1I2-R-4-2

# 遺伝子のオントロジー
## An Ontological modeling of gene

桝屋 啓志[*1]
Hiroshi Masuya

溝口 理一郎[*2]
Riichiro Mizoguchi

[*1] 理研バイオリソースセンター
RIKEN Bioresource Center

[*2] 大阪大学産業科学研究所
The Institute of Scientific and Industrial Research (ISIR)
Osaka University

In life science, gene is essential concept meaning molecular matter of genetic information to realize a body of organism and biological functions, which often termed as “blueprint of life”. However, an ontology fully representing multiple aspects of a gene is still not provided. In this study, we dissected biological and ontological definition of bearers of genetic information, including gene and alleles. Based on the analysis, we propose the basic way of modeling of an ontology represents common definition in classical and molecular genetics.

## 1. はじめに

「遺伝子(gene)」は、生命科学の基盤をなす最も重要な概念のひとつであり、1900 年代の遺伝学の黎明期に確立された。この後、20 世紀中盤の一連の研究により、遺伝子の分子的実体は核酸であり、遺伝情報が核酸の4種類の塩基の配列によって担われる事が明らかになった。現代では、生物体の構造や機能が、遺伝情報に基づいて「実現」されるしくみが、分子的実体に基づいて詳しく解明されるようになり、コンピュータを通じて、遺伝情報を生命科学コミュニティ全体で共有したことで、生物学は大きく発展した。

近年の生命科学では、大量に蓄積された情報を知識として体系的に扱う必要性が高まっている。そのためにオントロジーの重要性が認められるようになり、Open Biomedical Ontology (OBO) コンソーシアムでは、多くのオントロジーが作成されるようになった。

しかしながら、OBO のオントロジーでは、遺伝子という概念をデータモデルとして正しく表現できていない。各オントロジーにおける「遺伝子」の上位概念を見てみると、1)物(Cell Cycle Ontology: CCO)、2)生物学的巨大分子 (Foundational Model of Anatomy Ontology: FMA)、3)alias type (Proteome Binders: PAR)、4)配列の特徴としての領域(Sequence on-tology: SO)とばらばらである。それぞれは遺伝子の部分的な意味を示すと言えるかもしれないが、生命科学で一般に受け入れられている『遺伝子』の意味を導くには不十分である。

我々は、このような問題を解決するために、遺伝情報の担体たる遺伝子や、そのバリアントとして定義されるアレルという概念について、存在論的に詳細に考察を行い、包括的かつ一貫したオントロジー的モデルの構築を試みた。

## 2. 遺伝子という概念の分析

### 2.1 生物学における遺伝子概念

古典的な遺伝学では、遺伝子は「生物の遺伝的な形質を規定する因子たる粒子」と定義されていた。遺伝子の持つ遺伝情報は、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も形成すると考えられていた。遺伝学では、厳密な意味での「遺伝子」とは、遺伝因子を生物種レベルで捉えた場合の概念を指している。これに対して、遺伝子の種内バリアントは、「アレル」と呼ばれる。

一方、分子生物学では、遺伝子とは、ゲノム領域のひとつであることが判明し、ゲノム領域には、遺伝子とそうでない領域とがあることが判明した。つまり、ゲノム領域はその機能に応じて遺伝子、非遺伝子領域、エクソン、イントロン、プロモータ、エンハンサー等に分類され、遺伝子とは、遺伝情報の 1 単位として、1つの遺伝子産物をフルにコードするものと定義される。

このような歴史的に定義の変遷に関わらず、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解析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ない[Griffiths 2007]。それゆえ、古典から分子生物学に到るまで、一貫した遺伝子の概念があると考えるべきである。そのような遺伝子概念の本質を、次節以降で存在論的に分析する。

