

Title	遺伝学オントロジー
Author(s)	榊屋, 啓志; 溝口, 理一郎
Citation	人工知能学会論文誌, 29(3): 311-327
Issue Date	2014
Type	Journal Article
Text version	publisher
URL	http://hdl.handle.net/10119/12917
Rights	Copyright (C) 2014 人工知能学会. 榊屋啓志, 溝口理一郎, 人工知能学会論文誌, 29(3), 2014, 311-327. http://dx.doi.org/10.1527/tjsai.29.311
Description	

遺伝学オントロジー

The Ontology of the genetics

梶屋 啓志

Hiroshi Masuya

理化学研究所 バイオリソースセンター

RIKEN Bioresource Center

hmasuya@brc.riken.jp, <http://www.brc.riken.go.jp/lab/bpmp/>

溝口理一郎

Riichiro Mizoguchi

北陸先端科学技術大学 サービスサイエンス研究センター

Japanese Advanced Institute of Science and Technology, Research Center for Service Science

mizo@jaist.ac.jp

Keywords: ontology, data integration, biological science, gene, genetics

Summary

Gathering information using a computer has become more and more important as a component of research studies in life science. Recently, ontology provides us with one of the most important means of processing varieties of data and representing knowledge models. Currently, several biomedical ontologies have been constructed with the aim of integrating a variety of information produced by different fields of biology. Therefore, development of a common model/ontology for genes is one of the key issues in bioinformatics studies. However, an ontology fully representing multiple aspects of a gene is still not available. In this study, we dissected the biological roles of a gene and built an ontology that represents a consistent data model of the basic concepts of genetics, including genes, alleles, nucleic acid molecules, locus, genotype and phenotype. This is the first ontology to provide a foundation for the construction of a semantic data model for the concept of gene applicable to broad fields of life science including genetics, molecular biology, and population genetics.

1. はじめに

「遺伝子」は、現代の生命科学の基盤をなす最も重要な概念のひとつである。この概念は、1900年代初頭のメンデルの法則再発見の時代に、「遺伝情報の担体である仮説的な分子」として、遺伝子の種内バリエーションである「アレル (allele)」、染色体上の遺伝子の位置を示す「座 (locus)」などと共に確立されていた。20世紀中盤の一連の研究により、遺伝子の実体はDNAであり、遺伝情報がDNAの4種類の塩基の配列によって担われていることが明らかになった。これをきっかけに、遺伝情報の解読および操作の技術が劇的に発達し、現代ではほとんどの生命科学研究が遺伝子を基盤に行われている。

近年、生命科学で扱う知識やデータ量は急速に増大しており、機械可読な知識ライブラリとして、『オントロジー』の有効性が広く認められるようになった。生命科学における最初で最大のオントロジー、Gene Ontology (GO) は、遺伝子の機能注釈 (アノテーション) に必要な、生命プロセス (biological process)、細胞の構成物 (cellular component)、分子機能 (molecular function) の3つの概念に属する用語を分野横断的に収集し、それらの詳細な分類と相互関係を提供した [GO Consortium 00]。この成功を受けて、生命科学分野では、様々な種類のオントロジーが作成されるようになり、オントロジー間の相互運用も課題となっている。Open Biomedical Ontology (OBO) コンソーシアムのオントロジー開発者の連携グループ

OBO Foundry は、上位オントロジーを基盤とした主要なオントロジーの再構築作業を行っており、将来生命科学を広くカバーする包括的オントロジーが作成されることが期待される。これが実現されれば、データベース統合やテキストマイニング等、様々な用途のためのオントロジーが統一的に管理されるようになり、標準化によって、生命科学における多様なデータの相互運用性の向上、知識処理の有効活用、コスト軽減等に大きく貢献すると考えられる。

OBO Foundry では、生命科学の知識を表現するためには、単に物質的な概念だけでなく、抽象的な概念や、概念同士の関係性を体系立ててモデル化した知識フレームワークを構築することが重要であることも示されている [Smith 07]。ただし、Gene Ontology が遺伝子概念自体は扱っていないことも含め、現状では、遺伝学の基本的な概念を正しく表現できていない問題が残っており、人工知能分野からの知識表現技術導入による問題解決が期待されている。

1.1 既存の生物オントロジーの問題点

ここで、遺伝子概念のオントロジー化の難しさと、包括的オントロジー構築にあたって、現在提供されている生物系オントロジーの抱えている問題点を簡単に述べる。生物学における厳密な「遺伝子」とは、遺伝因子を生物種レベルで捉えた概念を指しており、同じ物を種内バリ

アントとして捉えた場合は、「アレル」と呼ばれる。この分類階層を簡単に示すと、図1の上側ようになる。「遺伝子」→「アレル」→「分子」の順序で各遺伝子の分類を細分化することとなり、「Gdf5 (遺伝子)」→「Gdf5^{Rgsc625} (アレル)」→「Gdf5^{Rgsc625} である1つの DNA 断片」という階層で実装されることになる。一方、遺伝子、アレルという概念は、特定の階層ランクを指すメタクラスと言うべき高次の分類に相当する。また、遺伝子はゲノムの部分領域のひとつであり、例えばアレルは、遺伝子のバリエーションだけでなく、遺伝子以外の非遺伝子領域や、遺伝子の部分領域のバリエーションも指す概念である。よって、それらを簡単に分類ツリーにすると図1のような複雑な多重継承となる。さらに遺伝子は、DNA、RNA の2種類の分子のいずれをも実体としうるので、分子の分類ツリーとは独立に扱うべきという問題もある。

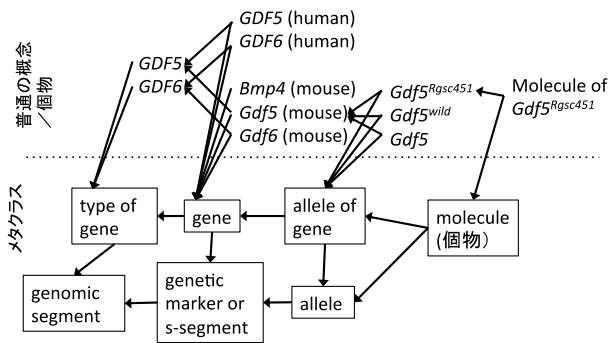


図1 「遺伝子」→「アレル」→「分子」という分類の階層化。遺伝子、アレル等の概念がメタクラス（四角で囲んだもの）となる。矢印は is-a リンクを示す。

現在、OBO コンソーシアムでは、遺伝子概念を含むオントロジーが数個提供されている。しかしそれらは、遺伝子の意味を部分的にとらえたものとして、それぞれ異なる視点で定義をばらばらに提供している状態であり、相互運用性を確立するに至っていない。例えば、「遺伝子」の上位概念を見てみると、1)「物」(Cell Cycle Ontology: CCO) *1、2)「生物学的巨大分子」(Foundational Model of Anatomy Ontology: FMA) *2、3)「alias type」(Protein Affinity Reagents: PAR) *3、4)「配列の特徴としての領域」(Sequence Types and Features: SO) *4 と、全く異なる概念である。

この中で、遺伝子を含めた遺伝学的概念を収録し、概念間のリンクを提供しているのは SO のみである。SO は、ゲノム配列アノテーションの際に各領域を区別するための語彙集として開発された。OBO Foundry では、SO を包括的オントロジーの一部として機能させるため、SO の全概念を、概念には依存するが個別の実体には依存しないで存在することができる抽象的な概念

(BFO の *GenericallyDependentEntity*) であると位置づけた [Smith 07]。SO における遺伝子やアレルは、「sequence feature」の下位概念として個々にインスタンスを持つ別個の概念として定義されている (図2)。

しかしながら、ゲノム配列アノテーション用のオントロジーである SO を、包括的オントロジーの一部として扱うことには、少なからず無理が生じていると言わざるを得ない。ゲノムのアノテーションでは、データとしてもたらされる DNA 配列情報が、実際のゲノム分子の上でどの部分なのかを示すことが目的であり、全てを DNA 配列の上に投影するように設計されている。例えば、図2の gene member region の下位概念に、transcript (転写産物) という概念がある。本来 transcript とは、遺伝情報から産物を作る際に、中間産物として合成される DNA 配列のコピー情報を持つ RNA 分子である。遺伝子と gene member region を結ぶ member_of リンクは、part_of の下位と定義されているので、DNA である遺伝子の部分が RNA 分子であることになり、間違いに見える。しかし、SO の transcript 概念が意図するところは、transcript となるゲノム上の配列を示すことであって、通常生物学者が考える transcript そのものではない。このような問題は、特定用途のために作られたオントロジーを包括的オントロジーとして扱う際に、少なからず誤解を生じる点であると考えられる。

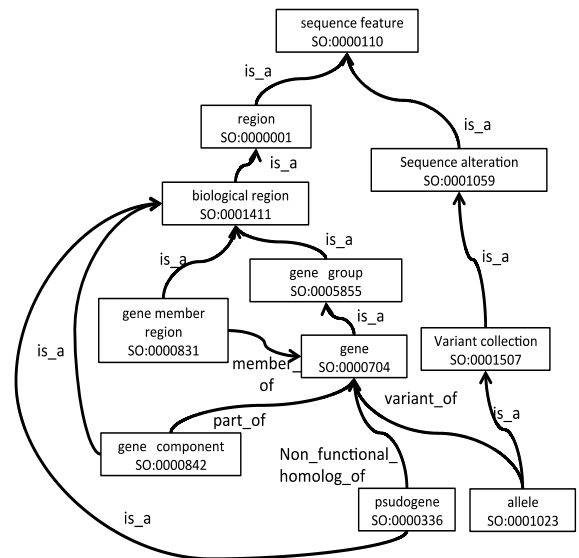


図2 SO で提供されている遺伝子と関連概念の相互関係

OBO Foundry でも、生命科学の包括的オントロジーとして、*GenericallyDependentEntity* である SO の概念を DNA 等の分子に再投影して、SO 語彙が示す分子のオントロジーを作成することの必要性が認識されており、この目的で、配列と同型の分子の分類階層を定義する Sequence Ontology: Molecules (SOM) が提供されている [Mungall 10-2]。しかしながら、SO の階層にある遺伝子

*1 <http://www.semantic-systems-biology.org/apo/>
 *2 <http://sig.biostr.washington.edu/projects/fm/>
 *3 <http://www.psidev.info/psi-par>
 *4 <http://www.sequenceontology.org>

やアレル概念を物の世界に投影した場合には、メタクラス問題や、同じ DNA 分子のインスタンスが、見方によって遺伝子ともアレルとも言える問題、つまり図 2 の *variant_of* リンクが、*is-a* リンクと同等となり、図 1 のような複雑な多重継承が生じる問題が浮上する。そのためか、SOM には、生命科学にとって不可欠なはずの遺伝子やアレルが含まれていない。

