The Virtual Fly Brain browser and query interface

Nestor Milyaev[1], David Osumi-Sutherland[2], Simon Reeve[2], Nicholas Burton[3], Richard A. Baldock[3] and J. Douglas Armstrong[1,*]

[1]School of Informatics, University of Edinburgh, Edinburgh, UK, [2]Department of Genetics, University of Cambridge, Cambridge, UK and [3]MRC Human Genetics Unit and Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh, UK

Associate Editor: Jonathan Wren

ABSTRACT

Motivation: Sources of neuroscience data in *Drosophila* are diverse and disparate making integrated search and retrieval difficult. A major obstacle to this is the lack of a comprehensive and logically structured anatomical framework and an intuitive interface.

Results: We present an online resource that provides a convenient way to study and query fly brain anatomy, expression and genetic data. We extended the newly developed BrainName nomenclature for the adult fly brain into a logically structured ontology that relates a comprehensive set of published neuron classes to the brain regions they innervate. The Virtual Fly Brain interface allows users to explore the structure of the *Drosophila* brain by browsing 3D images of a brain with subregions displayed as coloured overlays. An integrated query mechanism allows complex searches of underlying anatomy, cells, expression and other data from community databases.

Availability: Virtual Fly Brain is freely available online at www.virtualflybrain.org

Contact: jda@inf.ed.ac.uk

1 INTRODUCTION

With its relatively simple architecture and unparalleled level of genetic tractability, the brain of *Drosophila* is an ideal model for fundamental neuroscience research. However, in practice, navigating the *Drosophila* neurobiology literature and the various community databases is a real challenge. The literature's long history and diversity have resulted in a variety of often-conflicting terminologies. Information about useful research reagents is scattered across a number of databases that are not interoperable and rarely even cross-referenced. Consequently, finding, for example, data about the connectivity between two brain regions, or where genes are expressed or what resources exist is an arduous and difficult task. To realize the value of the scientific investment in *Drosophila* neuroscience, the community requires a robust anatomical framework with the supporting data structures and computational tools to exploit it. The first challenge is to agree on a common framework for gross anatomy. This has been addressed, for the first time by the 'BrainName' consortium, who proposed a revised nomenclature for the insect brain (K.Ito *et al.*, submitted for publication) with each term defined both textually and as a volume in a reference *Drosophila* brain.

Next we need to be able to construct and solve useful searches and integrate data from disparate sources. Using the web ontology language OWL 2 (http://www.w3.org/TR/owl2-primer/), it is possible to store complex relationships between terms and, with the help of reasoning software, to use those relationships to automate classification and drive queries. For example, the relationship between a class of neurons and a structure that it innervates can provide both a substrate for useful queries (what neurons innervate region X) and a criterion for classification (all neurons in class A innervate some region X). An ontology also provides a mechanism for common annotation of neuroanatomical data, allowing easy integration of data from disparate resources.

Aligned serial images, such as those from confocal microscopy, are a key tool for visualizing anatomical structures or regions of gene expression in the *Drosophila* nervous system [e.g. Ito *et al.*, 1997; Marin *et al.*, 2002; Yang *et al.*, 1995]. Demarcation (or 'painting') of known neuropil domains on top of such image stacks provides a powerful tool for research and study of the brain structure (Pereanu and Hartenstein, 2006; Rein *et al.*, 1999, 2002). Processing and viewing such data locally can be done using graphics software such as Amira (Stalling *et al.*, 2005) or ImageJ/Fiji (Abramoff *et al.*, 2004). There are also powerful specialist desktop-based systems for interactively exploring the *Drosophila* nervous system in 3D (Bruckner *et al.*, 2009; Peng *et al.*, 2010, 2011; Pettersen *et al.*, 2004). However, such datasets are also large and viewing them requires high-specification workstation hardware and a lot of storage rather than commodity computing.

