



Published in final edited form as:

*Neuroinformatics*. 2015 January ; 13(1): 127–129. doi:10.1007/s12021-014-9257-y.

## Doubling up on the fly: [NeuroMorpho.Org](#) meets Big Data

Sumit Nanda, M. Mowafak Allaham, Maurizio Bergamino<sup>#</sup>, Sridevi Polavaram, Rubén Armañanzas, Giorgio A. Ascoli<sup>\*</sup>, and Ruchi Parekh

Center for Neural Informatics, Structures, & Plasticity, Krasnow Institute for Advanced Study, George Mason University, Fairfax, VA (USA)

[NeuroMorpho.Org](#) is a centralized repository of neuronal reconstructions hosting data from a variety of species, brain regions, and experimental conditions<sup>1</sup>. This resource aims to provide dense coverage of available data by including all digital tracings described in peer-reviewed publications that the authors are willing to share<sup>2</sup>. Although most reconstructions to date are acquired manually or semi-manually<sup>3</sup>, the transition to quasi-automated methods is widely considered as necessary for long-term progress<sup>4</sup>. The 2010 DIADEM competition ([DiademChallenge.org](#)) helped foster considerable advances towards tracing automation<sup>5</sup> and was followed one year later by the large-scale reconstruction of more than 16,000 *Drosophila* neurons<sup>6</sup>. The public posting of all image stacks and corresponding digital tracings on [flycircuit.tw](#) after an additional year<sup>7</sup> constituted the first (and so far only) success in high-throughput digital morphology.

Although [flycircuit.tw](#) reconstructions are beginning to enable new analysis and discoveries by independent research groups<sup>8</sup>, these data were posted in a commercial format ([vsg3d.com/amira/skeletonization](#)) and lacked useful information such as the somatic brain region. Following the open invitation of [flycircuit.tw](#) to copy, transform, and redistribute the material for non-commercial re-use, here we announce inclusion of this dataset in non-proprietary SWC format, along with additional metadata and morphometric measurements, under “Chiang archive” in [NeuroMorpho.Org](#) version 6.0. With this major release, the number of [NeuroMorpho.Org](#) reconstructions more than doubles from 11,335 to 27,385.

## Data Conversion and Standardization

Each of the existing 16,050 reconstruction files (out of the 16,227 [flycircuit.tw](#) neuron pages) were initially converted from the posted Amira representation into the *de facto* community standard SWC, which is compatible with all freely available visualization, analysis, and modeling tools. Due to its large number, the whole dataset was processed with an automated variant of the [NeuroMorpho.Org](#) standardization pipeline ([neuromorpho.org/neuroMorpho/StdSwc1.21.jsp](#)). Accordingly, only the following irregularities were checked: trifurcations, long connections, overlapping points, and large radius. Essential metadata and measurements were then computed for each of these neurons to enable full search and browse functionality. The added information included somatic region assignment, neuron type classification (both further explained below), morphometric quantification<sup>9</sup>, and strain

<sup>\*</sup>Correspondence: [ascoli@gmu.edu](mailto:ascoli@gmu.edu).

<sup>#</sup>Current address: Laureate Institute for Brain Research, Tulsa, OK (USA)

(genetic) mapping against [flybase.org](http://flybase.org)<sup>10</sup>. Moreover, a second version of the reconstruction file was included for a subset of 617 neurons as made available from an independently published report using the same image stacks<sup>11</sup>.

## Brain Region Assignment

The structure of metadata for neuron location in [NeuroMorpho.Org](http://NeuroMorpho.Org) largely follows the typical organization of mammalian brains in regions, sub-regions, layers, and/or nuclei. In contrast, the somata in the fly central nervous system tend to line up on the neuropil surface<sup>12</sup>. As a consequence, it is not uncommon for cell bodies to lie near the border of two (or occasionally three) neuropils. We therefore leveraged [flycircuit.tw](http://flycircuit.tw) online records, query tools, and image stacks to map every soma of the 16,050 neurons to one, two or (rarely) three brain regions. We first interrogated the [flycircuit.tw](http://flycircuit.tw) text-based search engine to list all neurons within 0  $\mu\text{m}$  of each of the 58 represented regions. This operation returned 5533 neurons mapped to single regions and no neuron mapped to more than one region. Next we searched neurons within 10  $\mu\text{m}$  of each region. This step returned 8171 neurons mapped to single regions, including 3958 of the 5533 found at 0  $\mu\text{m}$ . The brain region matched between 0 and 10  $\mu\text{m}$  for 3894 of those 3958 neurons, and mismatched for only 64. Assuming the assignment at 0  $\mu\text{m}$  as the gold standard, these values correspond to 98.4% reliability for the assignments at 10  $\mu\text{m}$ . We therefore accepted the further assignment of the additional 4213 neurons. With the same approach, we assigned the somatic brain region of 2475 more neurons (with estimated 84.2% reliability) by searching within 20  $\mu\text{m}$ . When we attempted to search within 30  $\mu\text{m}$ , however, the estimated reliability dropped to 13.6%; thus, we rejected those assignments. Altogether, the above described process assigned the somatic brain region of 12,219 neurons.