このような問題に加えて、そもそも *Generically DependentEntity* のオントロジーとして見た場合にも、SO は、アレルが遺伝子以外の領域のバリエーションでもあることを表現できていないなど、元々遺伝学的な考察が足りておらず、多様な目的に対応するには概念間のリンクが足りないことも問題である。また、遺伝情報とはどんな内容を、分子のシンボルを用いてどのような構造（あるいは形式）でコードしたものなのか、といったことは、SO をはじめ既存のどのオントロジーも提供していない。

生命科学の知識を記述する包括的オントロジーにおいて、遺伝子をはじめとする遺伝学的な概念を表現するためには、分子としての特徴に加えて、それらが生物において果たす役割を表現し、記述する必要がある。そのためには、個体、集団、進化といった、生物学的概念も同時に定義する必要がある。さらに、広範囲の概念を、哲学的、工学的に正しい方法論で（オントロジー工学に則って）構造化（モデル化）する必要がある。

我々は、遺伝情報の担体としての遺伝子(*gene*)やアレル(*allele*)、座 (*locus*)、遺伝型/遺伝子型 (*genotype*)、表現型(*phenotype*)といった遺伝学の主要概念について、存在論的に詳細な考察を行い、法造のロール概念等、国内で工学的に確立された手法を用いて、包括的なオントロジー（＝知識モデル）の構築を試みた。その結果、一貫性を持ち、かつ遺伝子という概念が含む多様な役割（ロール）を、そのコンテキストと共に正しく捉えたオントロジーを作成したので報告する。

2. 遺伝子とは何か: 生命科学からの要請

本章では、生命科学における遺伝子および、いくつかの遺伝学的概念について、生物学の教科書的知識に基づきながら、対象の実在物としての本質は何か、また、どのような特徴を持っているかについて、哲学（存在論）的な視点から分析し、作成すべき遺伝学オントロジーの要件を論じる。

メンデルが遺伝の法則を発見した 1860 年代から、1900 年代にかけての古典遺伝学の時代において、遺伝子は、1) 「生物の遺伝的な形質を規定する単位となる因子（あるいは粒子）」と定義されていた[Gerstein 07]。遺伝子の担う遺伝情報は、それぞれの生物種の身体や生物機能の特徴を形成するとともに、個体差、家系による差などの生物種内のバリエーションを生み出している。特に、集団遺伝学の分野では、交配によって遺伝子交換が可能なグループ（＝種）に属する全個体の遺伝子の総体、「遺伝子

プール」という考えが重要で、プール内の遺伝子のバリエーション（＝アレル）の頻度を論じる。つまり、個々の遺伝子は、種内でアレルのプールを形成するという意味で、生物種の遺伝的なアイデンティティを担っている。

一方、現代では、遺伝子の実体が核酸であり、ゲノム上に遺伝子産物をコードする部分としない部分が存在することが明らかになったことで、遺伝子は、2) ゲノムのセグメントのうち、1 個（以上）の遺伝子産物をコードするものと定義されるようになった。

このような歴史的経緯を背景として、しばしば、遺伝子の定義は変遷していると言われる[Gerstein 07]。しかし、生命科学者は、遺伝子という概念を、メンデル遺伝から分子生物学に到るまでシームレスに使用することに何の困難も感じていない[Griffiths 07]。つまり、古典から分子生物学に到るまで一貫した意味を持つ遺伝子の概念があると考えることが自然である。

2.1 遺伝子は、分子の分類とは独立な概念である

古典遺伝学と分子生物学の両方で、遺伝子の物質的基盤は規定されていない。遺伝子の機能は、DNA、RNA どちらの分子でも担うことができ、核酸が 1 単位の遺伝情報も担わない場合には「遺伝子」とは呼ばれない[Gerstein 07]。つまり、遺伝子という概念の最も基本的な本質は遺伝情報を担うという生物学的役割にあって、分子の種類とは独立である。

2.2 遺伝子の生物学的役割は、2つに分類できる

生命の本質は、自己の維持と複製にある。そして、生命に不可欠な遺伝子にもまた、自己維持の際の役割と自己複製の際の役割の両方がある。自己維持のためには、遺伝子それぞれが独立に RNA 分子にコピーされるが、自己複製の為には、ひとつつながりの DNA 分子（ゲノム）にコピーされる。また、それぞれを稼動する分子プロセスも全く異なっている。

これらの違いをゲノムレベルで見直すと、ゲノムにおけるそれぞれの部分（セグメント）は、自己維持のためにそれぞれ異なる役割を持っており、その役割に従って分類されている（その一つが遺伝子である）。一方、自己複製のためには、全体がひと繋がりコピーされ、その際の誤り（変異）が次の世代に固定されることで、ゲノムの多様性を生み出す。変異が積み重なり生物同士が交配ができなくなると、個体間で遺伝情報の交換ができなくなり、遺伝子プールの分割、つまり、新たな生物種が生まれる。自己複製というコンテキストでは、ゲノムのセグメントは、個々の分子>アレル（種内の多様性のレベル）>遺伝子（生物種のアイデンティティ/種間の多様性のレベル）>オルソログ（生物種横断的なレベル）、というように多様性の方向に分類され、それぞれのレベルで生物学的多様性の役割を担う。

以上をまとめると、遺伝情報担体（ゲノムの部分：ゲ

ノムセグメント) の分類体系は、i) ゲノム領域の機能的役割と、ii) DNA 自己複製により生物学的多様性を生み出す役割との2方向で説明でき、遺伝子は、2つの分類の交点で示される(図3)。この分類は、実体である核酸分子が生物というコンテキスト(個体、種、生物群集といった異なる状況)において果たす役割に応じて分化しており、遺伝学における各概念の意味をよく説明している。

2.3 遺伝子の担う情報の構造

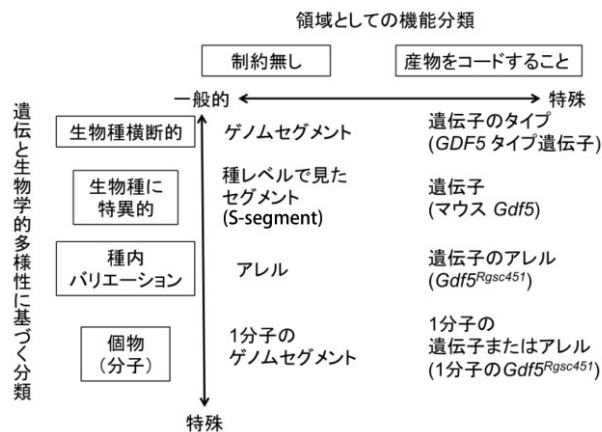


図3 ゲノムセグメントの分類における二方向性

化学物質としての核酸は、情報を担う機能とは関係なく存在している。一方で生命体では、遺伝情報はゲノムセグメント(実体は核酸)にとって本質的な要素となっている。従って、遺伝子の定義を学問分野横断的に行うには、「遺伝情報とは何か」を明確に定義し、それが分子の構造とどのような関係を持ち、結果として生命においてどのような役割を持つかを示すことが極めて重要である。

核酸が遺伝情報を担うことは、しばしば紙に書かれたテキストなど、人工の情報担体とのアナロジーで説明される。文章による工業製品の記述を例にすると、最も基本的なユニットは文字記号であり、その配列によって文章が形成され、製品構造に関する内容をコードしている。核酸の場合は、それ自身が記号列を担い、シンボルを担うメディアとして機能している。シンボル列であるDNA配列は、産物であるタンパクをコードしている。この2つの説明には、実体(紙、分子)、シンボル列(文字列、DNA配列)、記述される内容(製品の構造、タンパクの構造(仕様))といった共通のレイヤーがあり、それぞれの概念は類似している。

現在では、人類が生命から遺伝情報を抽出して参照し、さらにはそれを改変して生命にフィードバックすることも可能になった。このことを表現するには、上記の説明のレイヤー構造のように、シンボル列であるDNA配列を解読し、紙や電子媒体に転写可能であることがモデル

化できるべきである。

一方、同じDNA配列が、異なる化学反応によって、異なる情報をコードしうることも重要である。これを実現するために、同じシンボル列は異なる内容をコードしうることも記述可能でなければならない。これには、情報を記述する形式(シンボル列)と、そのコードする内容の分離が必要になる。

2.4 個々の遺伝子は分子か、『種』か?

生命科学では、遺伝子の個物レベルについて2つの捉え方がある。ひとつは、ここまでの議論の通り、物体として個物は個々の分子であると考えられる捉え方である。もう一つは、遺伝子プールと対応して、例えば「マウス *Gdf5*」という、言ってみれば遺伝子の名前単位の個物を指す。この捉え方では、アレルも同様に、「*Gdf5^{Rgsc451}*」という個物を指す抽象概念となる。例えば、多くのデータベースでは、データベースレコードの個物として遺伝子が存在する。また、SOにおける遺伝子やアレルも、それぞれが個物を持つクラスとして定義されている。生物学では、これら個物の捉え方はどちらも用いられるので、この2つを相互運用できるようにする必要がある。

2.5 遺伝子以外の遺伝学的概念

遺伝子やアレル以外に、遺伝学を構成する主要な概念として、変異(mutation)、形質(trait)、表現型(phenotype)、座(locus)、遺伝子型(genotype)等がある。これらは遺伝子と深く結びついており、遺伝学の説明に不可欠である。下記に、変異、形質、表現型、座、遺伝型について概略を述べる。

変異

「変異」には3つの意味がある。一つ目はDNA配列に変化を生じるイベントとしての「変異」、二つ目は、変異イベントの結果生じた配列、三つ目は、変異配列を持つ核酸分子そのものを「変異」と呼ぶこともある。この3つは密接に関連する一方、明らかに異なる概念であり、明確に区別される必要がある。さらに、変異配列は、同じ配列でも変異であったり、変異でなかったりするという問題がある(例えば、復帰変異: reversionでは、生じた配列は既存の野生型配列である)。このように、変異配列は、変異イベントのコンテキストに明確に依存する概念であり、特定の配列が、イベントの結果として生じたという事実をもって変異と呼ばれることを表現できる必要がある。

形質と表現型

「形質」および「表現型」に関しては、既に他の報告[Gkoutos 05, Smith 07, Mungall 10-1]において示されている通り、個々の生物体が持つ『性質』と定義できる[Masuya 11]。ただし、遺伝学では、形質と表現型を区別して扱う場合があり、形質は性質のタイプ(身体の長さ、体重、

毛色)などを指し、表現型は具体的な値(身体の長さ 10cm、体重 10g、毛色が茶色)を指す。値は、『量』に相当し、性質とは異なる体系で分類される概念を含んでいる。特に値は、数学的な尺度の分類、つまり、名義尺度、順序尺度、間隔尺度、比例尺度等によって分類される [Mizoguchi 10, Tarumi 08, Masuya 11]。