Another characteristic of such image data is that it is expensive to produce and therefore its sharing and distribution is of paramount importance. Obviously sharing such bulky data imposes serious requirements on data storage and data transfer bandwidth. The ideal way of distributing and sharing image data is via web interfaces. Most existing projects in the community use *ad hoc* solutions such as downsized, pre-compiled QuickTime movies or representative reconstructions [e.g. Flytrap (Kelso *et al.*, 2004); BrainTrap (Knowles-Barley *et al.*, 2010)]. Some of the more recent solutions also provide interactive 3D browsing functions such as volume rotation [FlyCircuit (Chiang *et al.*, 2010)] although the reduction of resolution from raw dataset to that observed on the browser still remains an issue.

To more fully exploit these 3D images, mechanisms are required to display the boundaries and extent of defined structures, such

---

*To whom correspondence should be addressed.

as the neuropils defined by the BrainName standard, and to link exploratory browsing of images to the construction of queries on the underlying data.

Here were present an online resource, Virtual Fly Brain, inspired by the pioneering work of the Heisenberg group (Rein *et al.*, 2002) that builds upon the BrainName nomenclature (K.Ito *et al.*, submitted for publication) and crucially provides a computational framework upon which we can integrate tools, data and research resources.

2 RESULTS

2.1 Conversion of BrainName controlled vocabulary into an ontology

For the BrainName nomenclature to fulfil its potential for aiding data integration, it needs to be available as a standard vocabulary that can be used to annotate the location of gene expression, phenotypes and other neuroanatomy-related data. The standard vocabulary for annotating expression and phenotypes in *Drosophila* is the *Drosophila* anatomy ontology maintained by FlyBase (Tweedie *et al.*, 2009), the main community resource database for *Drosophila*. This ontology has been used by FlyBase to annotate over 40 000 phenotypes and expression patterns in the nervous system, including several thousand for the adult brain. It has been used for annotation by outside resources such as RedFly (Gallo *et al.*, 2011) and BrainTrap (Knowles-Barley *et al.*, 2010). Researchers involved in the BrainName group, which covers a large section of *Drosophila* neuroscientists, are also currently using it.

We have adapted and updated the *Drosophila* anatomy ontology to incorporate the BrainName standard terminology and linked the resulting ontology terms back to the annotated standard BrainName stack. We have also populated the ontology with terms for an extensive set of published neuron classes in the adult brain. These are linked to terms for neuropil domains and tracts using a newly defined set of interlinked relations (D.Osumi-Sutherland *et al.*, submitted for publication). These relations allow us to capture fasciculation and innervation patterns for each neuron class. The ontology is expressed in the W3C recommended OWL2 language (http://www.w3.org/TR/owl2-primer/). This allows us to use the standard OWL reasoner FaCT++ (Tsarkov and Horrocks, 2006) to autoclassify neurons and query the ontology.

The resulting ontology provides a common annotation system for annotating any type of neuroanatomy-related data, including gene expression, phenotypes and images.

2.2 System architecture

The application architecture is presented in Figure 1. It consists of a lightweight client (browser based) that communicates to the server. The client sends requests via the Web and receives back image tiles, image properties (Husz *et al.*, 2009) or ontology/DB query results in text form (Section 2.4).

The imaging server delivers high-resolution 2D image sections as a series of tiles covering the area visible in the browser window. Since the tiles are generated and delivered only when required and the viewer area is of constant size, the system uses memory and bandwidth efficiently and ensures consistent response time for user operations at multiple levels of magnification and arbitrary orientation. The tiles are produced from a 3D
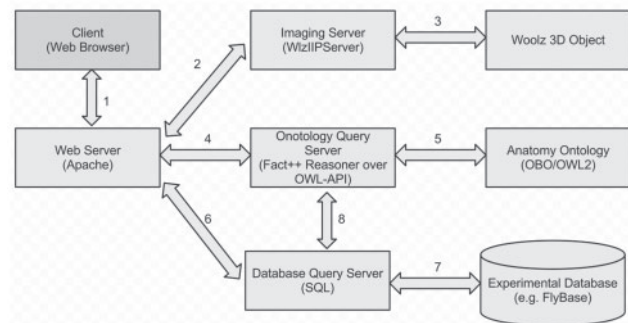

