

ガイドクラスタリングとネットワーク分割法 を用いた土壌メタゲノムからの新規キシロース代謝酵素の探索

根本賢作¹, Harry A Moesa^{2, 3}, 千葉啓和^{3, 4},
竹山春子¹, 藤渕航^{1, 3, 5, 6, 7}

メタゲノム解析は難培養性微生物の遺伝子の収集を可能にしたが遺伝子が由来するゲノムを判別することは難しい。我々はこの問題を解決するためにクラスタ解析とネットワーク分割アルゴリズムを併用した判別手法を開発した。難培養性微生物に由来する糖代謝遺伝子 *xyIA* はバイオエタノールの生産効率に寄与している。本手法を適用して土壌メタゲノムから *xyIA* 遺伝子が由来するゲノムを推定した。今学会でその成果を報告する。

A SEARCH FOR NEW XYLOSE METABOLIC ENZYMES IN SOIL METAGENOME BY GUIDE CLUSTERING AND NETWORK PARTITIONING.

Kensaku Nemoto¹, Harry A Moesa^{2, 3}, Hirokazu Chiba^{3, 4},
Haruko Takeyama¹, Wataru Fujibuchi^{1, 3, 5, 6, 7}

Metagenome analysis has provided a gene catalog derived from unculturable microorganisms but it is difficult to identify origin of retrieved gene; where do they come from? To address this problem, we developed a new method that combines a general clustering with a network partitioning algorithm named BRIM. *xyIA* gene encoding xylose (glucose) isomerase also contributes in the productivity of bioethanol. Using the method, we estimated a genome that contains *xyIA* genes. We will report these results in this conference.

1. 結論

現在、大腸菌や酵母菌といった様々な微生物が産業や学術研究に広く利用されている。実際、微生物が持つ遺伝子から産業に利用可能な酵素や抗生物質など価値の高い物質が発見される事例は多い[1]。その一方で人類が単離培養する事が可能な微生物種は微生物種全体の 0.1%にも満たない事が知られている[2]。難培養性微生物の培養には特殊環境や微生物間の相互作用が不可欠である。したがってその酵素は生体内では易培養性微生物由来の既知の酵素と同機能を示しながら異なる活性条件や物理的性質を示すと考えられている。

環境中に含まれるゲノム(メタゲノム)を網羅的に抽出するメタゲノム解析は培養工程を含まないという特徴があり培養が困難である微生物でもその遺伝子を収集することが理論上可能である[3]。実際の解析は2つのアプローチ 1)配列基準, 2)活性基準で実施される。配列基準の手法では標的遺伝子の保存領域からプライマーを設計し、PCR法を用いてメタゲノム中に潜在する保存領域を包括的に増幅させる。既知配列であれば理論上メタゲノムから包括的に標的遺伝子を収集することが可能である[4]。欠点としては標的遺伝子を発現させないために収集された標的遺伝子の機能及び活性が保証されていない点があげられ、標的遺伝子を収集しても既知の酵素より工業的に劣っていたケースも存在する。一方、活性基準の手法では 1)メタゲノム DNA を物理的に断片化, 2)プラスミドライブラリーに導入, 3)形質転換株の作成, 4)栄養源で標的遺伝子含む形質転換株を選抜。これらの手順を経て標的遺伝子を収集する[5]。この手法では酵素活性情報のみを用いて選抜しているため進化的な類縁関係が遠い遺伝子を収集することが可能である。ただし、使用する形質転換株によって陽性クローンの作成効率は非常に低いことが問題点としてあげられる。例えば、大腸菌を使用した実験ではアミダーゼでは 23,000 クローン中 1 クローン[6]、抗菌物質では 113,700 クローン中 2 クローン[7]が活性を示す結果となっている。メタゲノム中から収集された遺伝子がどの

¹ 早稲田大学大学院 先進理工学研究所

² 東京大学医科学研究所 ヒトゲノムセンター

³ 独立行政法人 産業技術総合研究所 生命情報工学研究センター

⁴ 独立行政法人 自然科学研究機構 基礎生物学研究所

⁵ 東京医科歯科大学大学院 生命情報科学教育部

⁶ 東京大学 新領域情報生命科学専攻

⁷ 京都大学 iPS 細胞研究所

¹ Faculty of Science and Engineering, Waseda University.

² Human Genome Center Institute of Medical Science University of Tokyo.