The searches within 10 and within 20  $\mu\text{m}$  also returned 1984 and 535 neurons, respectively, matching two or more brain regions. The last 1489 neurons could not be mapped to any region within 20  $\mu\text{m}$  and were considered as potentially residing in any of the 58 regions. The total number of neurons sums up to 16,227, but only 16,050 of these had an associated morphological reconstruction file. To assign the somatic location of every neuron matching two or more brain regions, we first calculated the Euclidean distance between that neuron and each of the neurons uniquely mapped to one of those regions (a subset of the 12,219 uniquely mapped neurons) using the corresponding [flycircuit.tw](http://flycircuit.tw) atlas coordinates. We then defined the proximity of the neuron to each of those regions as the sum of the inverse of the distance to every neuron in that region. For example, if a neuron matched both Mushroom Body (MB) and Medulla (Med) in one hemisphere, given the distances of that neuron from the  $n$  Mushroom Body neurons ( $d_{MB1}, d_{MB2}, d_{MB3}, \dots, d_{MBn}$ ) and from the  $m$  Medulla neurons ( $d_{Med1}, d_{Med2}, d_{Med3}, \dots, d_{Medm}$ ), the respective proximities are:  $\text{Prox}_{MB} = \sum_i (1/d_{MBi})$ , where  $i=1 \dots n$ , and  $\text{Prox}_{Med} = \sum_j (1/d_{Medj})$ , where  $j=1 \dots m$ . These proximity values are then used to determine the relative assignment probabilities for MB and Med:  $\text{Prob}_{MB} = \text{Prox}_{MB} / (\text{Prox}_{MB} + \text{Prox}_{Med})$  and  $\text{Prob}_{Med} = \text{Prox}_{Med} / (\text{Prox}_{MB} + \text{Prox}_{Med})$ .

If a neuron only matched two regions, we assigned the somatic location exclusively to one region if its probability was at least twice as high as that of the other region (that is, if one of the two regions had a probability of 66.67% or higher). Otherwise, we assigned the neuron

to both regions (representing a border location between the two). For neurons matching three or more regions, we devised a heuristic decision tree to associate one, two or three regions (representing their location within one region, at the border between two regions, and in the intersection of three regions, respectively) based on the corresponding assignment probabilities. The details of the decision tree are available at [NeuroMorpho.Org/neuroMorpho/techDocFlyData.jsp](http://NeuroMorpho.Org/neuroMorpho/techDocFlyData.jsp). Following this procedure, 14,409 neurons were mapped to one region, 1632 to two, and 9 to three.

To verify the results of this empirical process, we built an adjacency matrix of male and female brain regions based on the corresponding [flycircuit.tw](http://flycircuit.tw) template image stacks. We considered two regions adjacent if they share a border or if they are in close proximity without a third region in the middle. We employed this matrix to check that the double or triple somatic region assignments by the decision tree corresponded to adjacent cases, manually inspecting all doubtful cases against the original [flycircuit.tw](http://flycircuit.tw) images. Lastly, we translated the assigned regions to [NeuroMorpho.Org](http://NeuroMorpho.Org) metadata entries based on the [VirtualFlyBrain.org](http://VirtualFlyBrain.org) hierarchy<sup>13</sup>. For neurons assigned to more than one region, we mapped only the most likely region to the hierarchy, adding the other assigned region(s) as bordering locations in [NeuroMorpho.Org](http://NeuroMorpho.Org) metadata.

## Neuron Type Assignment

The distinction between principal (projection) cells and (local) interneurons was based on the [flycircuit.tw](http://flycircuit.tw) list of regions invaded by the neurite terminals of every neuron. We considered a neuron as an interneuron if 95% or more of its terminals were contained within the somatic region and its adjacent brain regions. Conversely, we marked a neuron as a principal cell if more than 5% of its terminals were found in non-adjacent regions. This definition yielded 10,079 principal cells and 5971 interneurons. We further sub-divided all neurons on the basis of their putative neurotransmitter and, lastly, by their birth date.