Fig. 1. *VFB Architecture.* The application consists of a lightweight client (top left) that communicates to the web server. The web server then relays the request to either the Imaging or Query Servers. The Imaging Server produces high-resolution 2D image sections from the Woolz 3D Object (2, 3). The Ontology Query Server is responsible for processing both simple and complex anatomy queries that are run against the Anatomy Ontology (4, 5). The Database Query Server processes queries to the underlying experimental database (6). Since the data annotation and querying is done with respect to the anatomy terms, the Ontology Server is asked first to retrieve a list of relevant anatomy terms (8), after which the SQL database is queried against that list (7). The summary here presents the simplest scenario as we can run multiple linked instances of data sources or to increase the efficiency of image data delivery.

image volume in Woolz object format [(Piper and Rutovitz, 1985); http://www.emouseatlas.org/emap/analysis_tools_resources/software/woolz.html] representing both the overall 'gray-level' structure and individual neuropil domains.

We run a full instance of the FlyBase CHADO Postgres database (Mungall and Emmert, 2007), kept in sync with the FlyBase update cycle. This contains a simple representation of the *Drosophila* anatomy ontology including term names, synonyms, definitions and references. It also contains a large volume of expression and phenotype data annotated using this ontology with links to genetic features (genes, alleles, transgenes) and source publications. As part of the Virtual Fly Brain project, we have extended the annotation of transgene expression in FlyBase for the adult brain to a near comprehensive set of published transgenes.

We run a separate ontology query server (Fig. 1) that uses the OWL-API [(Horridge and Bechhofer, 2009); http://owlapi.sourceforge.net/] to load the ontology and classify it using the OWL reasoner FaCT++ [(Tsarkov and Horrocks, 2006); http://code.google.com/p/factplusplus/]. Once the ontology has been classified, we can send description logic (DL) queries to the server, which uses data cached by the initial classification step to rapidly return a list of IDs of ontology terms that satisfy the query.

The query server is responsible for processing both ontology queries and annotation-related queries (Section 2.4).

2.3 Browser interface to the ontology and reference brain

The browser-side client is a configurable web application that allows the user to conveniently navigate around a painted 3D stack of a *Drosophila* brain and execute queries against the underlying brain anatomy ontology and external databases.
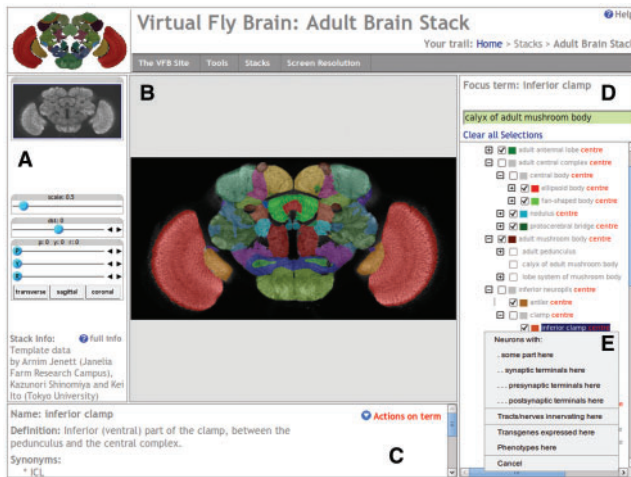
Fig. 2. *The Browser Interface.* The central area of the Viewer displays a slice through the image stack (**B**). Tools for user interaction with the image are in the left-hand side panel (**A**). These allow the user to specify the content and parameters of a current view and include navigation, zoom, depth and rotation. The navigation tool indicates the currently visible region of the image, and allows easy scrolling of the image by dragging. The zoom tool allows one to choose a resolution level for the displayed tiles. The rotation tool provides the means to specify the angle for the plane that is used for production of a 2D section from the 3D stack. Using this tool, the user has a choice to either define an arbitrary angle for the section or to select one of the three preset orientations. The depth tool allows the user to specify which of the series of parallel section planes to use for 2D rendering within the stack depth-wise. (**C**) Provides information on the currently selected term. (**D**) Includes an autocomplete search box and interactive ontology tree. Right/ctrl-clicking on terms in the tree (**E**) brings up a series of simple query options.