座

「座」は、染色体上の位置としての点や領域を意味している。この「位置」は、相対座標軸上の点や領域と同様に、種内の相同な染色体に対して、共通して用いられる。この概念は、現代の生物学に極めて重要な役割を果たしている。座の概念を用いることで、例えば、『A 遺伝子座に B 遺伝子が挿入されている』といったように、人工的な遺伝子改変をはじめとしてゲノム構成の多様性を記述できる。

遺伝子型

「遺伝子型」(あるいは遺伝型)は、特定の座におけるアレルの組として示される生物体の遺伝的な性質、すなわちゲノムの性質であり、表現型同様に値に相当するレベルで扱われる必要がある。また、性質である遺伝子型と、物であるアレルとの関係を説明する必要がある。

以上より、遺伝子および遺伝学のオントロジーは、下記のような要件を満たすべきであると考えられる。

- 実体である分子の概念と、その生物学的役割に基づく概念である遺伝子が独立に定義できること。例えば、遺伝子という一つの概念の実体に RNA と DNA がありうることを示せること。
- 上に述べた、セグメント分類の2方向性をオントロジーとして表現できること。各方向への特殊化において、属性継承が行われること。さらに、分類の多重性を生み出す生物学的な局面(コンテキスト)を明示すること。
- 遺伝子が持つ本質属性である遺伝情報の実体を記述すること。また、分子実体、分子の配列、そしてコードされる内容を区別できること。
- 2種類の遺伝情報と、上記の多重な分類との関係を明示すること。
- 遺伝子、アレルそれぞれに対して、分子がインスタンスとなる概念と、遺伝子、アレルそのものがインスタンスとなる概念を明示すること。
- 遺伝学における他の主要な概念、変異 (mutation) 形質 (trait)、表現型 (phenotype)、座 (locus)、遺伝子型 (genotype) 等の定義を示し、遺伝子との関係を明示すること。
- 遺伝子産物であるタンパク等へと実現される分子プロセスなどを記述できること。
- 広いドメインとの相互運用性を確保するために、少なくとも1つの上位オントロジーと相互

運用性を持つこと。

3. オントロジー工学に基づく遺伝子および遺伝学的概念の定義

上記の要件に基づいて、我々は、遺伝子および、主要な遺伝学的概念の、オントロジー工学に基づく定義を試みた。本オントロジー構築において、役割、情報、性質の記述が極めて重要になるため、我々は、溝口らの提唱する、ルール (役割)、表現、性質に関するオントロジーの理論[Mizoguchi 04, 05, Kozaki 07]および、これらの理論をまとめた、アリストテレス的な伝統的、常識的存在論を基礎とする記述的な上位オントロジーである Yet Another More Advanced Top-level Ontology (YAMATO)^{*5} [Mizoguchi 10]を参照した。以下に、ルール概念を用いた遺伝子概念のモデル化と、表現オントロジーによる遺伝情報モデル化の概要を示す。

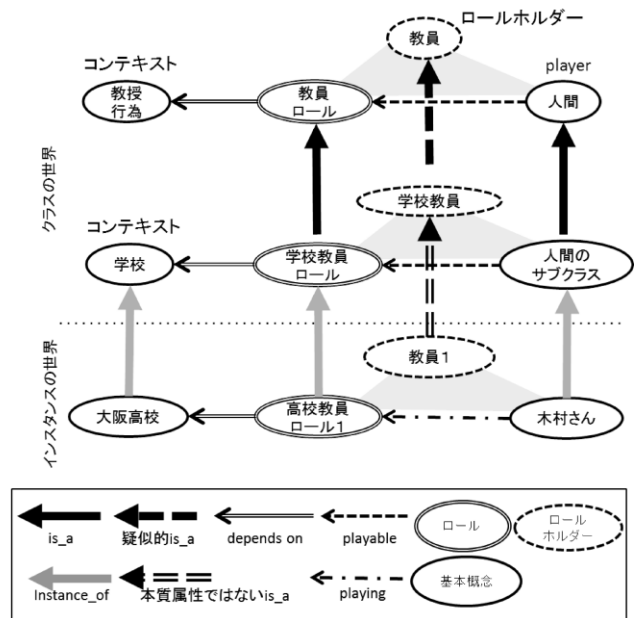


図4 ルールの基本モデル (Mizoguchi 2012 から改変)

3.1 ルール概念による、核酸の生物学的役割の記述

1.1 に述べたように、遺伝子概念の本質は遺伝情報を担うという機能的役割にあり、役割が生じるコンテキストに応じて分化している。従って、遺伝子のオントロジーの構築には役割 (ルール) の体系的な扱いが必要となる。ルールは、教師や妻のように、何かによって play されるものであり、特定の局面 (コンテキスト) に依存した概念である。

溝口らのルール理論[Mizoguchi 04, 05, 12, Kozaki 07]では、基本的な概念として、①コンテキスト、②ルール概念、③ (潜在的) player、④ルールホルダーの4つを定義

*5

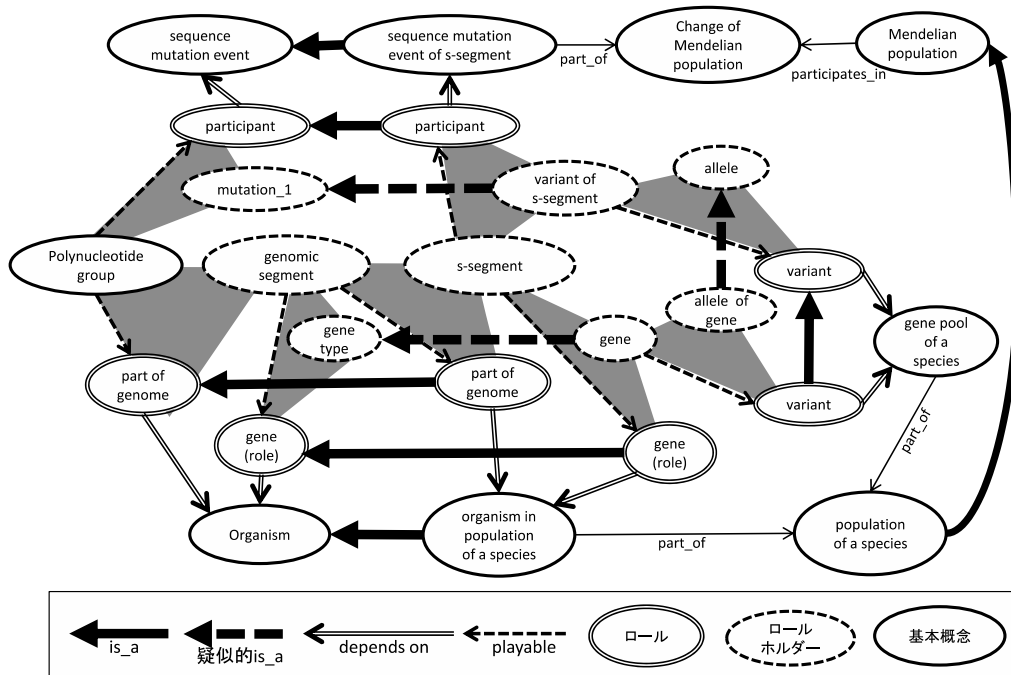


図5 各コンテキストとコンテキスト内で定義される基本概念, ロール, ロールホルダー間の関係性の概要

し区別することによって、コンテキストの記述とそこで生じる概念およびその分類を可能にしている (図4)。例えば、学校において生じる教師という概念は、「学校 (コンテキスト) で、人 (潜在的 プレイヤー) が、教師ロール (ロール概念) を play すると、その結果、教師 (ロールホルダー) となる」と記述される。

このように、いわゆるロールを、ロール概念とロールホルダーに分割することで、ロール概念のインスタンス生成と、ロールの play イベントによるロールホルダーの生成が分離されている。

また、ロールを play する実体であるロールホルダーの分類体系は、ロールを playable な実体の分類を本質としながらコンテキストに依存して生成される。つまり、教師ロールホルダーの実体が人であることが記述でき、さらには、高校 is-a 学校というコンテキストの分類に依存して、「高校教師は教師である」というロールホルダーの分類が可能となる。なお、ロールホルダー間の分類関係は、コンテキストにおいてプレイヤーがロールを play することによって定義され、クラスの下位にインスタンスを直接生成することを許す本質レベルでの is-a とは異なり、まず Player とロールのインスタンスを生成した後に、それらの二つのインスタンスを合体させることによって生成しなければならない。従って本稿では、この分類関係を、「疑似的 is-a」あるいは、「疑似的 instance_of」と呼ぶ。

我々は、この理論を用いて、実体である分子が生物学的コンテキストに依存して役割を持ち、さらに役割が細分化されていくことをモデル化した (図5)。

分子である *Polynucleotide group*^{*6} をクラス制約とする概念が、生物をコンテキストとした時に *Genomic segment* (ゲノム断片) というロールホルダーとなる。*Genomic segment* は、生物種の集団では生物種特有のロールホルダーである *S-segment* (種レベルのゲノム断片) となる。一方、ある *Polynucleotide group* は、変異イベントの結果物として *Mutation* というロールホルダーになる。この変異イベントがメンデル集団でおこる場合には、同じ原理で、*S-segment* が *Variant of S-segment* というロールホルダーとなる。これが種の遺伝子プールのコンテキストで、種内バリエーションである *Allele* というロールホルダーとなる。*Allele* はその種内頻度によって、*Major allele*、*Minor allele* 等の役割に応じたロールホルダーとなる。

さらに、*Genomic Segment* は、それぞれのコンテキストにおいて産物をコードすることで、順次、*Gene type*、*Gene*、*Variant of gene*、*Allele of gene* というロールホルダーとなる。これらの継承関係は、図3に示した2方向性の分類と対応する。

3.2 表現オントロジーによる遺伝子と遺伝情報のモデル化

遺伝情報のコーディングは、分子がシンボルとして機能することで成立している。DNA (あるいは RNA) のモノマーがシンボルを持ち、このシンボルで構成されるシンボル列によって、ヌクレオチド鎖分子は情報を担っている。

*6 遺伝学的には、遺伝子の実体は限定されていないが、現代では、遺伝子の実体は DNA あるいは RNA (厳密にはこれらのポリマー分子の部分: つまり *Polynucleotide group*) 以外は知られていないので、これを基盤に考える事とする。

溝口は、文章、音楽、映像などの内容を持つ記述、そこに埋め込まれる内容、またこれらの情報の担体としてのメディアについて、それらに含まれる概念の詳細な存在論的検討を行い、表現のオントロジーとして構築している[Mizoguchi 04, 05, 10, 12]。この表現オントロジーでは、①準抽象物^{*7}に属する文章、画像、楽譜等の内容を持つ記述 (content-bearing thing) を *Representation* (表現) と呼び、②これらの表現をメディアに転写した情報の担体物を *Represented thing* (表現物) と呼んで区別する。そして、一般的に、表現が③*Representation form* (表現形態) と④*Content* (内容) で構成されることを指摘している。また、表現の本質は、表現の主体が表現を作る時ではなく、それが「伝えられる」時に発揮されると定義されており、遺伝子のような表現主体の定義が困難な場合にも適用可能である。

