

## **Definitions of the concepts in genetics**

(The english version of extracts from Masuya H. and Mizoguchi R. (2014) The Ontology of the genetics, Transactions of the Japanese Society for Artificial Intelligence, 29, 311-327)

### **1. Overview**

As shown in Fig. 1 in Section 2.2 of the main text, the two directions of genomic segment classification each represent a systematic specialization in accordance with roles in organisms. These specializations of roles are based on the function of genomic segments as genetic information carriers, which is their essential quality, and also, these specializations are dependent on specialization of contexts in which the individual roles are manifested. Therefore, as discussed in Section 3.3 of the main text, we have defined concepts by using the role theory and the ontology construction tool Hozo that is capable of systematically describing role concepts based on this theory.

With Hozo, by introducing concepts such as "basic concept" that can be defined without depending on a context, "role concept" which is a role a basic concept plays in a specific context, and, additionally, "role holder" which is an instance bearing a role defined by a role concept, it is possible to describe, within a consistent framework, the fact that the role assumed by a basic concept also changes when a context changes.

In addition, with Hozo, the concept of "species" can be described by using #operation for the purpose of metamodeling. By adding the prefix "#" when selecting a class constraint, that concept is treated as a "species" instead of a class. This is equivalent to the class name created by using the "punning" function in OWL2.

In the following, definitions of concepts in genetics will be briefly described by dividing these concepts into seven levels, namely, 1) context needed to define a

concept, 2) object, 3) representation, 4) content, 5) representation method, 6) quality type, and 7) quality value. In the following, a prefix YAMATO: is used to represent concepts in upper levels defined in YAMATO.

## 2. Definitions of concepts

### 2.1 Basic concepts (contexts)

Concepts in genetics are established depending on various contexts. The major concepts that serve as contexts are indicated below.

- *Organism* =<sub>def</sub> a biological organism whose instances are individuals of the biological organism. It is one of *YAMATO: functional*.
- *Mendelian population* =<sub>def</sub> a population of individuals of a biological organism that are capable of sexual reproduction. It is one of *Biological population*, which is a type of *YAMATO: Group*. Its members are *Organism*. Because this *Organism* is restricted to one species class, it has an attribute *#Organism* (*Organism* serving as an instance in the form of a name) and plays a species role separate from that of member. This *#Organism* exactly corresponds to a biological species as a taxonomic hierarchy, and it serves as a role holder of *Biological species* by playing the *species* role.

In addition, a *Mendelian population* has a *Gene pool* as its component.

Component populations of a *Mendelian population* belong to the *Mendelian population* class.

- *Population of a species* =<sub>def</sub> a *Mendelian population* as the entirety of (a single) species. It is a sub-class of *Mendelian population* and also includes, as component populations, *Partial population of species*, which is a sub-class of *Mendelian population*.

- *Sequence mutation event* =<sub>def</sub> a mutation as a biological event in which a nucleic acid sequence is changed. It is a type of *YAMATO: event*. As its components, *Mutation start event* and *Mutation ending event*, which are role holders of *YAMATO: Instantaneous event*, and *Sequence mutation process*, which is a role holder of *YAMATO: Process*, are included. In *Mutation start event* and *Mutation ending event*, although *Polynucleotide group*, which is a *participant*, is the same, its sequence, which is a method of representation, is different between the two.
- *Change of Mendelian population* =<sub>def</sub> a process through which a *Mendelian population* undergoes changes over time. *Mendelian population* is included as a *participant*. In addition, as its component, *Sequence mutation event* is included.

## 2.2. Objects

- *Genome* =<sub>def</sub> one of *Subcellular component*. It is constituted of a set of *Polynucleotide molecules*.
- *Polynucleotide group* =<sub>def</sub> one of *Molecular entity* which is one of *YAMATO: Functional*. As a constituent element, *Genetic information entity* that plays the role of *realizing representation* is included.
- *Mutation\_1* =<sub>def</sub> a *Polynucleotide group* formed as a result of a mutation. It is a role holder that is created when *Polynucleotide group* assumes a *participant* role in the context of *Mutation ending event*.
- *Genomic segment* =<sub>def</sub> a role holder that is created when a *Polynucleotide group*, which is a molecular entity, plays a role of a genome component in the context of *Organism*.
- *Gene type* =<sub>def</sub> a role holder that is created when a genomic segment plays a gene role, which codes for a gene product and contributes to a biological function, in the

context of *Organism*. It holds two types of genetic information, namely, *Information for self-replication* and *Coding of gene product*, which are representations.