The most intuitive way to interact with brain anatomy is via the structures themselves, we therefore went to some effort to supplement tree-like ontology representations with interactive 3D brain representations. The image viewer at the centre of our browser interface (Fig. 2) fulfils this role. It serves 2D slices of a 3D image stack and along with a configurable colour overlay. In the current version of VFB, the only available stack is the BrainName reference stack and a configurable colour overlay defines the boundaries of BrainName regions. We will add reference stacks for other stages and parts of the nervous system as they become available.

A 'Google Maps-like' user interface makes the image viewer highly intuitive to use. This is supplemented with advanced controls, such as depth and stack orientation, to facilitate full 3D navigation through the volume. The interactive anatomy tree allows one to navigate the anatomy based on name or position in the part hierarchy. Clicking on a region of interest in a brain slice toggles the selection of the corresponding term in the anatomy tree and vice versa. From the tree, or by clicking on the image, users can toggle the display of multiple regions as colour overlays. Users can then move, rotate or zoom to any arbitrary view to explore the spatial relationships between the selected anatomical regions.

The autocomplete search box (Fig. 2D) provides a facility for quickly searching for anatomy terms. The list that drives autocompletion is populated using a DL query that finds all brain regions relevant to the image stack. In the case of the BrainName adult brain stack, the DL query finds all regions of synaptic neuropil that are part of the adult brain. For each anatomy term found, a list of all known synonyms is retrieved by querying the CHADO database and added to the autocomplete list. This allows searching on both the official names and synonyms with the same end result—a corresponding tree node for the selected term gets highlighted. This provides not only a convenient and forgiving way of searching, but also a quick linking mechanism between the colloquial brain term names and BrainName standard definitions. The structure of the anatomy tree is generated automatically from the anatomy ontology by traversing the part relationships for the 'adult brain'. This approach will allow us to easily extend Virtual Fly Brain to use other image stacks, and allows us to automatically update the search and tree when the underlying ontology changes for any particular stack.

The VFB application provides a basic atlas functionality allowing users to explore and interact with the FlyBase anatomy ontology and the BrainName nomenclature from which it is derived. There will be other atlases available in the community (e.g. K.Ito, personal communication) providing a range of visual methods to explore, interact with and learn fly brain anatomy.

2.4 Graphically driven ontology queries

Our primary aim was not to develop a reference atlas *per se*, but rather to provide an intuitive and interactive query interface with an underpinning logical framework upon which useful searches can be constructed and solved. The guiding premise is that any data source that is based around, annotated with or mapped onto the same ontologies should be searchable from a single site. In the first instance, we focused on the valuable phenotype and expression data curated from the literature and stored in FlyB. The simplest form of query on Virtual Fly Brain is initiated by selection of a tree term, either directly or by double clicking on a region on the image stack. When a tree term is selected, a request to the database is initiated, resulting in the detailed information record for that term being displayed in a 'Term info box' (Fig. 2C). Currently, this uses information stored in our local instance of CHADO retrieved based on the IDs of the found anatomy terms. In the future, we plan to extract this information directly from the ontology server. This will make our software more generically usable as it will not depend on having an ontology loaded into a CHADO database.

As more than one source of information is available for any selected term (e.g. anatomical, phenotypic, connectivity, etc.), each term has an associated contextual menu accessed by a right/ctrl click on the corresponding tree node (Fig. 2E). As additional data sources are integrated and thus new queries become feasible, this menu gets extended. Two classes of queries—anatomy and annotation—are currently accessible from this menu.