³ Computational Biology Research Center, Advanced Industrial Science and Technology.

⁴ National Institute for Basic Biology, National Institutes of Natural Sciences.

⁵ School of Biomedical Science, Tokyo Medical and Dental University.

⁶ Graduate School of Frontier Sciences, The Univ. of Tokyo.

⁷ Center for iPS Research and Application, Kyoto University.

ゲノムに由来するのか？同サンプル中のどの配列の近傍種なのか？この問題を解決する手法は非常に限られている。特にゲノムを単離出来ない難培養性微生物の場合は配列の相同性より判断するために BLAST[8]のデータベース検索ソフトに頼るしか手が無いのが現状である。

エネルギー供給における石油燃料の枯渇や地球温暖化の深刻化といった問題の解決策の一つとして、循環可能であり二酸化炭素の排出の少ないエネルギーが求められ、バイオエタノールが注目されている。原料である植物バイオマスは光合成により CO₂ を吸収して育つために地球温暖化に与える影響が少なく循環可能なエネルギーである。非可食部である木質を原料とするバイオエタノール生産の研究は以前から行われているが、発酵に使用される酵母は木質の成分ヘミセルロースを構成する五単糖、D-xylose を代謝出来ない。そのためバイオエタノール生産効率の上昇を目指し、バクテリアにおける D-xylose 代謝酵素発現遺伝子 *xyIA* を酵母内で発現させる研究が行われている [9][10]。

本研究では、土壌メタゲノムから単離株由来の *xyIA* とは活性条件が異なる難培養性微生物に由来する *xyIA* 遺伝子を取得することを目指した。配列基準の手法では収集が難しかった進化的に遠縁の配列を同定するためにクラスタリングとネットワーク分割アルゴリズム (BRIM) を併用した手法を開発した (図 1,2)。さらに本手法を活用しメタゲノムから取得した遺伝子に関して由来する生物種/郡の推測を試みた。

2. 実験方法

2.1 土壌サンプルの採取

筑波市の山中 2 箇所 (T1, T2) 及び白樺湖 (S1) で土壌サンプルの採取を行った。いずれも樹木が生育している土壌であり、T1 はブナ、T2 はスギ、S1 はシラカバが生育する土壌より採取した。

2.2 土壌サンプルからの未知 *xyIA* 遺伝子の抽出

茨城県筑波市の 3 箇所から採取した土壌から ISOIL(NIPPON GENE)を使用して DNA を抽出し、約 5kbp の断片にランディングした 3 種類の土壌 DNA サンプルをテンプレートとして PCR を行った。PCR プライマーは *xyIA* 遺伝子の保存領域を元に以下の 2 組を設計した (*xyIA* degenerate forward primer1 (20bp): 5'-TGGGGNGGNCNGARG GNTA-3', *xyIA* degenerate reverse primer1 (21bp): 5'-RAAYTSRTCNGTRTCCARCC-3', *xyIA* degenerate forward primer2 (24bp): 5'-TGTGTTTTGGGCGGNMKNKNGANG G-3', *xyIA* degenerate reverse primer2 (21bp): 5'-ATGGCCCGCCADNKKNKCRGTG-3'

