

Improvements to the *Drosophila* anatomy ontology

Marta Costa*, David Osumi-Sutherland, Steven Marygold and Nick Brown
FlyBase, Department of Genetics, University of Cambridge, Downing Street, Cambridge, UK

ABSTRACT

The *Drosophila* anatomy ontology (DAO) defines the broad anatomy of the fruitfly *Drosophila melanogaster*, a genetic model organism. It contains over 8700 classes, with close to half of these corresponding to neuroanatomical terms.

We are systematically reviewing the DAO classes, improving the textual information and classification. This includes adding definitions, comments and synonyms, as well as formal definitions, which results in a full classification in some cases. Classes belonging to each of the defined organ systems are reviewed together to improve consistency of free text and formalisation. So far we have reviewed 7 of the 11 organ system classes, resulting in 83% of classes having a definition.

1 INTRODUCTION

The *Drosophila* anatomy ontology (DAO) (Costa *et al.*, 2013) is an ontology that describes the wild-type anatomy of *Drosophila*, containing over 8700 classes. It is used by FlyBase (Dos Santos *et al.*, 2015), the gene and genomic database for *Drosophila*, for manual curation of phenotypes and expression patterns. Users are also able to query for this type of data, either through FlyBase or Virtual Fly Brain (Milyaev *et al.*, 2012). Having an accurate, encompassing and human-readable ontology is therefore essential to enable curators to choose the correct anatomy term, and for users to easily navigate the data.

When the DAO was first developed over 20 years ago, it did not include textual information or significant formalisation. A large effort has been undertaken in the last 9 years to improve this situation (Costa *et al.*, 2013). This work has resulted in 83% of classes now having a definition and the classification having been greatly improved. The DAO currently contains 46 object properties and over 18,000 subclass axioms, with over 2,500 equivalent class axioms, with around 50% of over 10,000 classifications being inferred.

New DAO classes are curated from the published literature, if enough evidence regarding their morphological characterisation, identity and if appropriate, function, are provided. Every class includes a definition, synonyms and comments each attributed to a source reference.

The neuroanatomy field has grown massively in the last few years thanks to technical advances, enabling researchers to identify and characterize the function of individual neurons and several projects are currently underway to map all neu-

rons in a variety of model organisms. We, in collaboration with the Virtual Fly Brain project, have focused on capturing this information in *Drosophila*. Currently, 46% of the DAO comprises terms that are part of the nervous system, including close to 2300 distinct neuron classes, over 220 neuroblast lineage clones and neuropils.

Here, we present our most recent work in improving the textual information and formalisation patterns of DAO classes.

2 RESULTS

In order to maintain consistency between related terms, we are reviewing existing classes by making use of their current classification into 11 different organ systems, such as the tracheal, muscle, adipose, etc. Around 80% of classes in the DAO had previously been classified as part of an organ system previously, thus making it easy to retrieve a list of classes to review. Work has proceeded class by class, improving both the textual information and formalisation. When necessary, we have sought advice from expert researchers.

The systematic review of classes uncovered several cases of redundancy and duplication, which were resolved by obsoleting one of the terms. For example, the classes Malpighian tubule Type II cell (see section 2.1) and excretory star cell were found to refer to same entity. In this case, the latter was obsoleted, and the name added as a synonym to the former.

We have concluded this review for 7 of the 11 organ systems (muscle, tracheal, reproductive, digestive, circulatory, excretory and adipose), corresponding to 570 classes. Work is ongoing to complete the remaining (muscle, nervous, endocrine and sensory).

2.1 Improving textual information

We have added textual definitions to 83% of DAO classes, an improvement of 10% since October 2013. The definition describes the general classification of the anatomical entity, its properties, and when appropriate, any distinguishing traits. These statements are supported by references, either cited in the text, or listed at the end with a publication identifier (mostly a FlyBase one: prefix FBrf followed by 7 dig-

* To whom correspondence should be addressed: m.costa@gen.cam.ac.uk

its).