Anatomy queries drive DL queries via the ontology server. They currently comprise a query for tracts innervating a selected neuropil, and a nested set of queries for neurons with increasing specificity. The most general of these is a query for all neurons with some part in the selected neuropil. A more specific query finds all neurons with synaptic terminals in the neuropil and more specific queries still find neurons with pre-synaptic or post-synaptic terminals. These queries rely on inference over the ontology using a set of interlinked relations for overlap, 'part of' and synapse location. Figure 3 shows one example of inference in a query for neurons with some part
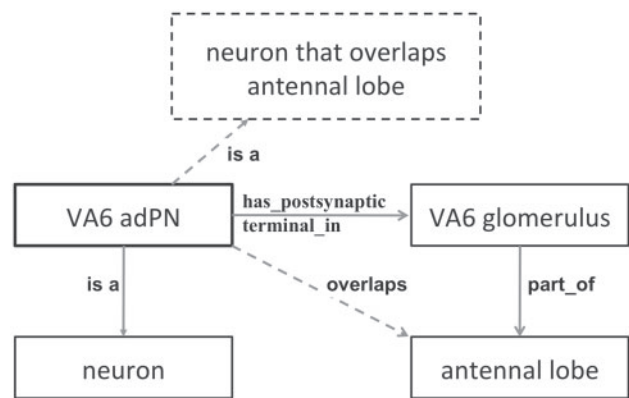
Fig. 3. *Reasoning used to answer the query: 'Find neurons with some part in the antennal lobe'.* Boxes represent anatomical classes, solid arrows asserted relationships between classes and dotted arrows inferred relationships. The query works by defining a new class—the class of neurons that overlap the antennal lobe (dotted box) and then asking reasoner software to find any class that **is a** subclass of this class. Using the asserted relationships shown and rules in the ontology (detailed below), the reasoner infers that neurons of class 'VA6 adPN' overlap the antennal lobe. This, combined with the assertion that VA6 adPN is a class of neuron is sufficient for the reasoner to infer that VA6 adPN is a 'neuron that overlaps the antennal lobe'. More formally: the reasoner returns classes, such as 'VA6 adPN', that satisfy the DL query 'neuron *that* **overlaps** *some* "antennal lobe" '. 'is a' (filled arrowheads) corresponds to *SubClassOf*. Relationships with unfilled arrowheads follow the pattern 'X *SubClassof* R *some* Y', e.g. 'VA6 glomerulus' *SubClassOf* **part_of** *some* 'antennal lobe'. Rules used (encoded as OWL property chains and property hierarchy) are as follows: if X **has_postsynaptic_terminal_in** Y and Y **part_of** Z then X **has_postsynaptic_terminal_in** Z; if X **has_postsynaptic_terminal in** Y then X h**as_synaptic_terminal_in** Y; if X **has_synaptic_terminal in** Y then X **overlaps** Y.

Fig. 4. *Query result example.* Output from a query designed to search for, and retrieve, information on transgenes expressed in a specific anatomical structure.

in the antennal lobe. These relations are discussed in detail in an accompanying paper (D.Osumi-Sutherland *et al.*, submitted for publication). The resulting lists are enhanced with extra information, such as a definition extract, pulled from the representation of the ontology in FlyBase CHADO.

Annotation queries, such as a query for expression or phenotype, begin with a DL query for all relevant anatomy terms. The resulting list is used to query FlyBase CHADO for annotated features and source publications (Fig. 4). For queries of transcript expression in a specified neuropil, the query for relevant anatomy terms returns all neurons that overlap that neuropil, as well as parts of the neuropil. This leads to some false positives in query results, as a transcript might be localized to part of a neuron that does not overlap the neuropil. But this is more than compensated for by the number of

correct results found by this method but not by queries for expression in parts alone. This is illustrated by the query results in Figure 4, which find expression of P{GaWB}Mz612 in the neuron VA6 adPN which was found by a clause in the initial DL query for relevant terms.

This approach to querying is quite different from typical uses of the Gene Ontology and other OBO ontologies by model organism databases. These databases use simple graph-based reasoning to group annotation to a query term with annotations to terms for all its subclasses and subparts. In this paradigm, a query for transgenes expressed in the antennal lobe could find transgenes annotated as expressed in parts of the antennal lobe, such as glomerulus VA1 but not in neurons that innervate the lobe, such as VA6 adPN (Fig. 4).