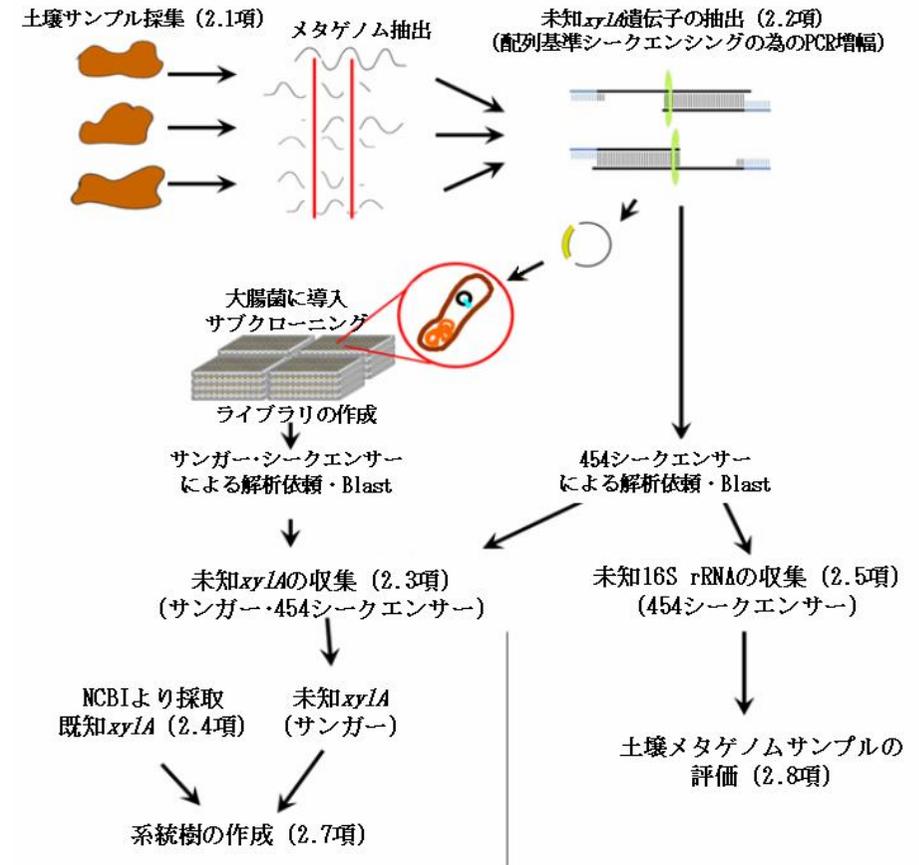


図 1. 実験フローチャート

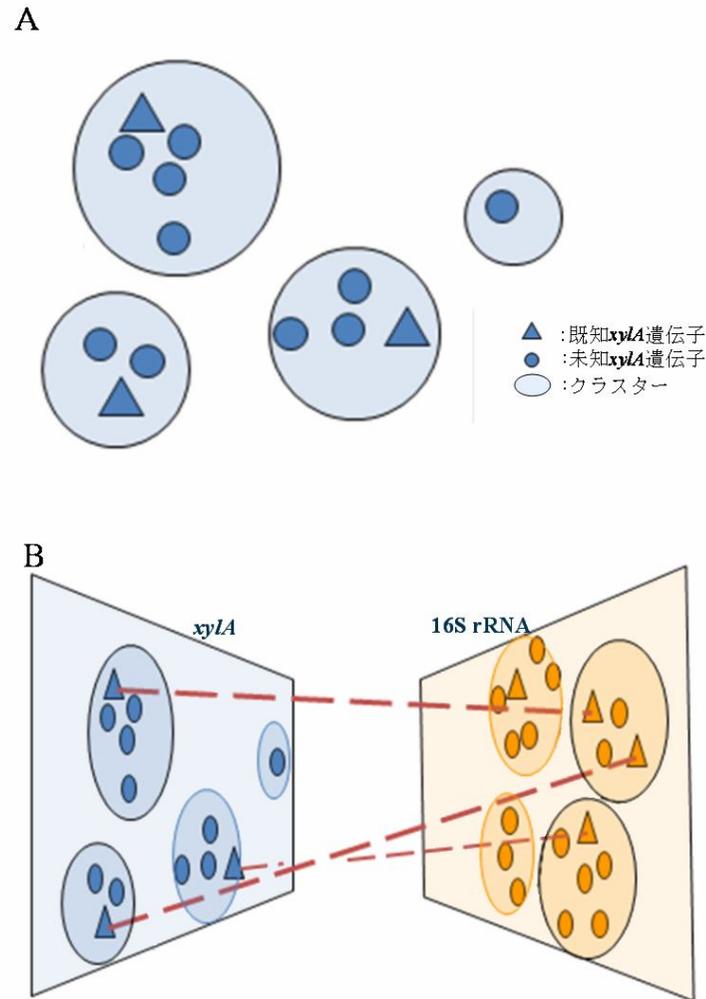


図2. クラスター解析及びBRIMアルゴリズムの適用におけるチャート
A) 配列データのクラスター解析の実行. B) 作成されたクラスター同士の
taxonomy ID による接続及びBRIMアルゴリズムによるネットワーク分割.