表現形態と内容は、共に準抽象概念であるが、明確に区別される。例えば、設計書における文章表現の表現形態は文字列、内容は製品の仕様である。ひとつの文字列は別の内容をコードしうる (例: “結構です”)。また、異言語で同じ内容を記述可能なので、それぞれ個別のインスタンスを持つ別の概念である。また、このオントロジーでは、文章の構成要素である文字についても同様の分析を行っている。例えば、英語における文字記号は、アルファベットの“G”が、紙に書かれた“G”形状の図形という③「形式」によって、内容である④シンボルの“G”を「内容」として表現した①文字表現である。そして、この表現が紙やメディアに書かれたものは②表現物である。

この表現オントロジーを参照して、我々は遺伝子を以下のようにモデル化した。まず、グアニンのヌクレオチド基という③「形式」で、抽象的な④内容としてのシンボルである *Symbol designated by nucleotide*、”G”を記述する、①シンボル表現 *Representation of nucleotide symbol* が存在する。これは、上記の文章中の文字と同等の概念である。

このようなシンボルが配列することにより、②表現である *Genetic information entity* (遺伝情報) が形成される。この表現では、上記のシンボルが配列した③形式である *5' to 3' Nucleotide sequence* によって、④内容である *Specification of gene product* (遺伝子産物の仕様) が表現されている。ここでの「仕様」とは、哲学的には *realizable entity* と呼ばれ、設計された実在物が作られる時に、その実現に対して本質的な寄与 (実在物の本質的な特徴の鑄型になる等) をする実在物である。

一方、DNA 分子が自身の複製の鑄型となることは、「コーディング」としては、極めてプリミティブであることに留意する必要がある。ここでは、核酸の配列が何の翻訳プロセスも無しに実現されるため、実現の元である *realizable entity* は、表現形態であるシンボル列そのもの

表 1 遺伝情報と、人工物としての表現の対比

表現物	表現 (表現形式 + 内容)	表現形式	内容	実現された実体
ヌクレオチド基	分子シンボル表現	ヌクレオチド基	分子シンボル	-
紙に書かれた分子シンボル	アルファベットによる分子シンボル表現	図形としてのアルファベット	分子シンボル	-
紙に書かれた文字	アルファベットシンボル表現	図形としてのアルファベット	アルファベットの意味するシンボル	-
DNA内の遺伝子	遺伝情報表現	塩基配列	タンパク構造の仕様	タンパク
紙に書かれた遺伝情報	アルファベットによる遺伝情報表現	塩基配列	タンパク構造の仕様	タンパク (機械で合成された)
紙に書かれた文章	文字列表現	文字列	意味内容	-
仕様書	文字列による仕様表現	文字列	製品仕様	工業製品
紙に書かれた楽譜	楽譜による音系列表現	音符列	仕様としての曲	音系列

が持つと考えるべきである。本研究では、これを *Specification of polynucleotide replication* (ヌクレオチド分子の複製仕様) とした。溝口の表現オントロジーでは、文字列自体が文体などのプリミティブな内容を持つとしており、これと対比できる。

このように、遺伝情報は、文書、楽譜等と同じ概念構造を持つ「表現」として形式化できる。表 1 に、遺伝情報と人工物としての表現について、表現物のコンポーネントである表現、表現形式、内容等の各レベルでの対応を示した。

3.3 抽象的な個物としての遺伝子

『遺伝子の個物とは何か』という問題は、概念定義の根本的な問いである。存在論的には、このような『抽象的な個物』について、2通りの扱いが可能である。一つ目は、『種』という概念である。これは、ものの見方によって、クラスがそれ以上分類できない個物 (=インスタンス) として扱われることに対応したものである。例えば「人種」のインスタンスとしては、「黄色人種」「白色人種」等が挙げられるが、これらの意味はほとんどクラスである。このような場合、オントロジーのクラスをインスタンスに変換する処理が必要となる。Web Ontology Language 2 (OWL2)の punning 機能で作られるクラスの名前や、法造の#オペレーションなどは、クラスから種への変換を実装したものである[Ohta 10]。

二つ目は、*realizable entity* としての『仕様』である。種と仕様は、哲学上は全く異なる概念であるが、クラスと深い関係があるという点では共通している。仕様は、設計された物のクラスに属するインスタンスのあり方を規定しているからである。例えば、遺伝子の実体としての分子が世界に存在する「存在論のプロセス」を考えると、半保存的な複製という設計物の実現という過程を通じて、上位の仕様である遺伝子の仕様を満たしながら、

*7 その存在に時間だけを必要とする存在物。

さらに、下位仕様としての「遺伝子の種内バリエーション」たるアレルの仕様を満たした上で、実世界上に存在するようになる。その意味で仕様は、遺伝子あるいはアレルといった実世界で実現されたモノが属するクラスを強く規定する概念であると言える。また、仕様は、明らかに自己複製を行う遺伝情報担体に本質的であることも重要な点である。

核酸の仕様である *Specification of polynucleotide* は、核酸の持つ遺伝情報: *Genetic information entity* を持っており、仕様の実現の際にそれが実現物であるゲノムセグメントに継承されるとした。また、2.1 節で述べた、ロールに基づくゲノムセグメントの分類によって、ロールホルダーの *Specification of polynucleotide* も同じく分類されると同時に、遺伝子、アレルそれぞれのレベルで、各々インスタンスとなるとした。また、*S-segment* のレベルでの仕様、*Spec_s-segment* は、種の遺伝子プールを規定する本質的属性としてリンクさせた。以上の構造によって、*Spec_s-segment* は、本節で議論している「抽象的な個物としての遺伝子」と同等の概念として扱う事ができる。

3.4 その他の遺伝学的概念の定義

その他、上の節で挙げた *Mutation* (変異) *Trait* (形質)、*Phenotype* (表現型)、*Locus* (座)、*Genotype* (遺伝子型) 等の遺伝学的概念についても、YAMATO に基づいて定義を行った。

上述の通り、変異には、3つの意味がある。まず、イベントとしての *Sequence mutation event* (配列の変化) を、開始および終了イベントをもつ *YAMATO:ordinary event* の下位概念として定義した。このイベントにおける終了イベントでの *participant* (参加物) ロールを演じる核酸分子を、*Mutated polynucleotide group* (分子としての変異) として定義した。さらに、*Mutated polynucleotide group* の持つ表現である *Genetic information entity* の配列を、*Mutated sequence* (配列としての変異) として定義した。*Mutated polynucleotide group* は、メンデル集団や種の集団コンテキストにおいて、*Variant of S-segment*、*Allele* などのロールホルダーのクラス制約となる。

形質と表現型に関しては、既に他の論文で報告している通り[Masuya 11]、性質タイプ (*YAMATO: generic quality*) と値 (*YAMATO: quality value*) をクラス制約として、生物コンテキストにおけるロールホルダーとして定義した。同様に、遺伝型もゲノムのコンテキストで *YAMATO: quality value* が演じるロールホルダーとして定義した。

ゲノムの性質である遺伝子型は、遺伝子産物や生物体の仕様ではなく、生物体やその部品が「どのような性質を持つか」を規定している。このことを示すために、遺伝型は表現型を実現する *realizable entity* である *Specification of phenotype* を持っているとして定義した。*Specification of phenotype* は、産物によって作られた生物

が持つ”性質の”仕様である。この仕様により、遺伝型が表現型を支配することが記述できる。

座に関しては、遺伝子などの任意のセグメントの名称がそのまま座の名称になる場合と、染色体末端からの塩基対の数で示される物理的位置などで示される場合とがある。これもまた、ある特定の性質を示すのに様々な尺度を用いることと同型である。染色体上の位置は、一次元座標上の点あるいは領域であることから、その本質は順序尺度で示すことのできる値であると定義した。

3.5 法造による近似オントロジーの実装

以上の概念定義を、オントロジー構築ツール法造 [Kozaki 01, 07] を用いて実装した。法造では、メタモデリングのために、「種」という概念を、#オペレーションを用いて記述できる。クラス制約を選択する時に、「#」という接頭辞を加える事で、その概念が、クラスではなく、「種」として扱われる。これは、OWL2 の *punning* 機能で作られるクラスの名前と同等である。また、仕様と実現される実体とのリンクである、属性継承機能を持った「*realization-of*」に関しては、現存のいかなるオントロジーエディターでも実装されていない。そこで我々は、法造において定義されている、アイデンティティの継承無しに属性継承を行うリンク、*IS-A* を用いてこのリンクを近似した。また、仕様と下位仕様の間にも、属性継承を伴うリンク「*is-an-upper-specification*」についても、これを実装するオントロジーエディターは存在しない。そこで、*IS-A* リンクを用い、上位仕様インスタンスをクラスとして実装することで、これらを近似した。

下記に、各遺伝学概念を、1)概念定義に必要なコンテキスト、2)物体、3)表現、4)内容、5)表現形式、6)性質タイプ、7)性質値、のレイヤーに分けて、それぞれの定義を簡単に説明する。本オントロジーは、*Genetics Ontology (GXO)* として、法造形式ファイル、OWL 形式ファイル (法造の変換機能により生成) とともに、ウェブサイト*⁸にて公開している。また、法造オントロジービューワのサイト*⁹において、Web ブラウザで閲覧可能である。

§ 基本概念 (コンテキスト)

遺伝学概念は様々なコンテキストに依存して成立している。下記に、コンテキストとなる主要な概念を示す

- **Organism** =_{def} 生物。インスタンスは生物の個体。*YAMATO: functional* のひとつ。
- **Mendelian population** =_{def} 有性生殖を通して結ばれた生物個体の集団。*YAMATO: group* の一種である *Biological population* のひとつ。*Organism* を *member* として持つ。この *Organism* は、ひとつの生物種クラスに限定されるため、*member* とは別に、

*⁸ http://www.brc.riken.jp/lab/bpmp/ontology/ontology_gxo.html

*⁹ http://hozoviewer.ei.sanken.osaka-u.ac.jp/HozoWebXML/?file_name=GeneticsOntology.ont

species ロールを演じる、*#Organism* (名前としてインスタンス化された *Organism*) を属性として持つ。この *#Organism* は、分類階層としての生物学的種そのものであり、*species* ロールを演じる事で、*Biological species* というロールホルダーとなる。

また、*Mendelian population* は、*Gene Pool* を部分として持つ。*Mendelian population* の部分集団は、*Mendelian population* クラスに属する。

- **Population of a species** =_{def} (一つの) 生物種全体としての *Mendelian population*。 *Mendelian population* のサブクラスであり、かつ、部分集団として *Mendelian population* のサブクラス、 *partial population of species* を持つ。