### 2.2 遺伝子概念の存在論的分析

遺伝子という概念の本質は、「遺伝情報を担う」という機能的役割にある。遺伝子の概念定義が、その分子実体の発見に先んじていたことは、これを象徴している。

遺伝子の機能的役割は2つに分類可能である。ひとつは、遺伝情報を、親から子へと継承することである。遺伝継承の一連のプロセスにおいて中心的な役割を果たすのが、DNA の「自己複製」と呼ばれるプロセスで、DNA2重鎖のそれぞれが鋳型となり、同じ配列を持つ2重鎖を合成することで、そこに書かれている遺伝情報をも「複製」する。さらに、遺伝継承プロセスのどこかで、変異(情報の変化)が起こる事で、遺伝的なバリエーションが生じ、その結果生じるゲノムセグメントの種内バリアントは、アレルと呼ばれる。遺伝子→アレルという概念の階層は、生物の遺伝的多様性のレベル、すなわち種としてのアイデンティティ、および、種内の個体差に対応している。

もう一つの役割は、生物個体の設計図として、タンパク質の設計をコードし、その合成のタイミングを規定する事である。この機能は、上記の遺伝子発現と、それに続く遺伝情報の翻訳のプロセスによって担われている。これは、分子生物学における、ゲノム領域の機能的な分類と一致する。

連絡先:桝屋啓志, 理研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, 〒305-0074 つくば市高野台 3-1-1, hmasuya@brc.riken.jp

以上をまとめると、遺伝情報担体(ゲノムの部分:セグメント)の分類体系は、1) 自己複製機能に基づく、バリエーション方向への分類と、2) ゲノム領域の機能的分類の2方向の分類があり、遺伝子とは、図1に示すように、2つの分類の交点である。

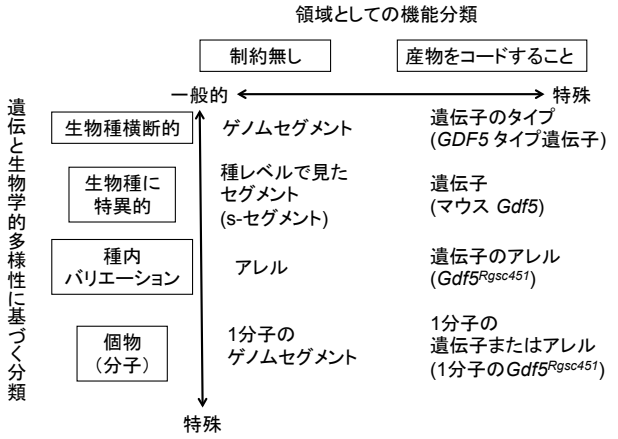



図1 ゲノムセグメントの分類における二方向性

## 3. 遺伝子の役割の詳細

### 3.1 2種類の遺伝情報の共通点と相違点

遺伝子の本質的な属性である遺伝情報は、ヌクレオチド分子の配列によって「記述」されている。特定の形式による情報記述は、「表現」と呼ばれる。遺伝子は、表現を「書き込んだ」具体物の一種と見なすことができ、下記に示すように、メディア(紙など)に、シンボル配列(テキストなど)で情報書き込んだ人工の表現物と相似的に説明することができる。

テキストよる情報の記述において、最も基本的なユニットは記号(シンボル)である。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"形状の図形によって、その抽象的な意味であるシンボルとしての"G"を示している。同様に、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。すなわち、グアニンのヌクレオチド基という「表現形式」で、抽象的な遺伝情報としてのシンボル"G"という「内容」を示している。

このようなシンボルが配列することにより、ヌクレオチド鎖は情報を担うことができる。その概念的な構造は、文書が文字列という表現形式で、仕様や物語などの内容を示すのと同一である。

同じヌクレオチド配列の形式で記述されながらも、自身であるDNA 鎖の配列を記述する自己複製の情報と、ポリペプチドあるいは機能的 RNA の一次配列仕様を記述する遺伝子産物のコーディングとしての情報とは、記述内容も含めた表現と考えた場合に明確に異なる存在である。事実、これらの2つの遺伝情報は、情報の読み出しから終了に至るまで、全く異なる分子プロセスにより実現されている。