- *S-segment* =<sub>def</sub> a role holder in which a *Genomic segment* plays a role that determines the identity of a species in the *Population of a species* context, which is a sub-class of a *Mendelian population*, that is, a population of individuals of a biological organism.
- *Gene* =<sub>def</sub> a role holder that is created when a *Gene type* plays the role (*gene* role) of bearing the identity of a species in the context of *Population of a species*.
- *Variant of s-segment* =<sub>def</sub> a role holder in which an *S-segment* plays a *variant* role which is a *participant* role specialized in a *Terminal event* in a *Mutation event of s-segment*, in the context of *Change of Mendelian population*.
- *Allele* =<sub>def</sub> a role holder in which *Variant of s-segment* plays a *allele* role as a variant within a species in the context of *Gene pool of population of species*.
- *Major allele* =<sub>def</sub> a role holder in which *Allele* plays a role of having the highest frequency in the context of *Population of a species* context.
- *Loss of function allele* =<sub>def</sub> a role holder in which *Gene allele* plays a role of having lost its function in the context of *Organism*.

### 2.3 Representations

There are two types of representations in genetic information, namely, a symbol representation that is represented by monomers of a nucleic acid and a representation of a gene product based on a sequence.

- *Representation of nucleotide symbol* =<sub>def</sub> one of *YAMATO: representation*. It possesses a *Nucleotide group*, which is a *Molecular entity*, as a representation method and, in addition, a *Symbol designated by nucleotide* as the content.

- *Genetic information entity* =<sub>def</sub> one of *YAMATO: representation*. It is constituted of a *5'-to-3' Nucleotide sequence* as a representation method (it is restricted in this way in accordance with reality because, currently, there is no known gene that is represented by a method other than a nucleic acid sequence) and a *Specification of gene product* as the content.

## 2.4 Contents

Contents indicated by the representations described above are defined below.

As discussed in the main text, when implementing this in a Hozo ontology, a "realization-of" link was approximated by using an *IS-A*, which is a link for inheriting an attribute without inheriting an identity. However, because an *IS-A* link cannot directly be linked to a role holder, it is linked to a basic concept at one level above it.

In addition, an "*is an upper specification*" associated with an attribute inheritance was also approximated by using an *IS-A* link.

- *Symbol designated by nucleotide* =<sub>def</sub> a symbol that is one of *YAMATO: symbol*.
- *Specification of gene product* =<sub>def</sub> a specification for realizing a molecule of a gene product. It is one of *YAMATO: specification plan*. It is connected with a *Molecule*, which is one of *YAMATO: molecular entity*, by a *realization-of* link.
- *Specification of polynucleotide* =<sub>def</sub> one of *YAMATO: Specification*, which is *YAMATO: Designed proposition*. It is a specification that specifies a *Genomic segment*, which is a *Polynucleotide group* synthesized in a living organism. It is linked to a *Genomic segment* that is realized depending on this specification with a

"*realization-of*" relationship<sup>\*1</sup>. This concept includes a *Genetic information entity*, which is inherited to a *Genomic segment* via a *realization-of* relationship.

- *Specification of polynucleotide replication* =<sub>def</sub> a role holder in which *Specification of polynucleotide* plays a *style of form* role (a primitive realizable entity which a sequence itself possesses: see Section 3.2 of the main text), which is a representation method possessed by the representation of a *Polynucleotide group* that serves as a template in the *Replication of polynucleotide* process.

- *Spec\_genomic segment* =<sub>def</sub> a role holder in which *Specification of polynucleotide replication* plays the role of a *style of form* in the context of *Genomic segment*. It is linked with *Specification of polynucleotide*, which is a specification at an upper level, via an *is\_an\_upper\_specification* link<sup>\*\*2</sup>.