). PCR産物はゲル電気泳動でサイズを確認した後に切り出し精製を行い、pCR-blunt TOPO II ベクター(invitrogen)に導入し、ヒートショック法で形質転換を行いSOC培地で37℃、1時間の復活培養の後抗生物質を含むLB寒天培地に植菌した。サブクロニングによって得たコロニーを土壌DNAサンプル毎に72個ピックアップし、抗生物質添加液体LB培地で培養した後にグリセロールストックを作成した。採取したサンプルはプラスミド抽出によってインサートが保持されていることを確認した。

2.3 土壌サンプルからの未知 *xyIA* 遺伝子配列の収集

クローニングホスト及びライブラリーには *E.coli* EC100 (Epicentre) を用いた。LB液体培地 (10 g / L Bacto trypton, 5 g / L yeast extract, 5 g / L NaCl, pH 7.0) と抗生物質カナマイシン (12.5 μg / mL) で培養した。プラスミドのインサート配列をサンガー法とPyrosequencing法で決定した(本解析はTAKARA BIOにて実施)。次に決定した配列データが *xyIA* であることを確認するために3種類のコドンフレームでアミノ酸に翻訳し、途中終止コドンを含む配列は疑似遺伝子とみなし取り除いた。最後にclassIで報告されている配列モチーフ1) GREG[YSTA], 2) [LIVM]EPKPx[EQ]P ([YSTA]はYSTAのいずれかのアミノ酸, Xは任意のアミノ酸を表す) を含む配列を選抜し *xyIA* 遺伝子とした[11]。また、酵素の分布を土壌サンプル毎に調べるために翻訳後の各配列データに対しクラスIのキシロース代謝酵素の活性部位(181E, 183K, 217E, 220H, 245D, 255D, 257D, 287D)の有無を確認した。

2.4 データベースからの *xyIA* 遺伝子配列の収集

NCBI (<http://www.ncbi.nlm.nih.gov/>) のGenBankデータベース(600,752種, 286,730,369,256配列の遺伝子データベース)より既知の *xyIA* 遺伝子のアミノ酸配列情報を入手した。"xylA"及び"xylose isomerase"をキーワードとして配列情報を検出し、既知 *xyIA* アミノ酸配列935本を収集した。次に「DNAデータベース(DDBJ/EMBL/GenBank=INSD)総覧と検索,ダウンロード」(<http://lifesciencedb.jp/ddbj/download.cgi>)より各配列情報の固有ID(Accession ID)と微生物種及び株レベルの固有ID(taxonomy ID)の対応データをダウンロードし、NCBIのFTPサーバ由来の微生物種対応表(categories.dmp)と関連付けて、全ての既知 *xyIA* 遺伝子に種のtaxonomy IDを付与し、既知 *xyIA* 遺伝子と定義した。

2.5 土壌サンプルからの16S rRNA配列の収集

16S rRNA遺伝子の保存領域からPCRプライマーを設計した(F8: 16SrRNA forward primer (20bp): 5'-AGAGTTTGATCCTGGCTCAG-3', R357: 16SrRNA reverse primer (15bp): 5'-CTGCTGCCTYCCGTA-3')。作成したプライマーを用いてPCR法により5kb長に調整した土壌DNAサンプルから増幅断片を収集しパイロシークエンサ

ー (454 シークエンサー) を用いて塩基配列を同定した。次に収集された配列が 16S rRNA 遺伝子であることを確認するため、塩基配列データベース nt に対して BLASTn を実行し、配列相同性が 80% 以下の配列を削除した。配列長を 250 塩基に揃え、トリミングした。

2.6 データベースからの 16S rRNA 配列の収集

2.4 項と同様に GenBank データベースより、既知の 16S rRNA の塩基配列情報を入手し、"16S"及び"16S ribosomal"をキーワードとして配列情報を検出し、2.4 項の既知 *xytA* アミノ酸配列と共通する taxonomy ID を持つ既知 16S rRNA の塩基配列 333 本を収集した。

2.7 OTU 数の推定

97%以上の相同性を確認した 16S rRNA 配列は同種由来の 16S rRNA と定義した[12]。続いて土壌サンプルに含まれる全 16S rRNA に対して、どれほどの割合を採取する事が出来たのか確認した。97%以上の相同性が確認された土壌由来の 16S rRNA 配列を一つのグループ (operational taxonomic unit, OTU) に纏め、OTU の数と土壌サンプル中の微生物の種数を対応付けた。遺伝子解析ソフト *mothur* (<http://www.mothur.org/>) を使用してコマンド `rarefaction.single()`により土壌由来 16S rRNA の本数より OTU 数を計算した[13] [14]。また、`rarefaction.single()`により生成された rarefaction curve のデータ及び *mothur* 付属のコマンド `rarefaction(calc=solow, size=2000)` (または `size=5000, size=10000`) を使用し、16S rRNA が 10,000 本増加した場合の OTU 数の収束を通してサンプル土壌中に含まれる 16S rRNA 遺伝子の種数の推定を行った。