以上をまとめると、上記のセグメントの機能分類の二方向性は、遺伝情報に由来する本来的な機能が、生物個体あるいは生物群集といった異なる背景や状況(コンテキスト)を通じてさらに詳細に分化した結果であると考えられる。以下の節ではこれをさらに詳しく分析する。

### 3.2 自己複製の機能的役割の詳細な分類

ゲノムセグメントの遺伝継承と生物学的多様性形成(図1縦軸)における機能的役割は、下記のようにさらに分類できる

(1) 生物種横断的な意味での機能:

ヌクレオチド分子は、生物体(細胞、器官、個体など)のコンテキストにおいて、ゲノムの一部分(ゲノムセグメント)として、「自己複製のための自身の設計情報を担う」という機能的役割を持つ。

(2) 生物種のアイデンティティを担う機能:

ゲノムのセグメントは、遺伝子プール(互いに繁殖可能な個体からなる集団(個体群またはメンデル集団) が持つ遺伝子の総体)の最大範囲たる生物種をコンテキストとして、「生物種特有の遺伝情報のアイデンティティを担う」という、より特殊化された機能的役割を持つ。

(3) 種内の遺伝情報バリエーションを担う機能:

同じく、ゲノムのセグメントは、生物種の遺伝子プールにおいて、種内バリアント(アレル)としての機能的役割を持つ。

### 3.3 ゲノム領域の機能分類

一方、ゲノム領域は、その配列に従って様々な機能を持ち、それによって分類できる(図1横軸)。生物のコンテキストの下で、遺伝子産物をコードするゲノムセグメントのことを遺伝子と呼ぶ。この概念では、上述のような「生物種に特異的」であるかは問題にならない。本論文では、これを「遺伝子タイプ」と呼び、種レベルで見た遺伝子とは区別する。

## 4. 遺伝子のオントロジー的モデリング

以上の分析に基づき、我々は、トップレベルオントロジー、Yet Another More Advanced Top-level Ontology [YAMATO, 溝口 2009]のフレームワークを用いて、遺伝子およびアレルのオントロジー的なモデリングを試みた。

オントロジー構築ツールには、役割(ロール)、役割の担い手(プレイヤー)およびコンテキストを体系的に記述することができる法造を用いた。

法造では、概念は、コンテキストに依存なしに定義される「基本概念」と、コンテキストに依存する「ロール概念」とに大別されている。あるインスタンスがロール概念で定義された役割を担った状態を、「ロールホルダー」と呼び、その役割の担い手となる概念は、基本概念もしくはロールホルダーから選ばれる [古崎 2007]。この基本的な体系を用いて、遺伝子のモデル化を行った。下記にゲノムセグメントに由来する主要な概念の定義について簡単に説明する。以下では、YAMATO で定義されている上位概念を、YAMATO:の接頭辞をつけて表現する。

1. ***Genetic information entity*** $=_{def}$ 表現である ***YAMATO: representation*** の一つ。表現形式としての ***molecular sequence*** と、内容としての ***YAMATO:specification*** から構成される。
2. ***Molecular sequence*** $=_{def}$ 表現形式である ***YAMATO:symbol sequence*** の一つ。***molecular symbol*** で構成される。
3. ***Molecular symbol*** $=_{def}$ シンボル ***YAMATO:symbol*** の一つ。
4. ***Representation of molecular symbol*** $=_{def}$ 表現である ***YAMATO:representation*** の一つ。表現形式としての ***radical group*** と、内容としての ***molecular symbol*** から構成される。
5. ***genomic segment*** $=_{def}$ ***molecular entity*** である ***polynucleotide group*** が、生物体である ***organism*** コンテキストで、ゲノムの部分という役割を演じるロールホルダー
6. ***gene type*** $=_{def}$ ***genomic segment*** が、***organism*** コンテキストで遺伝子としての役割を演じるロールホルダー。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。