An anatomy query–builder interface allows the user to build more complex anatomy queries, such as 'find all neurons that have pre-synaptic terminals in X and post-synaptic terminal in Y'. This interface includes an image viewer with simpler set of controls and the anatomy tree to enable graphical selection of the brain regions for individual query legs. The query builder part of the interface allows users to specify a type of relation (e.g. synaptic, post-synaptic, pre-synaptic) for each individual query leg and provides intermediate feedback on how many hits the resulting query would yield. As the query is built, a sentence is generated clearly stating the nature of the query. This is straightforward to do given the well-defined semantics of the ontology, but would be much more challenging for combinatorial queries from a conventional relational database.

3 CONCLUSIONS AND FUTURE DIRECTIONS

The Virtual Fly Brain site was developed with two major aims. First, it is designed as a hub for neuro-anatomical data integration. Secondly, it is intended to be an easily accessible and usable tool to disseminate community agreed anatomical standards. It currently focuses on the standard proposed by the BrainName consortium for the adult brain, but we plan to support standards for other regions and stages of the *Drosophila* central nervous system once they are available. Virtual Fly Brain achieves its dissemination aim by being accessible on the web, without downloading or installation of bespoke components, and by being easy to search, query and browse. We have demonstrated how anatomy, gene expression and phenotype queries can be easily constructed using the browser interface's context menu and query interface. Virtual Fly Brain achieves its data integration aim through its use and development of the *Drosophila* anatomy ontology. This provides both a means for storing information about *Drosophila* neuroanatomy in queryable form and a vocabulary for annotating data.

There are other, high-quality efforts to collect information about neuronal innervation patterns in the *Drosophila* brain into a queryable resource. The FlyCircuit database (Chiang *et al.*, 2010) does this using its own data comprised thousands of individual neuron image stacks and a query system built around its own, innovative image analysis system. It is, therefore, a unique resource whose data and analysis are complementary to ours. Another significant resource, FlyBrain Neuron Database (Shinomiya *et al.*, 2011), leverages the great anatomical expertise of the lab in which it is based along with a conventional relational database approach to record information about adult brain neurons. The combination of data and insights unique to the lab in which it originates along with various visualization tools means that this will continue to be a

valuable resource. However, where this database overlaps with our effort, our ontology-based approach has significant advantages for data integration, via use of the ontology in annotation, and querying, due to its more sophisticated semantics.

Virtual Fly Brain queries can be run against any third-party database that either contains data annotated with the *Drosophila* anatomy ontology or that can easily be mapped to this ontology. Our current version runs queries against FlyBase, instantly exploiting its vast store of data curated from the literature. Content is under constant revision and we are currently importing expression annotation that already uses our anatomy ontology from the BrainTrap database (Knowles-Barley *et al.*, 2010) into FlyBase and Virtual Fly Brain and have mapped FlyBrain Neuron DataBase records to our ontology. We are also auto-annotating neuron image data from FlyCircuit using cross-registration of the FlyCircuit and BrainName model brains. Links from relevant pages on our site to all three of these resources will be added in the near future, along with integration of FlyCircuit and BrainTrap data into our query system.

The application code is freely available, configurable and reusable for other projects and potentially also for other organisms. The web site is available at http://www.virtualflybrain.org. Please see http://www.virtualflybrain.org/site/vfb_site/tutorial.htm for video tutorials.

4 ACKNOWLEDGEMENTS

We acknowledge the efforts of the entire BrainName Nomenclature Working Group in proposing a unified nomenclature for the insect brain. In particular, we are indebted to the contributions of Kei Ito, Arnim Jennet, Greg Jefferis and Kasunori Shinomiya who generously provided images and commentary on annotations used throughout VFB. We also thank FlyBase for a very fruitful collaboration without which the VFB project would not be possible. We finally acknowledge the efforts of Michael Ashburner in helping us define the initial project. Ill health prevented him from contributing further.

Conflict of Interest: none declared.

REFERENCES

Abramoff,M.D. *et al.* (2004) Image Processing with ImageJ. *Biophotonics Int.*, **11**, 36–42.

