

メタゲノムデータからの生物種系統特異的な機能モジュールの抽出

EXTRACTION OF PHYLUM-SPECIFIC FUNCTIONAL MODULES FROM METAGENOMIC DATA

京都大学 薬学研究科 統合ゲノミクス分野 化学生命科学研究室 張 晨

背景と目的

メタゲノミクスは多数の微生物群を含む環境サンプルに対し、単離・培養の過程を経ずに直接ゲノム情報を網羅的に読み取って解析する方法論である。現在、主流な研究は海洋水、土壌、そしてヒトを対象に行われており、ヒトに関するメタゲノムサンプルは鼻、口腔、皮膚、尿生殖路、そして腸内と多岐に渡っている。その中でも、腸内のバランスはヒトの加齢状態や食習慣に密接に関わっていると考えられ、メタゲノミクスを応用した腸内サンプルからのヒトの潜在的な健康状態を予測するシステム等が期待されている。

本研究では、年齢・性別の異なる 13 名の日本人腸内細菌叢から得られたメタゲノムデータを利用し、各状態特異的に見つかる生物種群や機能遺伝子群の特徴抽出を目的とし解析を行った。メタゲノムデータに含まれる大量の遺伝子情報を効率的に解析するにはバイオインフォマティクスのアプローチが必須であり、そのための知識データベースや解析手法を有効活用することが求められる。KEGG MODULE データベース[1]では、生物種間で特に保存性の高い一連の生体反応経路やタンパク複合体を機能単位(モジュール)として定義している。モジュールを構成する遺伝子に対しメタゲノムデータをマッピングすることで機能分類に基づく生物種分布やサンプル間の違いを俯瞰し、特徴的なモジュールとそれらの因果関係を説明する成分の抽出を行った。

方法

Kurokawa ら[2]によって採取された 13 人の日本人腸内サンプルは 7 人の成人（男 4・女 3）と 2 人の子供（男 1・女 1）、そして 4 人の乳児（男 2・女 2）で構成されている。これらのデータから得られた 662,548 遺伝子に対し、KAAS[3]システムを用いて遺伝子機能のアサインを行った。遺伝子機能は KEGG Orthology (KO)に定義されている識別番号によって区別し、その番号を基に 571 の KEGG MODULE に対し遺伝子マッピングを行った。また同様の遺伝子データに対し BLAST 検索をかけ、ベストヒット遺伝子の由来生物種に基づき Phylum(門)階層で系統情報の付与を行った。モジュール、系統、サンプルの 3 つの情報が与えられた 153,309 個の遺伝子から、各対応を表す遺伝子のカウントテーブル表を作成し、多変量解析手法である O2PLS[4]を用いて特徴的なモジュールとそれらの因果関係を説明する成分の抽出を行った。(Figure 1)