7. ***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。
8. ***gene*** =$_{def}$ ***gene type*** が ***population of a species*** コンテキストにおいて、***s-segment*** で規定される役割を演じるロールホルダー。
9. ***variant of s-segment*** =$_{def}$が変異プロセスである ***mutation process in gene pool*** コンテキストにおいて、バリアントという役割演じるロールホルダー。
10. ***allele*** =$_{def}$ ***variant of s-segment*** が ***gene pool of population of species*** コンテキストで演じるロールホルダー。
11. ***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。
12. ***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

本オントロジーは、http://www.brc.riken.jp/lab/bpmp/ontology/ontology_gene.html.よりダウンロード可能である。図2に、法造オントロジーエディター画面 ***population of a species*** 概念と、そのメンバーとしての ***organism***、構成要素としての ***gene pool***、それぞれのコンテキストにおいて定義される、***s-segment***、***gene***、***allele*** 等の概念示す。

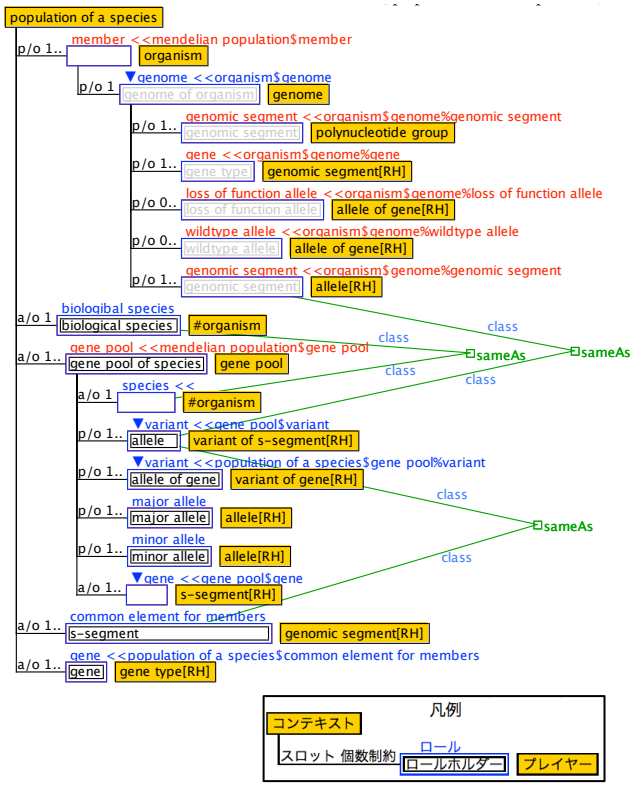
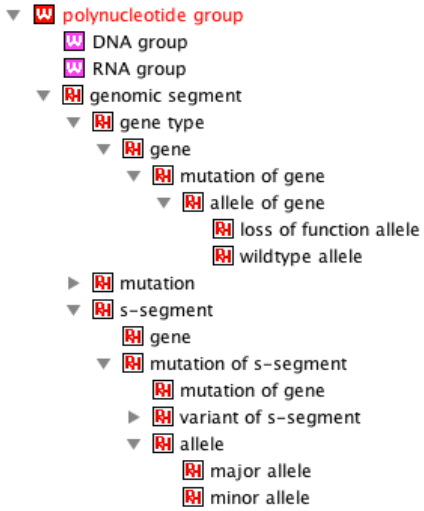



図2 法造による、生物種の集団(***population of a species***)コンテキストと、コンテキスト下で定義される概念のモデル

法造のロールホルダーとプレイヤーとの間には、コンテキストの下でのみ成り立つ継承関係があり、これを「IS_A」と示すと、“ロールホルダー IS_A プレイヤー”の関係が成り立つ[太田 2011]。これに基づいて、遺伝子やアレルの分類関係を推論して示す事ができる。この階層では、ゲノムセグメントの下位に ***s-segment*** があり、さらにその下位に、***gene***、***allele***、***allele of gene***、