- *Spec\_gene type* =<sub>def</sub> a role holder in which *Specification of polynucleotide* plays the role of a *style of form* in the context of *Gene type*. It is linked with *Specification of polynucleotide*, which is a *YAMATO: Specification* at an upper level, via an *is\_an\_upper\_specification* link.

- *Spec\_s-segment* =<sub>def</sub> a role holder in which a *Specification of polynucleotide* plays the role of a *style of form* in the context of *S-segment*. It is linked with *Specification*

---

<sup>\*1</sup> Because an *IS-A* link that is used as an approximation of a *realization-of* link links just between basic concepts and cannot link with a role holder, we actually linked it with a *Polynucleotide group*, which is a class constraint.

<sup>\*\*2</sup> Because an *IS-A* link that is used as an approximation of an *is\_an\_upper\_specification* link links just between basic concepts and cannot link with a role holder, we omitted it from the implementation.

*of polynucleotide*, which is a specification at an upper level, via an *is\_an\_upper\_specification* link.

- *Spec\_gene* =<sub>def</sub> a role holder in which *Spec\_s-segment* plays the role of a style of form in the context of *Gene*. It is linked with *Spec\_s-segment* via an *is\_an\_upper\_specification* link.
- *Spec\_allele* =<sub>def</sub> a role holder in which *spec\_s-segment* plays the role of *style of form* in the context of *Allele*. It is linked with *Spec\_s-segment* via an *is\_an\_upper\_specification* link.
- *Specification of polypeptide* =<sub>def</sub> one of *Specification of gene product*. It is a specification that specifies a primary structure (sequence) of a polypeptide that is synthesized in a living organism. It is linked with a *Polypeptide* that is realized depending on this specification via a *realization-of* link.
- *Specification of phenotype* =<sub>def</sub> one of *YAMATO: specification plan*. It is a specification that specifies *Phenotype*. It is linked with a *Phenotype* that is realized depending on this specification via a *realization-of* link<sup>\*3</sup>.

## 2.5 Quality types

A trait and a phenotype, a genetic quality and a genotype, locus, and so forth are classified as dependent entities, and they are differentiated into types and values of qualities in YAMATO. The following concepts are included in quality types.

---

<sup>\*3</sup> Because an *IS-A* link that is used as an approximation of a *realization-of* link links just between basic concepts and cannot link with a role holder, we actually linked it with *YAMATO: quality value*, which is a class constraint.

- *Trait* =<sub>def</sub> a role holder in which a *YAMATO: generic quality* plays the *quality* role in the context of *Organism*, and possesses a *Phenotype* as a value.
- *Genetic quality* =<sub>def</sub> a role holder in which *YAMATO: generic quality* plays the role of a *quality* in the context of *Organism*, and possesses a genotype as a value.
- *Chromosomal location* =<sub>def</sub> a role holder in which a *YAMATO: location*, that is, a *YAMATO: generic quality*, plays the role of a *location* in the *Genomic segment* context, and possesses a *Locus* as a value.

## 2.6 Quality values

Quality values that correspond to the quality types described above are defined as follows:

- *Phenotype* =<sub>def</sub> a role holder that is created when a *YAMATO: quality value* plays a *referring to* role in the context of *Trait*.
- *Genotype* =<sub>def</sub> a role holder that is created when *YAMATO: categorical* plays the *value* role the context of *Genetic quality*, and includes *Alleles* as a constituent element that it is dependent on.
- *Locus* =<sub>def</sub> a role holder that is created when *YAMATO: quality value* plays the *value* role the context of *Chromosomal location*.
- *Proper locus* =<sub>def</sub> a locus whose name is the same as the name of a gene located at the locus. It is a role holder in the case in which a *Locus* as a value has *YAMATO: categorical* as a class constraint.
- *Physical length locus* =<sub>def</sub> a physical position on a chromosome. A base pair or the like, which is a value of length, is used as the name of a locus. It is a role holder in the case in which a *Locus*, as a value, has *YAMATO: quantity* as a class constraint.