2.8 クラスタリング

取得した配列に対し解析ソフト *clustalW2* (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) を使用して NJ 法によるアラインメントを実行し、アラインメント後の配列を 2.2 項、2.5 項で使用した PCR プライマーに合わせてトリミングした。次にプログラム *mothur* を使用して `dist.seqs(fasta=データ名,output=square)`を実行した。全対全で配列類似度 (identity を採用)を算出し距離行列を出力した[15]。

距離行列に対し、R 言語のパッケージ *stats* に含まれる階層的クラスタ解析実行関数 `hclust(method="average",...)`を適用し群平均法によりクラスタリングを実行した。

群平均法はクラスタリングの手法の一つであり、二つのクラスタ間の距離 d を互いのクラスタの全要素同士の距離平均とする手法で、数式としては

$$d(C_1, C_2) = \frac{1}{|C_1||C_2|} \sum_{x_1 \in C_1} \sum_{x_2 \in C_2} d(x_1, x_2)$$

と表される。但し $d(x_1, x_2)$ はクラスタの要素 x_1, x_2 間の距離であり、 $|C|$ はクラスタ C 内の要素の個数である。

2.9 クラスタ間の接続

生成された *xytA* クラスタと 16S rRNA クラスタを系統情報に基づいて関連づけるために、既知 *xytA* と既知 16S rRNA を taxonomy ID を介して行列形式で対応付けた。具体的には既知 *xytA* を行に、既知 16S rRNA を列に並べ、互いの taxonomy ID が一致するセルに含まれる値を 1、一致しないセルに含まれる値を 0 とした行列を作成し、アルゴリズムへの入力に使用した。

次に両クラスタに対し BRIM (bipartite recursively induced modules) アルゴリズムを適用した。BRIM アルゴリズムの原理は以下の通りである。

頂点と辺より構成されるネットワークの分割において、分割の際の頂点の分け方に対し、重さ Q (modularity Q) が設定される (図 3)。 Q 値は切断される辺の重さの和であるため、多数の辺を切断する分割程 Q 値が高く、僅かな辺の切断による分割は Q 値が低くなる。 Q 値を最小にする分割方法を見つけるのがネットワークの分割問題である。ネットワーク分割の際に任意の頂点 ij 間の辺が分割されるとすると、 Q 値は

$$Q = \frac{1}{2m} \sum_{i,j} (A_{ij} - P_{ij}) \delta(g_i, g_j)$$

と表される[16]。但し m はネットワーク上の頂点の数、 A_{ij} はネットワークの頂点 ij を結ぶ辺の有無を 1 と 0 で示した行列、 P_{ij} は各辺の重さを表す補正行列、 $\delta(g_i, g_j)$ は頂点 i, j 間異なるクラスタに分配された場合は 1、同一のクラスタに分配された場合は 0 を示す。

最小の Q 値を求めることで、相同性の高い配列で構成されたクラスタをより相同性の高い二つのクラスタに分割する事が可能であり、既知遺伝子と同じクラスタに存在する遺伝子は既知遺伝子の近傍種、既知遺伝子の存在しないクラスタに含まれる遺伝子を難培養性微生物の遺伝子と推測した。

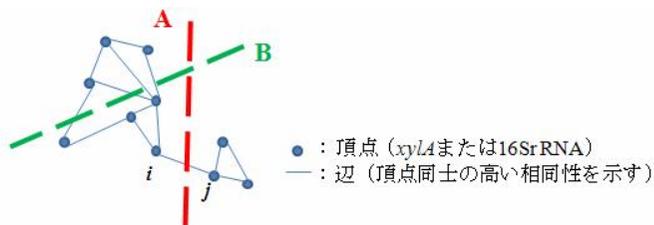


図3. ネットワークの分割の差による Q 値の高低
A)辺の切断が少なく Q 値の低い分割 B)辺の切断が多く Q 値の高い分割

3. 結論

3.1 土壌メタゲノムに潜在する *xylA* 遺伝子

3種類の土壌サンプル T1, T2, S1 からサンガーシーケンサーより配列長 200aa の未知 *xylA* アミノ酸配列を 140 本, 112 本, 112 本, 454 シークエンサーより配列長 200aa の未知 *xylA* アミノ酸配列を 17,524 本, 16,000 本, 12,512 本を回収した。回収した配列データを統合し *xylA* 遺伝子の保存領域を用いたフィルタリングを実行した。その結果, 各々の土壌サンプルより 6,156 本, 3,720 本, 3,046 本の未知 *xylA* アミノ酸配列を得た。