- **Sequence mutation event** =_{def} 核酸配列が変化する生物学的イベントとしての変異。*YAMATO:event* の一種。*YAMATO:instantaneous event* のロールホルダーである *Mutation start event* と *Mutation ending event*、および、*YAMATO:process* のロールホルダーである *Sequence mutation process* を部分として持つ。*Mutation start event* と *Mutation ending event* において、*participant* である *Polynucleotide group* は、同一であるが、それが持つ表現の形式である配列は異なっている。

- **Change of mendelian population** =_{def} *Mendelian Population* が時間とともに変化していくプロセス。*participant* として、*Mendelian population* を含む。また、部分として、*Sequence mutation event* を持つ。

§ 物体

- **Genome** =_{def} *Subcellular component* のひとつ。*Set of polynucleotide molecules* で構成される。

- **Polynucleotide group** =_{def} *YAMATO:functional* のひとつである *Molecular entity* のひとつ。構成要素として、*realizing representation* という役割を演じる *Genetic information entity* を持つ。

- **Mutation_1** =_{def} 変異した結果としての *Polynucleotide group*。 *Polynucleotide group* が、*Mutation ending event* コンテキストにおいて、*participant* ロールを担ってできるロールホルダー。

- **Genomic segment** =_{def} *Molecular entity* である *Polynucleotide group* が、*Organism* コンテキストにおいて、ゲノムの部分という役割を演じてできるロールホルダー。

- **Gene type** =_{def} *Genomic segment* が、*Organism* コンテキストで、遺伝子産物をコードし、生命機能に貢献するという *gene* ロールを演じてできるロールホルダー。表現である *information for self-replication* と *coding of gene product* の2つの遺伝情報を持つ。

- **S-segment** =_{def} *Genomic segment* が、生物個体の集団である *Mendelian population* のサブクラス、*Population of a species* (種の集団) コンテキストにおいて、種のアイデンティティを決める役割を演じるロールホルダー。

- **Gene** =_{def} *Gene type* が *Population of a species* コンテキストにおいて、種のアイデンティティを担うという役割 (*gene* ロール) を演じてできるロールホルダー。

- **Variant of s-segment** =_{def} *S-segment* が、*Change of mendelian population* コンテキストにおける、*Mutation event of s-segment* イベントの *terminal event* において、*participant* ロールが特殊化された、*variant* ロールを演じるロールホルダー。

- **Allele** =_{def} *Variant of s-segment* が *Gene pool of population of species* コンテキストで種内のバリエーションとしての *allele* ロール演じるロールホルダー。

- **Major allele** =_{def} *Allele* が *Population of a species* コンテキストにおいて、最も頻度の高いという役割を演じるロールホルダー。

- **Loss of function allele** =_{def} *Gene allele* が *Organism* コンテキストにおいて、機能欠損であるという役割を演じるロールホルダー。

§ 表現

遺伝情報では2種類の表現がある。核酸のモノマーが表現するシンボル表現と、配列による遺伝子産物の表現である。

- **Representation of nucleotide symbol** =_{def} *YAMATO: representation* のひとつ。*Molecular entity* である *Nucleotide group* を表現形式として、また、*Symbol designated by nucleotide* を内容として持っている。

- **Genetic information entity** =_{def} *YAMATO:representation* のひとつ。表現形式としての *5' to 3' Nucleotide sequence* (現在のところ、核酸配列以外を表現形式とする遺伝子は見つかっていないため、現実に即してこのように限定する) と、内容としての *Specification of gene product* から構成される。

§ 内容

上記の表現が指し示す内容を下記のように定義した。本文に示した通り、「*realization-of*」リンクは、法造オントロジーへの実装においては、アイデンティティの継承無しに属性継承を行うリンク、*IS-A* を用いて近似した。ただし、*IS-A* リンクはロールホルダーに直接つなげないため、ひとつ上位の基本概念につないでいる。

また、属性継承を伴う「*is_an_upper_specification*」についても、*IS-A* リンクを用いて近似した。

- **Symbol designated by nucleotide** =_{def} シンボル YAMATO:symbol の一つ。
- **Specification of gene product** =_{def} 遺伝子産物の分子を実現するための仕様。YAMATO:specification plan のひとつ。YAMATO:molecular entity のひとつである Molecule と、realization-of リンクで繋がっている。
- **Specification of polynucleotide** =_{def} YAMATO:designed proposition である YAMATO:specification のひとつ。生体内で合成される Polynucleotide group である Genomic segment を規定する仕様である。この仕様に依存して実現される Genomic segment とは、”realization-of”関係でつながっている*¹⁰。この概念は、Genetic information entity を持っているが、これは、realization-of 関係によって、Genomic segment に継承される。
- **Specification of polynucleotide replication** =_{def} Specification of polynucleotide が、Replication of polynucleotide (複製) プロセスにおいて、鋳型となる Polynucleotide group の持つ表現が持つ表現形式の style of form ロール (配列自体が持つプリミティブな realizable entity : 3.2 参照) を演じるロールホルダー。
- **Spec_genomic segment** =_{def} Specification of polynucleotide replication が、Genomic segment コンテキストにおいて、Genomic segment の自己複製の仕様という役割を演じるロールホルダー。上位仕様である Specification of polynucleotide とは、is_an_upper_specification というリンクで繋がっている*¹¹。
- **Spec_gene type** =_{def} Specification of polynucleotide が、Gene type コンテキストにおいて、Gene type の自己複製の仕様という役割を演じるロールホルダー。上位仕様である Specification of polynucleotide とは、is_an_upper_specification というリンクで繋がっている。
- **Spec_s-segment** =_{def} Specification of polynucleotide が、s-segment コンテキストにおいて、s-segment の自己複製の仕様という役割を演じるロールホルダー。上位仕様である Specification of polynucleotide とは、is_an_upper_specification というリンクで繋がっている。
- **Spec_gene** =_{def} Spec_s-segment が、Gene コンテキストにおいて、自己複製の仕様という役割

を演じるロールホルダー。Spec_s-segment とは、is_an_upper_specification というリンクで繋がっている。

- **Spec_allele** =_{def} Spec_s-segment が、Allele コンテキストにおいて、自己複製の仕様という役割を演じるロールホルダー。Spec_s-segment とは、is_an_upper_specification というリンクで繋がっている。

- **Specification of polypeptide** =_{def} Specification of gene product のひとつ。生体内で合成されるポリペプチドの一次構造 (配列) を規定する仕様である。この仕様に依存して実現される Polypeptide とは、realization-of リンクで繋がっている。

- **Specification of phenotype** =_{def} YAMATO:specification plan のひとつ。Phenotype を規定する仕様である。この仕様に依存して実現される Phenotype とは、realization-of リンクで繋がっている*¹⁴。

§ 性質タイプ

形質と表現型、遺伝的性質と遺伝子型、座などは dependent entity に分類され、YAMATO では、性質タイプと性質値に分化する。性質タイプとなる概念は以下である。

- **Trait** =_{def} YAMATO: generic quality が、生物体コンテキストにおいて、quality ロールを演じるロールホルダー。値として Phenotype を持っている。
- **Genetic quality** =_{def} YAMATO: generic quality が、生物体コンテキストにおいて、quality ロールを演じるロールホルダー。値として phenotype を持っている。
- **Chromosomal location** =_{def} YAMATO:generic quality である YAMATO: location が、Genomic segment コンテキストにおいて、location ロールを演じるロールホルダー。値として Locus を持っている。

§ 性質値

上記性質タイプに対応する性質値を以下のように定義する。

*¹⁰ realization リンクの近似として用いる IS-A リンクは、基本概念間のみリンクし、ロールホルダーとはリンクできないため、実際には、クラス制約である Polynucleotide group とリンクした。

*¹¹ is_an_upper_specification の近似として用いる IS-A リンクは、基本概念間のみリンクし、ロールホルダーとはリンクできないため、実装においては省略した。

*¹⁴ realization リンクの近似として用いる IS-A リンクは、基本概念間のみリンクし、ロールホルダーとはリンクできないため、実際には、クラス制約である YAMATO:quality value とリンクした。

- **Phenotype** =_{def} *YAMATO: quality value* が、*Trait* コンテキストにおいて、*referring to* ロールを演じてできるロールホルダー。
- **Genotype** =_{def} *YAMATO: Categorical* が、*YAMATO: genetic quality* コンテキストにおいて、*value* ロールを演じてできるロールホルダー。依存する構成要素として、複数の *Allele* を持っている。
- **Locus** =_{def} *YAMATO: quality value* が、*Chromosomal location* コンテキストにおいて、*value* ロールを演じてできるロールホルダー。
- **Proper locus** =_{def} 遺伝子名がそのまま名前となる locus のこと。*Locus* が、値として *YAMATO:categorical* をクラス制約とした場合のロールホルダー
- **Physical length locus** =_{def} 染色体上の物理位置。長さ値である base pair 等が locus 名として用いられる。*Locus* が、値として *YAMATO:quantity* をクラス制約とした場合のロールホルダー

溝口のロール理論では、player はロールを演じることでロールホルダーとなるため、ロールホルダーと player であるクラス制約の概念との間には、コンテキストの下でのみ成り立つ継承関係が成立する。この継承関係を *IS-A* で示すと、『ロールホルダー *IS-A* プレイヤー』という関係が成立する [Ohta 11]。法造では、この理論に基づいて推論を行い、遺伝子やアレルの分類階層を示す事ができる (図 6)。その階層は、図 1 に示したものと全く同型である。かつ、この階層は、上述の 1) ゲノム領域の役割と、2) 生物学的多様性を生み出す役割、つまり、図 3 における横軸と縦軸の双方に対応している。1) の分類はコンテキストに依存していないが、2) の分類は生物学的な種のコンテキストに依存している。

4. 考察

生命科学では、既存知識の再利用のための人工知能技術の開発が大きな課題であり、その用途は、文献からのテキストマイニング、大規模測定データの解釈、データベースの統合と様々である。これらの用途において利用される包括的オントロジーは、教科書の知識を忠実に表現しており、利用方法とメンテナンス手法が一元化されながらも、解釈系であるアプリケーションによって多様な用途に対応することが理想である。SO の例からも示唆される通り、特定の用途に向けて作成されたオントロジーは、しばしば、その用途に特化した概念定義の歪曲を生むため、近年では、生命科学分野でも存在論の原則に基づき、生命科学および遺伝学の教科書概念的な概念を純粋にモデル化することの重要性が指摘されている [Smith 07]。

