

ゲノム、メタゲノムの機能評価システムの開発とその応用

Development and application of functional evaluation system in genome and metagenome

化学研究所バイオインフォマティクスセンター化学生命科学 五斗進

背景と目的

海洋や土壌を中心とした環境や腸内などの生体環境に生息する微生物叢の実態を把握するため、メタゲノム解析がその最も有効な手段として用いられている。当初は 16S rRNA 遺伝子を指標とした種の多様性解析が主であったが、種の多様性の変化だけでは、実際の環境動態の把握にはつながらないため、種の多様性から機能の多様性解析へと研究の方向性をシフトさせなければならない。そこで本研究では、微生物叢が有する機能ポテンシャルを代謝反応や輸送、情報伝達を担う機能モジュール単位で詳細に評価する方法論を開発し、実際の環境メタゲノム解析に役立てることを目的としてシステム開発を進めている。開発した方法論をもとに実際の解析パイプラインを自動化した「MAPLE」システムを平成 25 年末に公開し、約 1 年間で 1,500 を超えるジョブリクエストがあった。本年度は、メタゲノム解析から得られる結果の解釈をさらに容易にするために、モジュールが部分的に充足された際のモジュールの存在意義を示唆する significance value、モジュールが 100%充足された場合の冗長度を示す module abundance を自動的に計算、表示する機能を MAPLE に追加することを主目的とした。

検討内容

Significance value の計算方法は、以下の 3 点に留意して設計した。1. 各モジュールを構成するオーソログ ID (K 番号) が他のモジュールでどの程度共有されているか。2. K 番号は KAAS によって付与されるが、K 番号が付与されない配列については BLAST の E-value も考慮する。3. K 番号が付与された配列の abundance を考慮する。

Module abundance については、まず各モジュールにマッピングされた配列のヒット数を各オーソログの平均長で割ることで補正を行う。この K 番号が付与された配列の abundance をもとに各モジュールの abundance を計算する。モジュールは通常複数の K 番号からなる複数の反応ステップで構成されているので、abundance が最も少ない反応ステップの値をモジュール全体の abundance として定義した。

結果と考察

図 1 に示したように、モジュール充足率 (MCR) の計算結果に加え、メタゲノム中のモジュールの存

在意義を示唆する significance value (Module significance)、遺伝子とモジュール(機能)の abundance を示す KO abundance と module abundance を自動的に計算し表示するようにした。また、この新機能を搭載した MAPLE 2.0.0 を平成 26 年 12 月末にテスト公開した。本年度中にゲノムネットを通じて一般公開する予定である (<http://www.genome.jp/tools/maple/>)。新バージョンでは以前の 10 倍の 100 万配列が解析でき、複数ジョブの結果を一つのジョブの結果として表示する機能も搭載した。また、一旦ダウンロードした結果をアップロードする機能も追加したので、目的に応じたフレキシブルなメタゲノム解析が可能となった。現在、このシステムを用いて、嫌氣的アンモニア酸化や脱窒能に違いのある環境メタゲノム解析を行っているが、複数機能の abundance に違いがあることが分かってきた。今後詳細な解析を行うことで、各機能の abundance の違いと環境中の窒素循環能との関係性が浮き彫りになると期待される。

ID	Type	Name	Components #	KAAS MCR % [Distribution]	KAAS module abundance (normalized) [Distribution]	KAAS module abundance (hit count) [Distribution]	KAAS KO abundance (normalized) [Distribution]	KAAS KO abundance (hit count) [Distribution]	Module significance (phylum) [Distribution]
Energy metabolism									
Carbon fixation									
M00165	Pathway	Reductive pentose phosphate cycle (Calvin cycle)	11	100.0	347	11	347 - 15792	11 - 808	0.969
M00166	Pathway	Reductive pentose phosphate cycle, ribulose-5P =>	4	100.0	347	11	347 - 13927	11 - 482	0.000
M00167	Pathway	Reductive pentose phosphate cycle, glyceraldehyde-	7	100.0	3497	117	503 - 15792	17 - 808	0.000
M00168	Pathway	CAM (Crassulacean acid metabolism), dark	2	100.0	3531	316	3531 - 10893	316 - 347	0.000
M00169	Pathway	CAM (Crassulacean acid metabolism), light	2	100.0	4961	432	4961 - 8255	432 - 534	0.000
M00172	Pathway	C4-dicarboxylic acid cycle, NADP - malic enzyme ty	4	75.0	3531	316	3531 - 8255	316 - 534	0.500
M00171	Pathway	C4-dicarboxylic acid cycle, NAD - malic enzyme typ	7	28.6	3531	316	3531 - 4961	316 - 432	0.999
M00170	Pathway	C4-dicarboxylic acid cycle, phosphoenolpyruvate ca	4	50.0	3531	190	3531 - 3567	190 - 316	0.938

図 1. 新機能を搭載した MAPLE 2.0.0 の解析結果の一例

謝辞

本研究は海洋研究開発機構の高見英人さん、荒井渉さん、三菱総合研究所の谷口丈晃さん、九州工業大学の竹本和広さんとの共同研究です。本研究プロジェクトのリーダーである高見さんを始め、プロジェクトの皆さんに感謝致します。また、本研究は化学研究所共同研究のサポートを受けています。

参考論文

H. Takami, (2014) New Method for comparative functional genomics and metagenomics using KEGG MODULES. *Encyclopedia of Metagenomics* (Nelson K. ed.) Springer, Tokyo Dordrecht, Heidelberg, London, New York, pp. 525-539.