そして、機能に基づくアレルの分類、***loss of function allele*** が分類される(図3)。

図3 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)

## 5. まとめ

### 5.1 遺伝子概念の知識モデル化

幅広い分野で共通利用できる汎用的なオントロジーを作成するには、特定の利用法や、分野特異性を捨て、客観的で本質に基づいた分類体系を作成することが望ましいと考えられている[溝口 2004]。その最も有効な方法の一つが、上位オントロジーへの準拠であり、これを用いたデータベース統合も行われている[桝屋 2010]。近年、OBO コンソーシアムにおいてもこのことが重要視されるようになり、OBO Foundry が中心となって、各ドメインオントロジーを、上位オントロジーである Basic Formal Ontology [BFO]の下位オントロジーとして組み入れ可能にするための変更作業を進めている[Smith 2007]。しかしながら、生命科学に必須の概念である遺伝子やアレルについては、シンプルな分類階層によるモデル化が困難である。

本研究では、遺伝子およびアレルの概念を、遺伝学、分子生物学および、オントロジー工学に基づいて詳細に分析し、「遺伝子とは何か」を表現する知識モデル(オントロジー)の作成を目指した。その結果、上位オントロジーYAMATO に基づいて、古典遺伝学、分子生物学、集団遺伝学など、ライフサイエンスの広い分野をカバーしうる一貫した概念モデルの構築を世界で初めて構築することができた。

YAMATO は、BFO や、Descriptive Ontology for Linguistic and Cognitive Engineering [DOLCE]と同様、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。YAMATO の一部である表現オントロジーは、様々な形式によって記述される contents bearing things=「表現」について、概念を構成する「表現形態」、「内容」という普遍的な構造を示している。また、表現を担う具体物(本や電子ファイルなど)は「表現物」と定義されている。YAMATO は、これらの抽象および具体物の分類と、それぞれがどのようなレベルでインスタンスとなるかについて詳細な理論を提供している[溝口 2004, 2009]。

親から子へ(あるいは、あるセグメントから、複製されたセグメントへ)伝えられる遺伝情報は、シンボル列という表現形式と、そのコードする内容(DNA 鎖自身の設計、あるいは他の分子の設

計）の組み合わせでインスタンスとなる。同じシンボル列（表現形式）にコードされていても、DNA 鎖自身の設計と、ポリペプチドの設計は別の表現である。このモデルによって、全てのセグメントが持つ自己複製のための情報と、遺伝子だけが持つ遺伝子産物のコーディングを区別することができた。また、表現物のモデルは、ある遺伝子の複数のインスタンスが、同じ遺伝情報を持つことのモデル化を可能にした。

本研究のもう一つの重要な要素は、ゲノムセグメントが異なるコンテキストにおいて、別のロールを担うことによって生じる、分類の多重性のモデル化である。遺伝情報の生物学的な役割とそのコンテキストを詳しく分析することで、ゲノムセグメントに由来する遺伝子、アレルなどの概念の多重な分類を、オントロジー上での多重継承を用いずに、体系的に整理することができた。これは YAMATO で理論化されたロール理論を具現化したツール、法造を用いることで実現できた[古崎 2007, 溝口 2009]。

## 5.2 関連する研究

冒頭に述べた、生物系オントロジーにおける遺伝子概念モデルの不一致の問題を解決するために、OBO Foundry や関連グループによって、いくつかの努力が行われている。

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」の関係性を体系化した。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[Hoehndorf 2009]。実際、3つの primitive term は、YAMATO および本研究における、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。彼らは、本研究では述べてられていない、配列を扱うための詳細なメレオロジーを、定理によって提供している。

しかし一方で、彼らの定理は、われわれが扱っている，表現が持つ「内容」としての遺伝情報そのものは扱っておらず、下に述べる SO 同様、遺伝子やアレルの多重なクラス階層を体系的に整理することはできない。