OBO Foundry では、上位オントロジー-BFO の下で各分野の用途別オントロジーを開発することで、既存オント

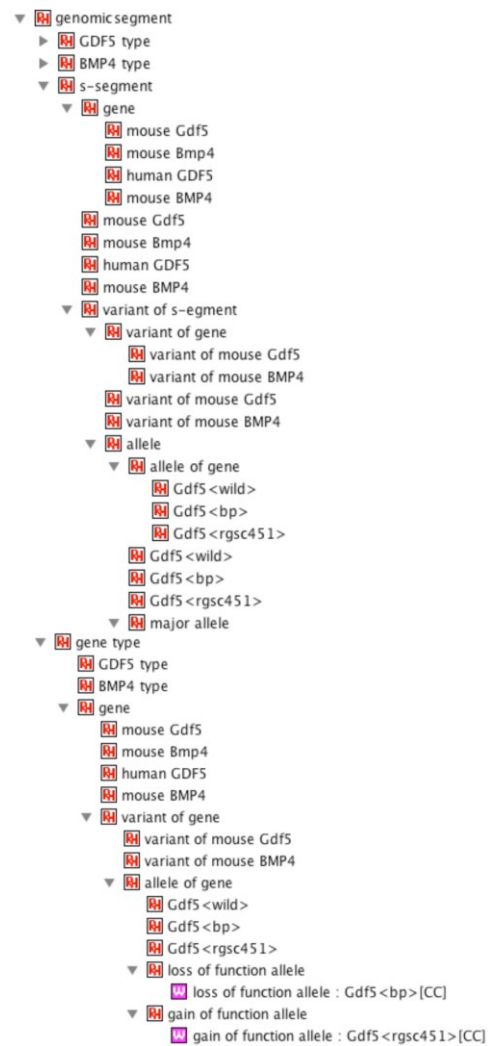


図 6 法造の推論機能で作成される、ロールホルダーの階層 (部分)

ロジーの知識フレームワークとしての問題を指摘し、各分野からボトムアップなかたちで同意を得ていくことにより、混乱の少ないゆるやかな情報統合と、オントロジー構築方法論に対する科学コミュニティの成熟を促すことを指向している。

これに対して、我々はオントロジー工学分野の最先端の技術を用いて、遺伝学に基づく生命科学の包括的オントロジーの作成にトップダウンで取り組んだ。我々が上位オントロジーとして採用した *YAMATO*、およびオントロジー構築環境である法造は、生命科学の分野では、一般的には用いられるには至っていない。しかし、あえて、法造のロール理論と情報のオントロジーを活用しつつ、法造を含めた一般的なオントロジーエディターでは実装されていない概念間のリンク (仕様と実現物のリンクである *realization-of* や、仕様間のリンクである *is_an_upper_specification* 等) をも必要に応じて用いる等、オントロジー工学の方法論を駆使することで、一見困難であり、ともすれば不可能だと言われる包括的オントロ

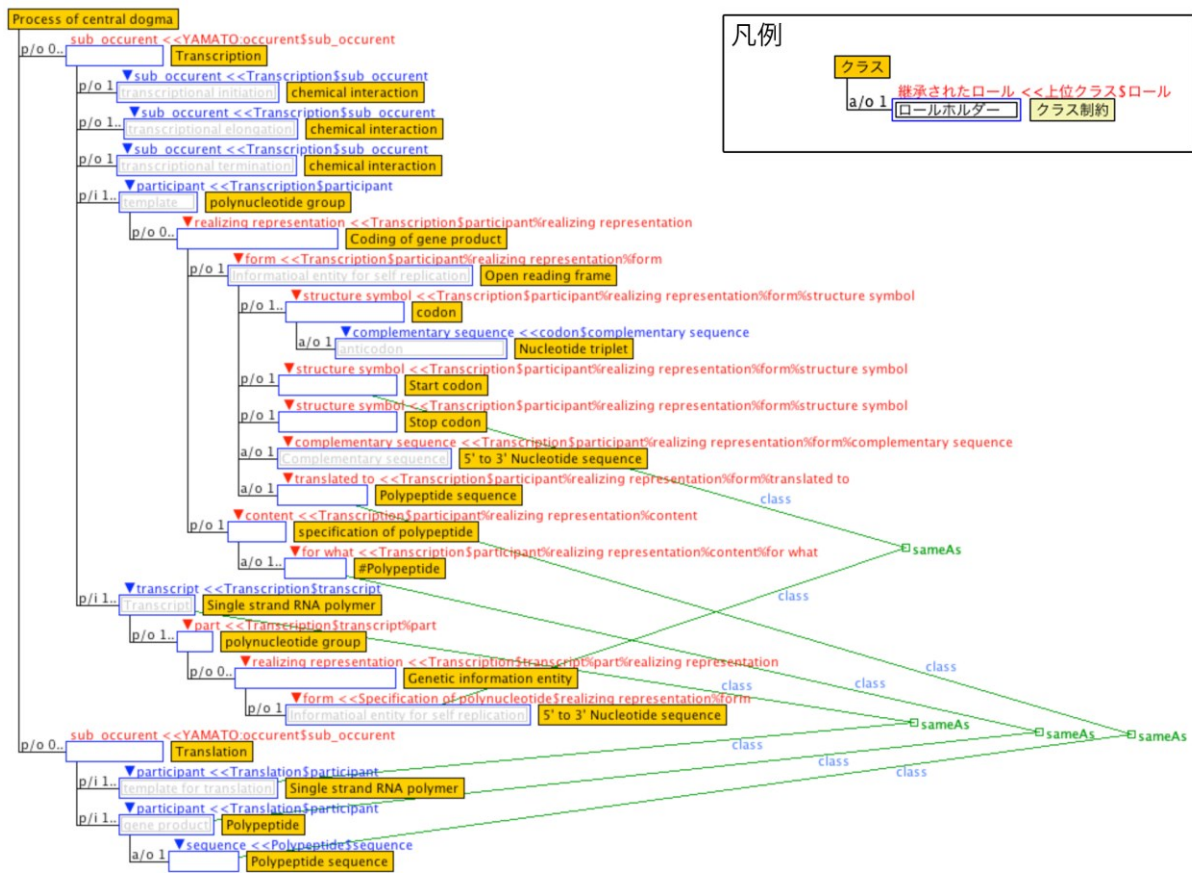


図7 本オントロジーを基盤として、法造エディターで記述したセントラルドグマプロセス（クラス）のモデル。法造クラス記述の凡例を右上に示した。

ジーの作成を目指した。

その結果、多重継承を多用しない等のオントロジー工学の原則を満足しつつ、古典遺伝学、分子遺伝学、集団遺伝学の異なる分野において、遺伝子の持つ意味の各側面を的確に表現し、主要な概念の相互関係を示すことができた。また、遺伝情報のメディア、記述形式、内容を区別して、遺伝子にコードされていることを包括的に示すことができた。

Gerstein らは、遺伝子の概念が「遺伝情報の単位」(1800年代後半～1900年代初頭)、「染色体上の座を占める情報担体」(1910年周辺)、「分子の実体」(1940年周辺)、「転写されるコーディング領域」(1960年周辺)、「Open Reading Frame をもつ配列」(1980年周辺)と移り変わったとしている[Gerstein 2007]。本オントロジーは上記の内容をひとつの概念で網羅するとともに、本質を捉えており、後述する将来の変化も含め、遺伝子の定義を体系的にカバーすることも可能だと考えられる。

4.1 包括的オントロジーにおけるロール概念の有用性

本研究では、遺伝子概念について、図3に示す自己の維持と複製が複雑に分化した役割を丁寧に追跡すること

で、図5に示すような生物学的役割の体系をモデル化した。ここでは、特定のコンテキストにおいて定義されたロールホルダーが、さらに他のコンテキストにおいて player となってロールホルダーを生成するような複雑な継承関係がある。この継承関係は、法造の推論機能によって、図1のような多重継承の階層として出力できる。これは、GXO が推論処理の障害をばらむ多重継承によって構成されること意味しているのではなく、従来の技術では多重継承でしか表現し得なかった分類の階層を、それぞれの分類がどのようなコンテキストで生じるかを明示的に体系化しており、かつ、用途に応じてその体系を単純化によって出力できることを示している。すなわち、GXO は、SOM が抱えるような物体レベルでゲノムセグメントのクラス階層が多重継承になってしまう問題に対して、強力な解決方法を示している。

この特徴によって、GXO は、古典遺伝学、分子遺伝学、集団遺伝学等、広い分野の遺伝子を記述できる。例えば、古典遺伝学の時代には、遺伝子の実体は確定しておらず、表現型とその遺伝様式を観察することで存在が示されたが、GXO では、Genomic segment のクラス制約を特定しないで、個々の遺伝子の名前を記述することで対応できる。一方、分子遺伝学では、Genomic segment のクラス制

オントロジー	全概念数	1対1で対応する概念数	少なくとも上位概念のレベルでは対応すると考えられる概念数	対応に検討を要する概念
Gene Ontology (GO)	40522	2	28957	molecular function (注)
Chemical Entities of Biological Interest (ChEBI)	43964	8	42790	role(注)
Phenotypic Quality (PATO)	2458	1 (最上位の概念)	2458 (本文4.4参照)	
Mammalian Phenotype (MP)	10105	1 (最上位の概念)	10105	
Cell Type (CL)	2178	1 (最上位の概念)	2178	
Common Anatomy Reference Ontology (CARO)	59	1 (最上位の概念)	59	
Foundational Model of Anatomy Ontology (FMA)	75148	1 (最上位の概念)	75148	
Mouse Adult Gross Anatomy (MA)	3029	1 (最上位の概念)	3029	
Uber anatomy ontology (UBERON)	12216	1 (最上位の概念)	12216	
Plant Ontology (PO)	1704	2 (最上位の概念)	1704	

注:原理的にYAMATOにマッピング可能と考えられるが、詳細な検討が必要と考えられる

表2 本オントロジーとマッピング可能な生物オントロジーの例

約を入力することで分子実体が記述できる。さらに、集団遺伝学における、遺伝子プールのアレル頻度等の記述も可能である。さらに GXO では、それぞれの記述において、種レベルの記述をする場合は遺伝子、種内バリエーションを記述する場合はアレル、遺伝情報の変化と捉えた場合には変異と呼ぶことなどを一貫した体系で説明できる。

このような体系は、YAMATO で理論化されたロール理論を具現化したツールである法造[Kozaki 07, Mizoguchi 04, 05, 10, 12]を用いることで実装可能となった。包括的オントロジーでは、一般的かつ本質的な概念定義と、特化した用途とを両立させることが必要な要件となる。本研究では、ロール概念を用いることで、一つの包括的オントロジーの中で多様なコンテキストにおいて様々な役割を持つ遺伝学概念を、一貫性を持って定義する事ができた。

生命科学では、化学、物理学等、他の学問での成果を導入し、生命のコンテキストにおいてどのような役割を演じるかを論じることが多い。そのことが、遺伝学概念の定義において、複雑に入り組んだロールの継承関係が必要となる一要因と考えられる。本研究において、生命科学のニーズに高度に応えることができたことで、他の分野においても、相互運用性の高い包括的なオントロジー構築に、ロール概念が大きく貢献することが期待される。

4.2 表現オントロジーを基盤とした遺伝情報のモデル化

ロール概念の導入による概念定義に加えて、本研究では、溝口の表現オントロジーを基盤にして、遺伝情報のオントロジー化に取り組んだ。ここでは、SO を基盤とした同様の取り組みとの相違点、優位点を論じる。