表1 *xylA* 遺伝子配列における活性部位数の土壌毎の比較

| 土壌サンプル | 活性部位数(個) | | | | | | | 計 |
|--------|----------|------|------|------|---|---|---|------|
| | 2 | 3 | 4 | 5 | 6 | 7 | 8 | |
| T1 | 91 | 2711 | 244 | 0 | 0 | 0 | 0 | 3046 |
| T2 | 155 | 86 | 5674 | 241 | 0 | 0 | 0 | 6156 |
| S1 | 107 | 332 | 273 | 3308 | 0 | 0 | 0 | 3720 |

3.2 土壌メタゲノムサンプルの評価

三種類の土壌サンプルから得られた土壌由来 16S rRNA の OTU 数は 354, 457, 397 だった。Solow's estimator を使用した予測では 10,000 本の土壌由来 16S rRNA を追加することで, それぞれの OTU 数は 893, 944, 936 と推測された。またサンプル土壌からの 16S rDNA の抽出率は 39.6%, 48.4%, 42.4%であった。

表2 16S rRNA 遺伝子の OTU 数及び配列追加時の OTU 数の推定

| 土壌サンプル | 16Sr RNA | OTU 数 | 配列追加時 OTU | 抽出率(%) |
|--------|----------|-------|-----------|--------|
| T1 | 1280 | 354 | 893 | 39.6 |
| T2 | 1584 | 457 | 944 | 48.4 |
| S1 | 2170 | 397 | 936 | 42.4 |

3.3 未知 *xylA* の推定の為の土壌中の *xylA* の同定

xylA 遺伝子と 16S rRNA 遺伝子をサンプルデータとしてクラスター解析を行った。その結果, 470, 837, 1225, 190, 454, 819 個の *xylA* 遺伝子クラスターと 853, 858, 859, 888, 887, 888 個の 16S rRNA 遺伝子クラスターが得られた。*xylA*-16S rRNA 間の相関がみられたクラスターは 75, 72, 57, 74, 68, 49 個で, 最多で 584 本の配列を含む *xylA* 遺伝子サンプルのクラスターが得られた(表3)。また本実験においては 16S rRNA 遺伝子配列のデータベースを利用した解析を行うため 16S rRNA 遺伝子サンプルのクラスターと対応関係を得た *xylA* 遺伝子サンプルのクラスターを既知 *xylA* 遺伝子サンプルのクラスター, 対応関係を得られなかった *xylA* 遺伝子クラスターを未知 *xylA* 遺伝子サンプルのクラスターとし, 解析及び考察を行った。BRIM によって得た *xylA*, 16S rRNA の各クラスターを遺伝子系統樹, 配列データの距離行列, 既知配列における taxonomy ID によって対応付け, 未知 *xylA* の近傍種の推定の基準とし, 既知 *xylA* と同じクラスターに含まれない未知 *xylA* を難培養性微生物の *xylA* として 16S rRNA クラスターとの関連付けを行った。

表3 クラスター解析及び BRIM アルゴリズム適用結果

| | 土壌サンプル | T1 | | | S1 | | |
|-------------|-------------|----------|----------|-----------|----------|----------|----------|
| | | 配列相同性 | | | 80% | 90% | 95% |
| <i>xylA</i> | 既知遺伝子クラスター数 | 75(86) | 72(85) | 57(80) | 74(79) | 68(78) | 49(79) |
| | 未知遺伝子クラスター数 | 145(384) | 308(752) | 394(1175) | 68(111) | 173(376) | 275(740) |
| | クラスター数(合計) | 220(470) | 380(837) | 451(1255) | 142(190) | 241(454) | 324(819) |
| 16S | 既知遺伝子クラスター数 | 31(80) | 32(85) | 31(86) | 21(79) | 31(78) | 30(79) |
| | 未知遺伝子クラスター数 | 208(773) | 208(773) | 208(773) | 253(809) | 253(809) | 253(809) |
| | クラスター数(合計) | 239(853) | 240(858) | 239(859) | 274(888) | 284(887) | 283(888) |

