらの発現量の差は極めて大きなものであった。また、予測領域以外からの発現も認められ、新規 non-coding RNA の可能性が考えられた。

### ( 3 ) Aa non-coding RNA 解析

Chastity フィルタリングを通過したクラスターの取得総塩基数は 2,356 ( Mbp ) となり、総リード数は 23,324,964 で、フィルタ済み総塩基数に占める Q30 以上の Higt Quality Base の割合は 93% ( Quality 平均値 ; 35.56 ) であった。さらに、トリミング後の配列長が Read1, Read2 共に 50 塩基以上となる read pair 情報のみを抽出すると、総リード数は 23,059,576 ( 98.9% 通過 ) となり、この各リードを Aa ATCC700685 のレファレンスゲノム ( GCA\_000604045.1 ) 上へ Mapping した。

発現量を表す FPKM 値をもとに解析をした結果、LANL で予測された全 630 ncRNA 中 307 ncRNA ( 48.7%、平均 size: 304 bp ) は予測された領域内からの発現が認められたが、残りの 323 ncRNA ( 51.3% ) には発現が全く認められなかった。また、FPKM を 10 以上と限定すると、発現が認められる ncRNA は 90 種 ( 14.3% / 630 ncRNA ) となり、その内訳は FPKM 値が 10 以上 ~ 100 未満が 76 ( 12.0% )、100 ~ 1000 が 12 ( 1.9% )、1000 以上は 2 遺伝子 ( 0.3% ) であった。また、Rfam での他種の既知 ncRNA 情報に基づくゲノム上への BLAST 検索の結果、104 種類の ncRNA 候補があり、うち 24 種に発現が認められ、かつ rRNA や tRNA でないものが 6 種あり、新規候補 ncRNA とした。

### ( 4 ) Td non-coding RNA 解析

Chastity フィルタリングを通過したクラスターの取得総塩基数は 2,305 ( Mbp ) となり、総リード数は 22,818,408 で、フィルタ済み総塩基数に占める Q30 以上の Higt Quality Base の割合は 94% ( Quality 平均値 ; 36.01 ) であった。さらに、トリミング後の配列長が Read1, Read2 共に 50 塩基以上

上となる read pair 情報のみを抽出すると、総リード数は 22,617,222 (99.1%通過)となり、ここで得られた各リードを Td ATCC35405 のレファレンスゲノム (NC\_002967) 上へ Mapping した。

発現量を表す FPKM 値をもとに解析をした結果、LANL で予測された全 643 ncRNA 中 349 ncRNA (54.3%, 平均 size; 287 bp) は予測された領域内から発現が認められたが、残りの 294 ncRNA (45.7%) には発現が全く認められなかった。また、FPKM 値が 10 以上のものに限定すると、発現が認められる ncRNA は 246 種 (38.3% / 643 ncRNA) となり、その内訳は FPKM 値が 10 以上~100 未満が 154 (30.0%), 100~1000 が 80 (12.4%), 1,000 以上は 12 遺伝子 (1.9%)であった。また、Rfam での他種の既知 ncRNA 情報に基づくゲノム上への BLAST 検索の結果、63 種類の ncRNA 候補があり、うち 14 種に発現が認められ、その中で rRNA や tRNA でないものが 8 種あり、新規候補 ncRNA とした。

#### (5) Tf non-coding RNA 解析

Tf のデータ採取に関しては 2 度のトラブルがあった。Tf は難培養であることが知られている。少量の液体培地での培養では問題がなかったが、大量培養を施したのち RNA 試料を分離・精製し NGS 解析したところ、他菌種がコンタミしていたことが判明し、再度、培養からやり直した経緯があった。また、大量に試料を準備するため液体培地での培養にこだわったが、液体培養ではコンタミしていなくても、何度試しても菌の増殖が極めて遅く (OD600 で約 0.25~0.3), 時間をかけても集菌するタイミングがつかめず、結局、インタクトな RNA を精製することが困難であったことから、寒天培地による培養に変えて行った。寒天培地上では、さらにコロニーの増殖速度が遅く、うすすらとした一層のコロニーしか形成されないため、プレートを 50 枚ほど用意し、すべてのプレートから少量ずつ掻き取ることで最終的にコンタミのない試料を得た。さらに、NGS 解析において当初予定をしていた本菌株のレファレンスゲノムと取得データ合わず、再度ゲノム情報を確認したところ、平成 27 年 8 月に登録されていた菌株が変更されていた。そこでパーソナルシーケンスではあるものの、実験に供した菌株と同様のものをレファレンスゲノム (2015.09.14 リリース) として再解析を進めた。

Chastity フィルタリングを通過したクラスタの取得総塩基数は 2,366 (Mbp) となり、総リード数は 23,421,780 で、フィルタ済み総塩基数に占める Q30 以上の Higt Quality Base の割合は 96% (Quality 平均値; 36.85) であった。さらに、トリミング後の配列長が Read1, Read2 共に 50 塩基以上となる read pair 情報のみを抽出すると、

総リード数は 23,305,532 (99.5%通過)となり、各リードを Tf ATCC43037 のレファレンスゲノム (GCF\_001006485.1) 上へ Mapping した。

発現量を表す FPKM 値をもとに解析をした結果、LANL で予測された全 731 ncRNA 中 412 ncRNA (56.4%, 平均 size; 307bp) は予測された領域内から発現が認められたが、残りの 319 ncRNA (43.6%) には発現が全く認められなかった。また、FPKM >10 とした場合には発現が認められる ncRNA は 307 種 (42.0%) となり、その内訳は FPKM 値が 10 以上~100 未満が 201 (27.5% / 731 ncRNA), 100~1000 が 91 (12.4%), 1000 以上は 15 遺伝子 (2.1%) であった。また Rfam での他種の既知 ncRNA 情報に基づくゲノム上への BLAST 検索の結果、61 種類の ncRNA 候補があり、うち 16 種に発現が認められたが rRNA か tRNA であった。

#### (6) マイクロアレイによる non-coding RNA および遺伝子発現解析

全 coding RNA (遺伝子) に加え、解析結果より得られた non-coding RNA の両方のプロンプを設計し、発現解析用のカスタムマイクロアレイを作製した。各環境変化に対しての遺伝子および non-coding RNA の発現プロファイリングを検証し現象として捕らえカタログ化を進めている (現在進行中)。

#### 5. 主な発表論文等

[雑誌論文] (計 0 件)

[学会発表] (計 0 件)

[図書] (計 0 件)

[産業財産権]

○出願状況 (計 0 件)

○取得状況 (計 0 件)

[その他]

特記事項なし

#### 6. 研究組織

##### (1) 研究代表者

平塚 浩一 (HIRATSUKA, Koichi)

日本大学・松戸歯学部・教授

研究者番号: 80246917