Hoehndorf らは、述語論理の定理の組み合わせによって、分子そのものの配列である *Molecular sequence*、電子メディア上での配列としての *Syntactic sequence*、抽象的な配列パターンである *Abstract sequence* の3つの

「primitive term」と呼ばれる基本概念の相互関係を体系化した。彼らは、この定理システムは主たる上位オントロジーと互換性があるとしている。これら3つの primitive term は、我々の採用する上位オントロジー YAMATO における、それぞれ、「分子をメディアとする表現物としての分子」、「電子ファイルをメディアとする電子世界の表現物」、「表現形式としての配列」に相当する。この基本分類に基づいて、彼らは、配列を扱うための詳細なメレオロジー、公理化することによって提供している[Hoehndorf 09]。

しかし、彼らの構築したモデルは、表現が持つ「内容」、つまり、realizable entity としての遺伝子産物の仕様を扱っていない。従って、同じ形式(例:同じ文字列)による記述を入力しても解釈系によって異なる出力が可能であることや、異なる記述形式によって同じ内容を表現可能であることは原理的に扱う事ができない。生物界では、ひとつの DNA 分子上に、複数の遺伝子産物がコードされていることがしばしばある。また逆に、真核生物では、コーティング領域がイントロンで分断されており、ひとつの遺伝子産物をコードするのに複数の形式(DNA 配列)が考えられる。これらの現象は、彼らのモデルでは扱うことは不可能である。この問題を解決するためには、YAMATO の表現オントロジーのように、「表現形式としての配列」と、「内容としての遺伝子産物の仕様」とを明確に区別しながら、その両方を実装することが不可欠である。また、*Molecular sequence* は、SO の階層を持ち込む限り SOM と同様の多重継承問題をはらんでおり、これには上述のロールによるモデル化が必要になることを付け加えたい。

4.3 『セントラルドグマ』の記述

GXO は、ロール概念と表現オントロジーの理論を用いることで、柔軟で強力な表現力を持っている。その一例として、遺伝情報が実現される根本原理「セントラルドグマ」の記述について説明する。セントラルドグマは、遺伝情報が読み出され、翻訳されて、遺伝子産物として

のポリヌクレオチドを合成する一連の生命プロセスである。このプロセスでは、DNA 分子の持つ DNA 配列情報が転写プロセスにより RNA に伝達され、最終的に翻訳プロセスによりタンパク質へと変換される。

図 6 に、法造を用いて記述されたセントラルドグマの概念図を示す。大まかには、全体プロセスとしてのセントラルドグマがあり、部分プロセスとして、転写、翻訳プロセスを持っている。転写プロセスでは、鋳型としての *Polynucleotide group* の持つ配列が、転写産物である *mRNA* へとコピーされ、翻訳プロセスではその *mRNA* を鋳型とした最終産物として、*Polypeptide* が仕様通りに実現されること、全体として *Polynucleotide group* から *mRNA*、*Polypeptide* へと情報が伝達されることが記述できた。

GXO では、上述のように、図 3 に示したような遺伝子関連の概念の階層を備えており、例えば、転写プロセスの鋳型である *Polynucleotide group* のクラス制約に、コンテキストに応じて遺伝子やアレルを代入できる。従って、物体レベルで定義したプロセスを用いて、分子生物学、集団遺伝学など、様々な用途に適した記述を行うことが可能である。

これに対して、OBO のオントロジーでも、BFO のフレームワーク上で、GO の *biological process*、*Chemical Entities of Biological Interest (ChEBI)* および、SOM の分子概念を用いて類似の記述を行うことができる。しかしながら、転写から翻訳への情報の流れは、SO におけるゲノム配列間の関係としてのみ定義されている。ここでは、遺伝子、転写産物、ポリペプチド配列等が定義されているが、これは文字通りの意味ではなく、「・・・という特徴を持つゲノム配列」という意味であり、転写、翻訳等のプロセスは、配列概念間の、*transcribed_from*、*ribosomal_translation_of* というリンクで表現されている。よって、この関係を GO の *biological process* や、そこに参加 (*participate_in*) する分子、そして、それらが持つ情報にマッピングすることが必要であるが、そのようなフレームワークは示されていない。

また、表現のオントロジーを持たないため、遺伝情報や配列を *informational entity* (BFO の *Generically DependentEntity* の下位概念) と定義することはできても、GXO のようなシンボルから配列が形成されるような詳細な記述はできない。さらに、表現、表現形式、内容の区別が全く無いために、DNA 複製との同時記載など、同じ配列から異なる産物が合成されるような場合、さらには、*realization-of* を用いた情報から実体の実現を記述しようとした際に記述対象とする概念に齟齬を生じ、結果として記述内容に矛盾が生じることが考えられる。

4.4 データベース統合の基盤としての期待

以前、我々は、OWL1.0 に準拠した、GXO のプリミティブなバージョン“GXO-Lite”を用いた生命科学分野のデ

ータベース統合を試みた[Masuya 10]。各データベースのテーブル、レコード、カラムを、OWL のクラス、インスタンス、プロパティに変換し、包括的オントロジーのフレームワークの下で、クラスやプロパティの継承関係を定義した。これは、多様なデータベースのテーブル、レコード、カラムを、オントロジーでアノテーションすることに相当する。そうすることによって、各データベースのレコード、カラムが何を指しているかを明示し、データ間のリンクを体系化することができた。このような、内容本位のデータの分類は、統合データベースのメンテナンス等のコストの軽減にも貢献すると考えられる[Masuya 10]。

しかしながら、GXO-Lite ではロール概念を用いておらず、遺伝子の定義に関しては SO と同様に、遺伝子、アレルがそれぞれインスタンスを持つというシンプルな定義を行っているため、現実をモデル化するオントロジーとしては矛盾を抱えており、データベース統合の用途においても、遺伝子とアレル、分子の関係や、化学分野の概念である DNA 等の物質名が、生物学では遺伝子と呼ばれる等の高度なリンクを提供するには至らなかった。

本研究で洗練された GXO は、OWL バージョンも提供しており、同様の手法でデータベース統合基盤に用いることが可能である。ただし、OWL に変換されたロール概念の構造が複雑なため、実用に向けてはロールの解釈/可視化ツールの実装が必須であろう。そのような解釈系と組み合わせる時に、GXO は幅広いデータを相互に結びつける極めて強力な基盤を提供できると期待される。

4.5 既存のオントロジーとの相互運用

本オントロジーが準拠する上位オントロジー YAMATO は、異なる上位オントロジー間の相互運用性をもたらすことを目指して開発されている。現存する上位オントロジーを詳細に考察し、特に属性に関する概念に関しては、各オントロジーが含む概念を 1 つのクラスツリー上で分類するとともに、各概念の相互関係を記述している[Mizoguchi 10]。本節では、YAMATO の特長を生かした、生命科学における他のオントロジーとの相互運用の可能性について論じる。

第一に、YAMATO は、他の上位オントロジー (特に、生命科学で多く用いられる BFO) と直接マッピング可能な概念を多く含んでおり、OBO コンソーシアムをはじめとする生命科学ドメインで提案されているオントロジーのいくつかは、最上位の概念を YAMATO の概念とマッピングすることで、全ての概念をそのまま使用する事ができる (表 2)。

§「表現型」の相互運用

ただし、いくつかの概念については、より高度なマッピングが必要となる。例えば、生命科学分野では

Phenotypic Quality オントロジー (PATO) ^{*15} が用いられている[Gkoutos 05, Mungall 10-1]。表現型が性質の一種であることは広く合意されているが、性質そのもののオントロジーは、上位オントロジー毎に異なっている。PATO が準拠する BFO では、性質のタイプと値の間に明確な区別は無く、それぞれは *Quality* 概念のクラスとインスタンスに相当する[Grenon 04]。つまり、「マウス A の体長 10cm」は、シンプルに「マウス体長」という *Quality* のクラスに属する。

一方、Descriptive Ontology for Linguistic and Cognitive Engineering (DOLCE) ^{*16} や YAMATO では、性質タイプと値とを、それぞれ区別された別の概念として定義している[Gangemi 02]。例えば、同じ長さの体長を持つマウス A と B は、体長クラスに属する固有の体長インスタンスを持つが、それぞれが「10cm」というひとつの長さ値インスタンスを共有していることになる。特に YAMATO では、特定の性質を様々な尺度水準（尺度比例尺度、順序尺度、名義尺度等）[Stevens 46]によって示すフレームワークが定義されており、ひとつの性質インスタンスが定性値や定量値を持ち、値同士の相互関係などが記述できる[Mizoguchi 10]。

溝口は、これらの性質概念の相互関係を明確化し、YAMATO に実装している。BFO における *Quality* は、特性 (YAMATO: *Property*) に対応する。さらに YAMATO: *Property* は、性質タイプ (YAMATO: *Quality type*) と値 (YAMATO: *Quality value*) とで構成されており、各性質概念を相互にマッピングできる[Mizoguchi 10]。我々は、この理論を用いて、PATO の持つ生物表現型の概念を YAMATO にインポートし、定性値/定量値の変換などが行える生物表現型の知識ベースのプロトタイプ作成に成功している[Masuya 11, 13]。

§ Locus 概念

また、locus の上位概念である「位置」も注意が必要である。一般的に位置という語句に対応する概念は、オントロジーでは 2 種類考えられる。1 つは、BFO の次期バージョン^{*17} において概念化されている *Immaterial entity* である。これは、洞窟、穴など、実体の無い実体とでも呼ぶべき物であり、地理上の位置を示すことに用いられる。

一方、YAMATO では、位置を *Immaterial entity* としての位置ではなく、特定の座標における任意の位置、つまり YAMATO: *Dependent entity* の下位概念 (BFO の *Dependent entity* と同義) として定義しており、性質と同様に、様々な尺度で示すことが可能である。この 2 つの位置の概念は、「位置」という語彙の示す (independent な) *Entity*、および *Dependent entity* の両側面を概念化したも

のであり、互いに深く関係している。従って、染色体上の位置である座も、*Immaterial entity*、*Dependent entity* の値のどちらでも記述が可能であることになる。本研究では、YAMATO: *Dependent entity* としての *Locus* を定義したが、理論上 *Immaterial entity* としての定義も可能である。

§ SO との相互運用の課題

SO は、配列に関する多くの語彙を含んでおり、本オントロジーにインポートする方法論を確立することが強く望まれるが、実際には、我々が区別している、表現形式、表現、内容、性質、その他様々な概念が混入していると見られ、BFO との相互運用性についても、確立途上の状態である。同様に、YAMATO フレームワークとの連携も困難であると考えられ、まず、SO 内での概念分類をさらに検討することが必要と考えられる。本研究の成果は、そのための指標の一部を提供している。