※ () 内は 1 本の配列で生成されたクラスターを含めた個数

4. 結言

本研究ではバイオインフォマティックスのアプローチを用いて土壌メタゲノム由来の配列の *xylA* 未知/新規遺伝子の探索を試みた。 *xylA* は活性部位に基づいてクラス I (*Streptomyces* 種), クラス II (*Thermotoga* 種, *Bacillus licheniformis*, *Escherichia coli*) の 2 種類に分類される。酵素の分布を土壌サンプル毎に調べるため、翻訳後の各配列データに対しクラス I のキシロース代謝酵素の活性部位(181E, 183K, 217E, 220H, 245D, 255D, 257D, 287D)の有無を確認した[17]。また、クラス II では活性部位に保存されたアミノ酸が存在しないために選抜基準として採用しなかった。

クラスター解析からは、土壌サンプルから抽出された *xylA* 遺伝子は最多で 1,255 グループ、16S rRNA 遺伝子配列は最多 876 グループに分類できることが明らかになった。また *xylA*-16S rRNA 間で taxonomy ID によって対応付けられたクラスターは両遺伝子サンプルのクラスターで最多 75 個得られた。一方で対応付けられなかった *xylA* 遺伝子サンプルのクラスターに関しても複数本の *xylA* 遺伝子サンプルが含まれるクラスターが得られた。

3.2 項の結果の OTU 数より土壌メタゲノム由来の 16S rRNA 遺伝子サンプルは最少で 354 個、最多で 457 個のクラスターを形成すると考え、それらのデータに既知の 16S rRNA 遺伝子サンプル 333 本を加えて、16S rRNA 遺伝子サンプル全体でのクラスターは 354 個から 790 個の範囲の値になると予想した。結果として予測値以上のクラスターが得られてしまったが、これは BRIM アルゴリズムによる過分割の結果だと考えられる。 *xylA* 遺伝子サンプルに関しては *xylA* 遺伝子を持たない微生物も存在するため、得られるクラスターは 16S rRNA のクラスター数より少ないと考えられ、16S rRNA の予測値と合わせて、90%の配列相同性を基準としてクラスターを生成したデータが *xylA* 遺伝子の同定に適していると判断した。

90%の配列相同性を基準としてクラスターにおいて、複数の配列より構成され、かつ既知の遺伝子と関係付けられたクラスターはサンプル S1 で 68 個、サンプル T1 で 72 個得られた。これらのクラスターに含まれた未知 *xylA* 配列を既知微生物の近傍種の遺伝子だと推定した。また、同様に複数の配列より構成され、かつ既知の遺伝子と関係付けられなかったクラスターはサンプル S1 で 173 個、サンプル T1 で 308 個得られた。これらのクラスターに含まれる遺伝子を難培養性微生物の遺伝子だと推定した。

以上よりクラスター解析のアルゴリズムによる *xylA* 遺伝子サンプル、16S rRNA 遺伝子サンプルのクラスターの対応付けの試みは現在のところ成功していると言える。

作成した系統樹や遺伝子サンプルデータの分布から、残った taxonomy ID で対応付けられないクラスター同士の対応関係の推定を行う事が本研究の今後の目標である。

研究の結果によってはメタゲノム解析において、本アルゴリズムを使用した遺伝子の微生物種推定に有用である事が期待される。今後は遺伝子配列の同定結果を元に、実

験株からの採取が難しい難培養性微生物の *xylA* 遺伝子配列を選出し、酵母のエタノール発酵を利用したバイオエタノール生産技術の研究に利用する予定である。

5. 謝辞

本研究を進めるにあたり、御指導と御高配を賜りました独立行政法人産業技術総合研究所 生命情報工学研究センター 細胞機能設計チーム 田中道廣博士に深く感謝の意を表します。