Bruckner,S. *et al.* (2009) BrainGazer–visual queries for neurobiology research. *IEEE Trans. Vis. Comput. Graph*, **15**, 1497–1504.

Chiang,A.S. *et al.* (2010) Three-dimensional reconstruction of brain-wide wiring networks in Drosophila at single-cell resolution. *Curr. Biol.*, **21**, 1–11.

Gallo,S.M. *et al.* (2011) REDfly v3.0: toward a comprehensive database of transcriptional regulatory elements in Drosophila. *Nucleic Acids Res.*, **39**, D118–D123.

Horridge,M. and Bechhofer,S. (2009) The OWL API: a Java API for working with OWL 2 ontologies. In *Proceedings of the 6th International Workshop on OWL: Experiences and Directions (OWLED 2009), Chantilly, VA, United States*, CEUR Workshop Proceedings, Aachen, Germany.

Husz,Z.L. *et al.* (2009) Woolz IIP: a tiled on-the-fly sectioning server for 3D volumetric atlases. *Adv. Vis. Comput.*, **5875**, 924–933.

Ito,K. *et al.* (1997) The Drosophila mushroom body is a quadruple structure of clonal units each of which contains a virtually identical set of neurones and glial cells. *Development*, **124**, 761–771.

Kelso,R.J. *et al.* (2004) Flytrap, a database documenting a GFP protein-trap insertion screen in Drosophila melanogaster. *Nucleic Acids Res.*, **32**, D418–D420.

Knowles-Barley,S. *et al.* (2010) BrainTrap: a database of 3D protein expression patterns in the Drosophila brain. *Database*, **2010**, baq005.

Marin,E.C. *et al.* (2002) Representation of the glomerular olfactory map in the Drosophila brain. *Cell*, **109**, 243–255.

Mungall,C.J. and Emmert,D.B. (2007) A Chado case study: an ontology-based modular schema for representing genome-associated biological information. *Bioinformatics*, **23**, i337–i346.

Peng,H. *et al.* (2010) V3D enables real-time 3D visualization and quantitative analysis of large-scale biological image data sets. *Nat. Biotechnol.*, **28**, 348–353.

Peng,H. *et al.* (2011) BrainAligner: 3D registration atlases of Drosophila brains. *Nat. Methods*, **8**, 493–500.

Pettersen,E.F. *et al.* (2004) UCSF Chimera–a visualization system for exploratory research and analysis. *J. Comput. Chem.*, **25**, 1605–1612.

Piper,J. and Rutovitz,D. (1985) Data structures for image processing in a C language and Unix environment. *Pattern Recognit. Lett.*, **3**, 119–129.

Pereanu,W. and Hartenstein,V. (2006) Neural lineages of the Drosophila brain: a three-dimensional digital atlas of the pattern of lineage location and projection at the late larval stage. *J. Neurosci.*, **26**, 5534–5553.

Rein,K. *et al.* (1999) A quantitative three-dimensional model of the Drosophila optic lobes. *Curr. Biol.*, **9**, 93–96

Rein,K. *et al.* (2002) The Drosophila standard brain. *Curr. Biol.*, **12**, 227–231.

Shinomiya,K. *et al.* (2011) Flybrain neuron database: a comprehensive database system of the Drosophila brain neurons. *J. Comp. Neurol.*, **519**, 807–833.

Stalling,D. *et al.* (2005) Amira: a highly interactive system for visual data analysis. In Johnson,C.R. and Hanson,C.D. (eds) *The Visualization Handbook*. Elsevier Academic Press, Orlando, pp. 749–767.

Tsarkov,D. and Horrocks,I. (2006) FaCT++ description logic reasoner: system description. *Lect. Notes Artif. Intell.*, **4130**, 292–297.

Tweedie,S. *et al.* (2009) FlyBase: enhancing Drosophila Gene Ontology annotations. *Nucleic Acids Res.*, **37**, D555–D559.

Yang,M.Y. *et al.* (1995) Subdivision of the Drosophila mushroom bodies by enhancer-trap expression patterns. *Neuron*, **15**, 45–54.