Comments are added when appropriate, for one of two reasons. The first is to provide relevant information relating to the experimental setup when, for example, investigating the function of a neuron. The second is to clarify the relationship between competing nomenclatures.

Synonyms from the published literature are added to each class, together with references. The addition of synonyms has particular relevance to anatomy ontologies, for which competing nomenclatures often exist.

An example of the textual information for a class is below:

name: Malpighian tubule Type II cell

definition: "Morphologically distinct cell type found only in the initial, transitional and main segments of the Malpighian tubules interspersed with Type I cells. Type II cells are smaller and flatter than Type I cells, with shorter (main segment) or no (initial region) apical microvilli. Type II cells originate from a subset of caudal visceral mesoderm cells that overlie the tubule primordia as they evert from the hindgut. By stage 15, Type II cells have been incorporated in the tubules and adopt epithelial characteristics. In the mature tubules there are on average 110 Type II cells." [FlyBase:FBBrf0064792, FlyBase:FBBrf0102373, FlyBase:FBBrf0160477, FlyBase:FBBrf0222532]

comment: These cells are involved in primary urine production via the presence of ion channels that allow chloride and water to enter the tubule lumen (O'Donnell et al., 1998).

synonyms: "excretory star cell" EXACT; "Malpighian tubule stellate cell" EXACT [FlyBase:FBBrf0030988]

2.2 Improving formal definitions

In systems such as the tracheal, in which certain structures are repeated in each metameric unit, adding a formal definition to each of these terms significantly increases the robustness of error checking procedures. An example of some of the relationships that are added is below:

name: adult abdominal spiracular branch

intersection_of: FBbt:00003071 ! adult spiracular branch

intersection_of: connected_to FBbt:00003040 ! adult lateral trunk

intersection_of: connected_to FBbt:00004814 ! adult abdominal spiracle

intersection_of: part_of FBbt:00003024 ! adult abdominal segment

In other instances, adding a formal definition allows for full classification of terms. This becomes particularly relevant for neuroanatomy, a field in which new neuron types are being frequently described. Having a formal definition for a class ensures that new terms are correctly classified, provided that enough information is available, such as developmental origin.

An example of a neuron class that can be fully classified based on expression (which identifies this subset of very well studied neurons) and developmental origin is below. This formalisation pattern, or a similar one (excluding only the neuroblast information), was used to define the 104 classes of adult *fruitless* neurons.

name: adult fruitless aDT-b (female) neuron

intersection_of: FBbt:00005106 ! neuron

intersection_of: develops_from FBbt:00050148 ! neuroblast CREa1 (female)

intersection_of: expresses FlyBase:FBgn0004652 ! fruitless

intersection_of: part_of FBbt:00110416 ! adult fruitless aDT-b (female) lineage clone

3 DISCUSSION

We have reviewed terms that belong to 7 of the 11 organ systems in the DAO, improving the textual information essential for casual users, and the formalisation necessary to easily maintain a correct classification and to prevent the introduction of errors. Reviewing related classes as a group helps to maintain consistency in the ontology, both in terms of free text and the formalisation patterns used.

Future work will focus on completing the systematic review of the DAO by revising the classes in the remaining 4 organ systems.

REFERENCES

- Costa, M., Reeve, S., grumbling, G. and Osumi-Sutherland, D. (2013). The *Drosophila* anatomy ontology. *J. Biomedical Semantics*, 32-4.
- Dos Santos G, Schroeder AJ, Goodman JL, Strelets VB, Crosby MA, Thurmond J, Emmert DB, Gelbart WM; the FlyBase Consortium. (2015). FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. *Nucleic Acids Res.* doi: 10.1093/nar/gku1099
- Milyaev, N., Osumi-Sutherland, D., Reeve, S., Burton, N., Baldock, R. A. and Armstrong, J. D. (2012). The Virtual Fly Brain browser and query interface. *Bioinformatics* 28, 411-5.