本研究を進めるにあたり快く受け入れていただき、御助言、御協力を頂いた同チームの皆様にも厚く御礼申し上げます。

参考文献

- 1) Yun J, Kang S, Park S, Yoon H, Kim MJ, Heu S, Ryu S. 2004. Characterization of a novel amylolytic enzyme encoded by a gene from a soil-derived metagenomic library. *Appl Environ Microbiol.* 70:7229-35.
- 2) Amann RI, Ludwig W, Schleifer KH. 1995. Phylogenetic identification and in situ detection of individual microbial cells without cultivation. *Microbiol Rev* 59:143-169
- 3) Cowan D, Meyer Q, Stafford W, Muyanga S, Cameron R, Wittwer P. 2005. Metagenomic gene discovery: past, present and future. *TRENDS in Biotechnology* Vol.23 No.6 321-9
- 4) Lena Tasse, Juliette Bercovici, Sandra Pizzut-Serin, et al. 2010. Functional metagenomics to mine the human gut microbiome for dietary fiber catabolic enzymes. *Genome Res.* published online.
- 5) Karin Hjort, Maria Bergstrom, Modupe F. Adesina, Janet K. Jansson, Kornelia Smalla & Sara Sjöling. 2009. Chitinase genes revealed and compared in bacterial isolates, DNA extracts and a metagenomic library from a phytopathogen-suppressive soil. *FEMS Microbiol Ecol* 71 (2010) 197-207
- 6) Gabor EM, de Vries EJ, Janssen DB. 2004. Construction, characterization, and use of small-insert gene banks of DNA isolated from soil and enrichment cultures for the recovery of novel amidases. *Environ Microbiol* 6 :948-958
- 7) Lim HK, Chung EJ, Kim JC, Choi GJ, Jang KS, Chung YR, Cho KY, Lee SW. 2005. Characterization of a forest soil metagenome clone that confers indirubin and indigo production on *Escherichia coli*. *Appl Environ Microbiol.* 71:7768-77.
- 8) Scott McGinnis and Thomas L. Madden. 2004. BLAST: at the core of a powerful and diverse set of sequence analysis tools. *Nucleic Acids Research*, Vol. 32, Web Server issue.
- 9) Catharine J. Moes, Isak S. Pretorius and Willem H. Zyl. 1996. Cloning and expression of the *Clostridium thermosulfurogenes* D-xylose isomerase gene (*xyLA*) in *Saccharomyces cerevisiae*. *Biotechnology Letters* Volume 18, Number 3, 269-274.
- 10) M Walfridsson, X Bao, M Anderlund, G Lilius, L Bülow and B Hahn-Hägerdal. 1996. Ethanolic fermentation of xylose with *Saccharomyces cerevisiae* harboring the *Thermus thermophilus xylA* gene,

which expresses an active xylose (glucose) isomerase. *Appl. Environ. Microbiol.* December 1996 vol. 62 no. 12 4648-4651

11) Aloysius Willhelmus Rudolphus Hubertus Teunissen. 2011. XYLOSE ISOMERASE GENES AND THEIR USE IN FERMENTATION OF PENTOSE SUGARS. Patent Application Publication.

12) Patrick D. Schloss and Jo Handelsman. 2005. Introducing DOTUR, a Computer Program for Defining Operational Taxonomic Units and Estimating Species Richness. doi: 10.1128/AEM.71.3.1501-1506.2005 *Appl. Environ. Microbiol.* March 2005 vol. 71 no. 3 1501-1506

13) Marko Kuyper, Maurice J. Toirkens, Jasper A. Diderich, Aaron A. Winkler, Johannes P. van Dijken, Jack T. Pronk, *Article first published online: 7 APR 2006. Evolutionary engineering of mixed-sugar utilization by a xylose-fermenting *Saccharomyces cerevisiae* strain.

14) PD Schloss, SL Westcott, T Ryabin. 2009. Introducing mothur: open-source, platform-independent, community-supported software for describing and comparing microbial communities. *Appl. Environ. Microbiol.* December 2009 vol. 75 no. 23 7537-7541

15) Patrick D. Schloss, Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, Brian B. Oakley, Donovan H. Parks, Courtney J. Robinson, Jason W. Sahl, Blaz Stres, Gerhard G. Thallinger, David J. Van Horn, and Carolyn F. Weber. 2009. Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied And Environmental Microbiology*, Dec. 2009, p. 7537–7541

16) Michael J. Barber. 2008. Modularity and community detection in bipartite networks. arXiv:0707.1616v3 [physics.data-an] 5 Nov 2007

17) Richard D. Whitaker, Yunje Cho, Jaeho Cha, H. L. Carrell, Jenny P. Glusker, P. Andrew Karplus, and Carl A. Batt. 1995. Probing the Roles of Active Site Residues in D-Xylose Isomerase. *The journal of biological chemistry*.