

オントロジーを活用した培地デザインシステムの提案

Proposing an Ontology Based Design System of Microbial Growth Media

川島 秀一*¹
Shuichi Kawashima

岡本 忍*¹
Shinobu Okamoto

*¹ ライフサイエンス統合データベースセンター
Database Center for Life Science

We have been developing Growth Medium Ontology (GMO), which describes nutritional and other components of growth media used for growing microorganisms. We have also created a linked open data of growth media based on the GMO. In this paper, we propose a design system of microbial growth media for microorganisms of which culture methods are unknown such as newly isolated microorganisms by using our LODs and ontologies.

1. はじめに

大学共同利用法人情報システム研究機構 ライフサイエンス統合データベースセンター(以下 DBCLS)は、生命科学分野に蓄積されてきた大量のデータについて、セマンティックウェブ技術を用いて、より効率的にデータを活用するために必要な基盤技術開発や Linked Open Data, オントロジーといったデータの整備を行っている[山口 2010]。また、そうしたデータを、一般の利用者も利用できるように、データベースアプリケーションの開発も行ってきた。一例として、原核生物ゲノムに関連した情報を統合したデータベース TogoGenome*¹ があげられる。TogoGenome は、RDF 化されたデータに対して、ファセット検索やキーワード検索等を行うことができるサービスである。DBCLS では、TogoGenome で利用するために、原核生物に関する様々なデータの RDF 化や、RDF 化に伴って必要となるオントロジーの開発も行ってきた。例えば、原核生物の表現型に関するオントロジーである Microbial Phenotype Ontology (MPO) [川島 2013] や、微生物の培養培地の成分情報等を記述するためのオントロジーである Growth Medium Ontology (GMO) などである [川島 2014]。

研究室で微生物の実験を行う際には、一般には目的の微生物を培養できることが必須である。微生物を培養するためには、その生育に必要な栄養源が含まれる液体、またはそれを寒天などで固めたゲル状固形物である培地が用いられる。微生物が必要とする栄養素は種によって異なるため、様々な成分組成の培地が存在している [Atlas 2010]。研究対象の微生物の培養条件が知られているものについては、既存の情報にあたることで実験を進めることができるが、新規の微生物を分離してきた場合には、既存の情報や研究者の経験などから、適切な培地を類推して試していく必要があり、この過程が困難な場合も少なくない。我々は、これまで、培地に利用される成分についてのオントロジーを開発し、培地の成分組成情報の RDF 化を行ってきた。それぞれの培地には、適応された生物種も記載されている。また、ゲノム配列が決定された微生物種については、MPO を用いて、表現型の情報を RDF 化してきた。これらのリソースを用いることで、新規微生物の培養に利用できる培地を提案する可能性が開かれた。本稿では、そのような観点から、我々のリソースを用いた培養培地のデザインを行うシステムを提案したい。

2. 微生物培養培地オントロジー

これまで継続的に Growth Medium Ontology (GMO) を開発しており、BioPortal*² で公開してきた。2015 年 3 月時点のバージョンでは、970 ほどのクラスが定義されている。また、それを利用して、製品評価技術基盤機構 NBRC で公開されている培地のうち 643 培地、および理化学研究所 JCM で公開されている培地のうち 849 培地、計 1492 の培地の成分組成について RDF 化を行った。

3. 培地デザインシステムの提案

新規微生物の培養に利用できる培地をデザインする際に、当該微生物に関して、どういった情報が利用可能か、ということは重要な要素である。未培養の状況であることを考えると、多くの情報は得られないことが想定される。ただし、最近では DNA シーケンサーの進歩により、環境サンプルに含まれる DNA 配列を直接読むことも可能になってきたことから、目的微生物種のおおよその系統分類情報があらかじめ得られる場合もある。従って、16S rRNA 配列情報から推測できる程度の系統分類情報は利用できるものとして、以下の議論を進める。また、目的微生物を分離した環境の情報、例えば、温度、酸素の有無、pH、塩濃度、栄養源の豊富さなど、も利用できることも想定している。

3.1 培地の成分組成の類似性

進化的に十分近く、類似した環境に生息する微生物種同士が、同じ、または似た成分組成の培地で培養できると仮定することは、一般的である。従って、対象生物と進化的に近い生物種で、培養に成功しているものがあれば、その培地、もしくは成分組成が類似した培地の情報を提案することは、合理的である。培地の類似性に一般的な指標はないが、成分組成の類似性を以下のように提案したい。2つの培地間の成分組成の類似性指標としては、例えば、Jaccard 指標を利用することが考えられる [Levandosky 1971]。ここで注意すべきは、成分の表現の粒度である。培地成分の記述では、例えば、化学化合物だと、一般名で書かれている場合、水和物の違いも記述されている場合、水溶液として記述されている場合等の表記の揺れがある。GMO では、原典の記述をできるだけ表現できるようにこれら全てを異なるクラスとして定義している。しかし、化学化合物の場合、これ

連絡先: 川島秀一, ライフサイエンス統合データベースセンター
〒277-0871 千葉県柏市若柴 178-4-4 東京大学柏の葉キャンパス駅前サテライト 6 階, kws@dbcls.rois.ac.jp

*1 <http://togogenome.org/>

*2 <http://biportal.bioontology.org/>

らは溶液中では成分的な意味では同じであり、その粒度で成分を区別してしまうことは、類似性の観点からは適切ではない。GMO では、化学化合物の水和物や水溶液等は、その化合物の一般的な概念を表す上位クラスと is-a 関係で定義されているので、類似性を計算する場合は、原典に書かれている記述ではなく、一般的なクラスを利用することとした。同じような問題は、化学化合物以外でも存在し、例えばペプトンのような天然物由来の成分でも、単純に Peptone と書かれている場合から、Meat peptone と由来成分が含まれるもの、Bacto peptone のように商品名で記述されているもの等、粒度に違いがある。天然物の場合、類似性指標にどの概念粒度を使うのが適切かは、化学化合物のように自明ではないが、ある程度の抽象化が必要であろう。今回は、培地成分を表す Component クラスと is-a 関係にあるクラスを類似性指標に用いた。ただし、Component クラスと is-a 関係にある Peptone と Extract (肉汁等の天然物からの抽出物)、Water については、それぞれに is-a 関係にあるクラスを類似性指標に用いた。