4.6 今後の課題

以下に、本オントロジーの課題を考察する。本研究では、「機能単位」として、ひとつ以上の遺伝子産物をコードすることをもって、遺伝子と、遺伝子以外のゲノム断片を区別するとした。ただし、『機能単位』の曖昧さ由来する問題が残されている。現在では、多くの遺伝子が、スプライシングやプロモーターの違いによって複数の「転写単位」を持っていることが知られているが、本研究はそれらひとつひとつを遺伝子として定義できる可能性を示している。このこと自体は、ある見地からは正しいと言えるが、多数意見ではないであろう。複数の転写単位を 1 つにまとめて遺伝子というためには、それらを 1 つの塊として解釈する「機能単位」という観点が必要である。本研究では、コンピュータ解釈可能な機能単位の定義を示すには至っていない。しかしながら、生命科学分野においてこの問題自体が極めて曖昧であるので、現状では、オントロジーへの実装としては十分と考えられる。

近年、従来考えられていたよりも、遺伝子の構造が多様であることが分かってきた。例えば、ポリペプチドをコードしない機能 RNA が非常に多く存在すること。また、転写領域の大規模解析によって、転写単位は従来考えられていたものより長大なものが存在し、複数の遺伝子を跨ぐものも多くあることも判明した[Carninci 05]。さらには、脊椎動物の *Sonic hedgehog* 遺伝子などのように、数 Mb 離れた他の遺伝子内に調節領域を複数持っており、長大な領域に跨がって存在する例も知られてきた[Sagai 09]。このような例では、染色体の高次構造をコントロールして遺伝子発現調節を行うことも示唆されており、染色体を跨がって遺伝子産物をコードする遺伝子が存在する可能性がある。このように、従来の分子遺伝学の常識では考えられないような遺伝子の様式が知られつつあるが、それが遺伝子と呼ばれる限り、古典遺伝学における遺伝子の定義を基盤とすることは間違いなく、GXO の一

^{*15} http://obofoundry.org/wiki/index.php/PATO:Main_Page

^{*16} <http://www.loa.istc.cnr.it/DOLCE.html>

^{*17} <http://code.google.com/p/bfo/>

部を拡張するだけで対応可能と考えられる。例えば、染色体を跨いだ遺伝子に関しては、ゲノム断片のあり方(1つの繋がった領域か、あるいは複数の分割された領域か)という問題に帰着する。この場合、遺伝子ロールのクラス制約となるゲノムセグメントのカーディナリティを変更するだけで対応ができる。このように、新たな発見によって定義の変化が起こる場合にも、本オントロジーは小規模な改変で知識フレームワークとして機能し続けることができる。

5. おわりに

本研究は、複雑な知識を存在論的に解析することにより、オントロジーとして構築可能であることを示した。特に生命科学分野に対しては、遺伝学概念の知識表現、そして知識表現の方法論の基盤として重要な礎石となると期待される。また、本オントロジーは、ドメイン外の知識との統合が容易であり、セマンティック Web などの大規模な情報統合に対しても有効である。今回、遺伝学で特に基本的な少数の概念を定義したが、今後さらなる概念を実装し、遺伝学と生命科学の世界をさらに詳細に記述できるような表現力を強化することで、ゲノムを対象とした知識情報処理に対して幅広く貢献できると期待される。

◇ 参考文献 ◇

- [Carninci 05] Carninci P., Kasukawa T., Katayama S. et al. The transcriptional landscape of the mammalian genome. *Science*. 309, pp.1559-1563 (2005).
- [Gangemi 02] Gangemi, A., Guarino, N., Masolo, C., Oltramari, A., Schneider, L.: Sweetening Ontologies with DOLCE, Knowledge Engineering and Knowledge Management: Ontologies and the Semantic Web, 13th International Conference. pp.166-181. (2002)
- [Gerstein 07] Gerstein MB, Bruce C, Rozowsky JS, Zheng D, Du J, Korb J, Emanuelsson O, Zhang ZD, Weissman S, Snyder M. What is a gene, post-ENCODE? History and updated definition. *Genome Res*. 17, pp.669-681 (2007).
- [Gkoutos 05] Gkoutos GV, Green EC, Mallon AM, Hancock JM, Davidson D: Using ontologies to describe mouse phenotypes, *Genome Biology*, 6, R8 (2005).
- [GO Consortium 00] Gene Ontology Consortium: Gene Ontology: tool for the unification of biology, *Nature Genetics*, 25, pp.25-29 (2000).
- [Grenon 04] Grenon P., Smith B.: SNAP and SPAN: towards dynamic spatial ontology. *Spatial. Cognition and Computation*. 4, pp.69-103. (2004)
- [Griffiths 07] Griffiths, Paul E. and Karola Stotz "Gene", in Michael Ruse and David Hull (eds.), *Cambridge Companion to Philosophy of Biology*, Cambridge: Cambridge University Press, pp.85-102 (2007).
- [Hoehndorf 09] Hoehndorf R, Kelso J, Herre H: The ontology of biological sequences. *BMC Bioinformatics* 18, No.10, pp.377, (2009).
- [Kozaki 01] Kozaki, K., Kitamura Y., Ikeda M. and Mizoguchi, R. An Ontology Editor in Hozo - Treatment of "Role" and "Relationship", *Proc. of the Pacific Asian Conference on Intelligent Systems 2001 (PAIS 2001)*, pp. 256-264 (2001).
- [Kozaki 07] K. Kozaki, E. Sunagawa, Y. Kitamura and R. Mizoguchi: Role Representation Model Using OWL and SWRL, *Proc. of 2nd Workshop on Roles and Relationships in Object Oriented Programming, Multiagent Systems, and Ontologies*, Berlin, July 30-31, pp.39-46, (2007)
- [Masuya 10] Masuya H., Makita Y., Kobayashi N., Nishikata K., Yoshida Y., Mochizuki Y., Doi K., Takatsuki T., Waki K., Tanaka N., Ishii M., Matsushima A., Takahashi S., Hijikata A., Kozaki K., Furuichi T., Kawaji H., Wakana S., Nakamura Y., Yoshiki A., Murata T., Fukami-Kobayashi K., Mohan S., Ohara O., Hayashizaki Y., Mizoguchi R., Obata Y., Toyoda T., The RIKEN integrated database of mammals, *Nucleic Acids Reseach*. 39, pp.D861-D870 (2010)
- [Masuya 11] Masuya H., Gkoutos G.V., Tanaka N, Waki K, Okuda Y, Kushida T., Kobayashi N, Doi K, Kozaki K, Hoehndorf R., Wakana S, Toyoda T., and Mizoguchi R.: An Advanced Strategy for Integration of Biological Measurement Data, *Proceedings of 2nd International Conference on Biomedical Ontology (ICBO2011)*, pp.79-86 (2011)
- [Masuya 13] 榎屋 啓志, 古崎 晃司, 大江 和彦, 溝口 理一郎, コンテキストに依存した定性値を扱う生物表現型統合データベースの試作, *人工知能学会第 27 回全国大会*, 3I1-2, (2013)
- [Mizoguchi 04] Mizoguchi, R. Tutorial on ontological engineering - Part 3: Advanced course of ontological engineering, *New Generation Computing*, OhmSha & Springer, 22, No.2, pp.198-220. (2004)
- [Mizoguchi 05] 溝口理一郎: オントロジー工学, オーム社 (2005)
- [Mizoguchi 10] Mizoguchi, R., YAMATO: Yet Another More Advanced Top-level Ontology, *Proceedings of the Sixth Australasian Ontology Workshop Adelaide*, pp.1-16 (2010)
- [Mizoguchi 12] 溝口理一郎: オントロジー工学の理論と実践, オーム社 (2012)
- [Mungall 10-1] Mungall CJ, Gkoutos GV, Smith CL, Haendel MA, Lewis SE, Ashburner M: Integrating phenotype ontologies across multiple species, *Genome Biology*, 11, R2 (2010).
- [Mungall 10-2] Mungall, C. J. Batchelor C. Eilbeck K.: Evolution of the Sequence Ontology terms and relationships, *Journal of Biomedical Informatics* 44, pp.87-93, (2010)
- [Ohta 11] 太田 衛, 古崎 晃司, 溝口 理一郎: 実践的なオントロジー開発に向けたオントロジー構築・利用環境「法造」の拡張 - 理論編 - 人工知能学会論文誌, Vol.26 No.2, pp.387-402, (2011)
- [Ohta 10] Ohta, M., Kozaki, K. and Mizoguchi, R. An Extension of an Environment for Building/Using Ontologies "HOZO" Toward Practical Ontology Engineering, *Proc. of the 10th IASTED International Conference Software Engineering and Applications (SEA2010)*, pp.725-066 (2010).
- [Sagai 09] Sagai T, Amano T, Tamura M, Mizushima Y, Sumiyama K, Shiroishi T. A cluster of three long-range enhancers directs regional Shh expression in the epithelial linings. *Development*. 136, No.10, pp.1665-1674. (2009).
- [Smith 07] Smith B, Ashburner M, Rosse C, Bard J, Bug W, Ceusters W, Goldberg LJ, Eilbeck K, Ireland A, Mungall CJ, et al.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration, *Nature Biotechnology*, 25, pp.1251-1255, (2007)

[Stevens 46] Stevens, S.S. On the Theory of Scales of Measurement. *Science* 103, pp.677-680. (1946)

[Tarumi 08] 垂見, 古崎, 來村, 溝口: 知識構造化システムにおける機能と性質に関するオントロジー的考察 人工知能学会第 22 回全国大会 ,3F1-1,(2008)

[担当委員: 武田 英明]

2013 年 7 月 4 日 受理

著者紹介



榎屋 啓志 (正会員)

1993 年東北大学大学院理学研究科博士前期課程生物学専攻修了。1996 年総合研究大学院大学生命科学研究科博士後期課程遺伝学専攻修了。日本学術振興会特別研究員を経て、1999 年理化学研究所ゲノム科学総合研究センター研究員、2008 年

より同バイオリソースセンター・マウス表現型知識化研究開発ユニット・ユニットリーダー。マウス発生遺伝学的実験解析を経て表現型データベースの開発と標準化に従事。2008 年日本遺伝学会奨励賞、同学会用語編集委員。日本遺伝学会、日本実験動物学会、日本分子生物学会、各会員。



溝口 理一郎 (正会員)

1977 年大阪大学大学院基礎工学研究科博士課程修了。大阪電気通信大学工学部講師、大阪大学産業科学研究所助手、助教授を経て 1990 年より教授。2012 年より、北陸先端科学技術大学院大学教授。工学博士。知識処理全般、特に知的学習支援システム、

オントロジー工学の研究に従事。1985 年 Pattern Recognition Society 論文賞, 1988 年電子情報通信学会論文賞, 1996 年本学会創立 10 周年記念論文賞, 1999 年 ICCE99, 2006 年 ICCE2006 Best Paper Award, 2005 年大川出版賞, 2006. 2013 年本学会論文賞, 2010 年教育システム情報学会論文賞受賞。本学会理事, 同編集委員会委員長, 同会長, Semantic Web Science Assoc. Vice-President 等を歴任。