また最近、SO は、DNA 配列の、メレオロジカル、空間、時間的な特徴を扱うよう改訂された。さらに、SO の拡張として、配列と同型の分子の分類階層を定義する Sequent Ontology: Molecules (SOM)を提供している[Mungall 2010]。しかしながら、SO 自体の階層分類は、遺伝子やアレルのクラス階層の問題を解決することはできない。例えば、 ＜アレル variant_of 遺伝子＞の関係は、遺伝子からアレルへの属性継承ができないことや、遺伝子でない s-segment（遺伝マーカーなど）のバリアントがアレルとなることに対応できない。多重継承を使わずにこのような問題を解決するには、各ゲノムセグメントの生物学的な役割を明確化することにより、遺伝学のコンセプトに基づいた概念の分類方法を示すことが必須であろう。つまり、本研究で示したオントロジーは、配列本位でのゲノムセグメント分類とは、知識表現や情報統合において、相補的に貢献すると考えられる。

## 5.3 今後の課題

遺伝学は、生命の本質的な理解のために極めて重要な概念を提供する。生命科学情報を整理統合するためには、遺伝子概念を軸として、遺伝学の他の概念、すなわち、染色体上の位置である「locus: 座」、生物個体の遺伝的タイプを示す「genotype: 遺伝型あるいは遺伝子型」、遺伝型に対応する生物体の特性である「phenotype:表現型」等のオントロジー化が必須である。

OBO が提供するオントロジーとの相互運用性も極めて重要な課題である。法造の出力機能では、ロールを含むデータモデルを、OWL と SWRL の形式へと変換することが可能である[古崎 2007]。また、最近のトレンドとして、OBO の中心的なオントロジーは、BFO との親和性を高めつつある。YAMATO では、存在論的な議論に基づいて、複数の上位オントロジーに含まれる各概念間の相互運用性を提供しつつあり、その成果を利用することで、本研究の提案する枠組みをバイオインフォマティクス分野で広く利用可能にしていくことが望まれる。

## 参考文献


[BFO] http://www.ifomis.org/bfo

[DOLCE] http://www.loa.istc.cnr.it/DOLCE.html

[Griffiths 2007] Griffiths, P.E. and Karola Stotz: "Gene", in Michael Ruse and David Hull (eds.), Cambridge Companion to Philosophy of Biology, Cam-bridge: Cambridge University Press pp85-102., (2007).

[Hoehndorf 2009] Hoehndorf R, Kelso J, Herre H: The ontology of biological sequences. BMC Bioinformatics 18;10:pp377, (2009)

Mungall, C. J. Batchelor C. Eilbeck K.: Evolution of the Sequence Ontology terms and relationships , J Biomed Inform. 44, pp87-93, (2010)

[Smith 2007] Smith, B., Ashburner, M., Rosse, C., Bard, J., Bug, W., Ceusters, W., Goldberg, L. J., Eilbeck, K., Ireland, A., Mungall, C. J., The OBI Consortium, Leontis, N., Rocca-Serra, P., Ruttenberg, A., Sansone, S.-A., Scheuermann, R. H., Shah, N., Whetzel, P. L., and Lewis, S.: The OBO Foundry: coordinated evolution of ontologies to support biomedical data integration. Nat Biotechnol, 25, pp1251–1255 (2007).

[YAMATO] http://www.ei.sanken.osaka-u.ac.jp/hozo/onto_library/upperOnto.htm

[太田 2011] 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)

[古崎 2007] Kozaki K., Sunagawa E., Kitamura Y. and Mizoguchi R.: 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)

[桝屋 2010] 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 Res. 39, D861-D870, (2010).

[溝口 2004] Mizoguchi, R.: Tutorial on ontological engineering - Part 3: Advanced course of ontological engineering, New Generation Computing, OhmSha&Springer, Vol.22, No.2, pp.198-220, (2004).

[溝口 2009] Mizoguchi, R.: Yet Another Top-level Ontology: YATO, Proc. of the Second Interdisciplinary Ontology Meeting, pp.91-101, (2009)