現在、利用可能な培地について、成分組成の類似性指標としては上記の基準で Jaccard 指標を利用し、Ward 法で階層型クラスタリングを行った。次に、今回採用した類似性指標が適切であったのか調べるために、各クラスターに属する培地で培養される生物種に着目した。それぞれの培地には、培養を行った生物種名が紐付いている。クラスターの特徴を概観するためにその生物種が属する門レベルの分類を用いて、各クラスターに所属する培地が培養可能な生物種が門レベルで何種あるのかを調べた(図 1)。図 1 は、クラスタリングの結果を、100 クラスター、200 クラスター、300 クラスターに分割した結果について、各クラスターに紐付いている門の種類数を横軸に、クラスターの頻度を縦軸にとったグラフである。例えば、横軸の 1 の箇所に数え上げられたクラスター(100 クラスターに分割したときは 7 クラスター、200 クラスターでは 49 クラスター、300 クラスターでは 97 クラスター)は、各クラスターに所属する培地により培養される生物種は全て、特定の一種類の門に所属する生物の培養にしか利用されていないことを意味している。ここから言えることは、より小さなクラスターに分割していくと、そこに所属する培地が培養する生物種の門レベルでの偏りが大きくなるということである。すなわち、進化的な類似性と、今回利用した類似性指標による培地の類似性には、一定の相関があると考えられる。従って、目的生物種に進化的に近い生物種に利用されている培地群を、培地候補として提案することは一定の合理性があるといえる。

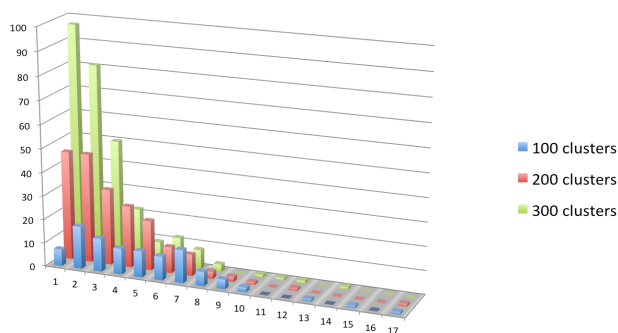


図 1 門レベル培養生物種数と成分組成による培地のクラスターの頻度分布

3.2 微生物生育環境や表現型からの培地の条件予測

培地を調製する際には、栄養としての成分組成も重要であるが、同時に pH、嫌気状態、塩濃度、栄養濃度等、至適生育環境に類似した状態を調整することも不可欠である。TogoGenome では、生物種の既知表現型情報に関して MPO を用いて記述しており、この情報は培地の設計にも利用できる。

一般に培地を調整する際には、培養する生物種の至適 pH を考慮して、それに近い pH になるように調整する。培地 RDF には、pH 情報も記述されているので、目的とする生物の生育 pH が分かっている場合は、各培地が pH の観点から適しているかどうか判定することができる。また、GMO に Buffer(緩衝液)として登録されている成分には、典型的な pH が記述されているので、この情報も利用できる。

次に、嫌気微生物を培養するための嫌気培地については、例えば炭酸ガス培養器の利用や、培地への還元剤の導入など様々な処理を施す必要がある。

微生物の中には、栄養素が多い環境では生育できないものも存在する(貧栄養細菌)ため、培地に加える栄養素の濃度も重要な要素である。特に、ペプトンや抽出液等の成分については、注意を払う必要がある。

4. まとめ

ここまで、(1) 培養対象微生物の系統分類情報が利用できる場合は、できるだけ近縁の生物種に適用される培地を利用する、またそれらの培地に対して、適当な類似性指標を用いた成分組成が類似した培地を利用する (2) さらに分離環境の情報や、表現型の情報も培地の調整に利用する、という観点で、目的微生物のための培地を提案する方法を提案した。これらの方法に必要な情報は、順次、我々により LOD 化されており、計算機的に処理を行うことが可能である。

謝辞

本研究は、DBCLS のメンバーおよび、JST NBDC 統合化推進プログラム「ゲノム・メタゲノム情報統合による微生物 DB の超高度化推進」(代表 黒川 東京工業大学教授)メンバーの方との共同作業の中で行われているものです。また、培養培地の情報は、製品評価技術基盤機構 NBRC および 理化学研究所 JCM が構築されているものを利用して頂いています。

参考文献

- [山口 2010] 山口敦子, 片山俊明: 我が国のデータベース構築・統合戦略(第2回)データベースを統合利用するための基盤としてのセマンティックウェブ技術, 細胞工学 Vol. 30, No. 11, pp. 1210-1215, 学研メディカル秀潤社, 2011.
- [川島 2013] 川島秀一, 岡本忍: 微生物表現型オントロジーおよび LOD の開発, 第 27 回人工知能学会全国大会, 2013.
- [川島 2014] 川島秀一, 岡本忍: 微生物培養培地オントロジーの開発, 第 28 回人工知能学会全国大会, 2014.
- [Atlas 2010] Ronald M. Atlas, Handbook of Microbiological Media Fourth Edition, CRC Press, 2010
- [Levandowsky 1971] Michael Levandowsky, David Winter: Distance between sets, Nature 234 (5): 34-35, 1971