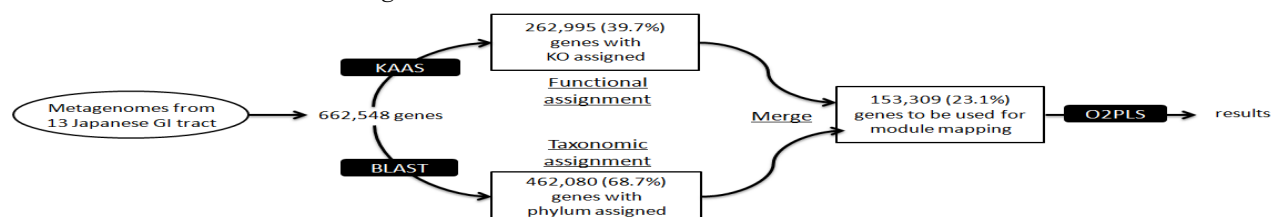


Figure 1. 解析の流れ

結果

今回調べた 13 名の日本人腸内サンプルにおいてモジュールを区別せずにヒット遺伝子全体で統計を取って傾向を見ると、*Bacteroidetes* 由来の遺伝子が大人や子供では 15.25%～55.39%の割合で占められているのに対し、乳児ではその割合が 0.18%～1.71%とわずかであった。また 4 人の乳児では母乳か人口乳かによる違いが顕著であり、前者のみで育った乳児(In-B、In-E)は *Actinobacteria* 由来の遺伝子が 64.00～73.74%、後者のみで育った乳児(F1-U)は *Gammaproteobacteria* 由来の遺伝子が 85.00%と特定の生物門グループによる偏りが大きい、両方で育った乳児(In-M)はそれぞれ中間的な割合を示した。乳酸菌が属する *Firmicutes* 由来の遺伝子はどのサンプルにおいても比較的大きな割合(12.19～60.09%)で存在していることが示された。(Figure 2 左図)

次に O2PLS によって検出された特徴的な違いを持つモジュールについて調べた。(Figure 2 右図) 糖伝達輸送体(M00207)とペプチド・ニッケル伝達輸送体(M00239)は他のモジュールと違い、大人や子供でも *Bacteroidetes* 由来の遺伝子がほとんど含まれていなかった。またその区別化に寄与している成分として、*Alphaproteobacteria*, *Chloroflex*, *Deinococcus*, *Dictyoglomi*, *Fusobacteria*, *Thermobaculum*, *Thermotogae* といったマイナー生物門グループクラスターの存在が確認された。解糖系モジュール(M00001)、クエン酸回路モジュール(M00173)、リボゾーム複合体(M00178/M00179)、アミノアシル tRNA 合成モジュール(M00359/M360)も高い有意性で検出されたモジュールであり、これらに関しては *Actinobacteria*、*Bacteroidetes*、*Betaproteobacteria*、*Deltaproteobacteria*、*Gammaproteobacteria* といったメジャーな生物門グループの組成と大人・子供と乳児間における違いが寄与していることが O2PLS から示された。

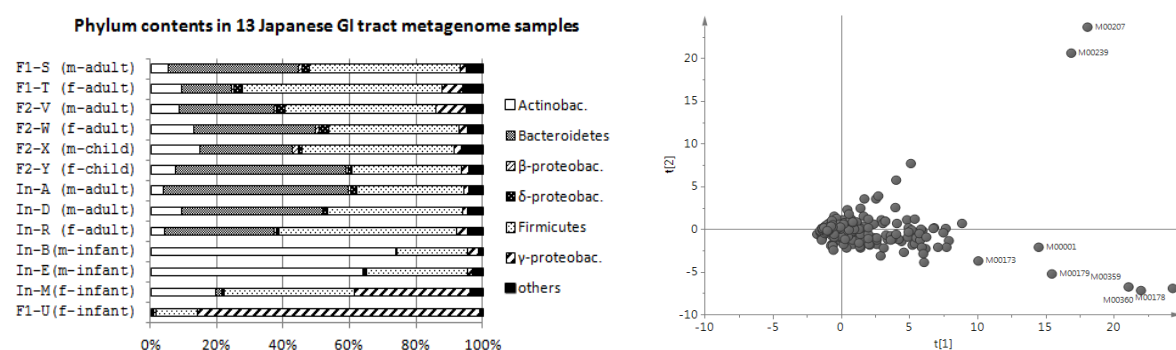


Figure 2. 生物門グループ別遺伝子数の割合(左図)と O2PLS による特徴的モジュールの抽出

- 参考文献
- [1] Tanabe M *et al.*, Using the KEGG database resource., *Curr Protoc Bioinformatics*. 2012 Jun;Chapter 1:Unit1.12
 - [2] Kurokawa K *et al.*, Comparative metagenomics revealed commonly enriched gene sets in human gut microbiomes., *DNA Res.* 2007 Aug 31;14(4)
 - [3] Moriya Y *et al.*, KAAS: an automatic genome annotation and pathway reconstruction server., *Nucleic Acids Res.* 2007 Jul;35(Web Server issue)
 - [4] Kirwan GM *et al.*, Building multivariate systems biology models., *Anal Chem.* 2012 Aug 21;84(16)