## Future implications

The content of [NeuroMorpho.Org](http://NeuroMorpho.Org) has grown continuously from just shy of 1000 reconstructions in the August 2006 beta release to 11,335 in version 5.7 (May 2014), paralleling considerable developments in the field<sup>14</sup>. This increase has been accompanied by steady rise in downloads, secondary discoveries, and citations<sup>15</sup>. The newly included Chiang archive constitutes the largest available dataset and is the second “atlas-like” collection of neurons in the repository (after *C. elegans*). All 16,050 latest reconstructions in version 6.0 can now be browsed, searched, visualized, inspected, and downloaded as the rest of the repository content.

This major release significantly shifts the balance of available data. Up to version 5.7 the dominant species were rodents, accounting for nearly two-thirds of reconstructions (evenly split between rat and mouse). In version 6.0, *Drosophila* alone comfortably sweeps up the absolute majority. In the same vein, neocortical neurons, until recently constituting half of the digital tracings, now represent a mere one fifth of the available data, and pyramidal neurons followed the same numeric fate. Similarly, bright field microscopy of sectioned slices from Golgi staining or intracellularly injected dyes was up to this last release by far

the most popular preparation in digital morphology. The dominant experimental approach now becomes whole-mount confocal microscopy with genetically labeled fluorescent proteins.

On the one hand, [NeuroMorpho.Org](#) content will likely even out, at least in the short-term. More than 8000 additional neurons from recent publications are already in the processing pipeline for future releases, including individual archives of more than 2000 reconstructions each from studies in culture<sup>16</sup> and in vitro<sup>17</sup>. We typically identify more than one thousand new reconstructions every month through systematic literature searches and we receive several hundred of those upon request for public sharing in the same time span. Thus, within one or two years the distribution of metadata is expected to evolve further and adjust substantially. On the other hand, the Chiang dataset also constitutes a possibly historical shift in technology from manual to automated reconstructions. In the likely future prospect of a definitive transition to fuller automation, it is expected that the size of typical individual archives (and eventually of the entire repository) will grow by one order of magnitude or more. Given the successful impact [NeuroMorpho.Org](#) has already had in the neuroscience community, we expect that this major update and future anticipated releases will foster substantial research advancement.

## Acknowledgements

The authors are grateful to the NCHC (National Center for High-performance Computing) and NTHU (National Tsing Hua University), Hsinchu, Taiwan, for publicly sharing their data. This work is supported by NIH R01's NS39600 (BISTI) and NS086082 (CRCNS) from NINDS, ONR MURI 14101-0198, and Keck NAKFI to GAA.

## References

1. Ascoli GA, Donohue DE, Halavi M. [NeuroMorpho.Org](#) - A central resource for neuronal morphologies. *J. Neurosci.* 2007; 27:9247–51. [PubMed: 17728438]
2. Halavi M, Polavaram S, Donohue DE, Hamilton G, Hoyt J, Smith K, Ascoli GA. [NeuroMorpho.Org](#) implementation of digital neuroscience: dense coverage and integration with the NIF. *Neuroinformatics.* 2008; 6:241–52. [PubMed: 18949582]
3. Halavi M, Hamilton KA, Parekh R, Ascoli GA. Digital reconstructions of neuronal morphology: three decades of research trends. *Front Neurosci.* 2012; 6:49. doi: 10.3389/fnins.2012.00049. [PubMed: 22536169]
4. Svoboda K. The past, present, and future of single neuron reconstruction. *Neuroinformatics.* 2011; 9:97–8. [PubMed: 21279476]
5. Liu Y. The DIADEM and beyond. *Neuroinformatics.* 2011; 9:99–102. [PubMed: 21431331]
6. Chiang AS, Lin CY, Chuang CC, Chang HM, Hsieh CH, Yeh CW, Shih CT, Wu JJ, Wang GT, Chen YC, Wu CC, Chen GY, Ching YT, Lee PC, Lin CY, Lin HH, Wu CC, Hsu HW, Huang YA, Chen JY, Chiang HJ, Lu CF, Ni RF, Yeh CY, Hwang JK. Three-dimensional reconstruction of brain-wide wiring networks in *Drosophila* at single-cell resolution. *Curr Biol.* 2011; 21:1–11. [PubMed: 21129968]
7. Lee PC, Chuang CC, Chiang AS, Ching YT. High-throughput computer method for 3D neuronal structure reconstruction from the image stack of the *Drosophila* brain and its applications. *PLoS Comput Biol.* 2012; 8(9):e1002658. doi: 10.1371/journal.pcbi.1002658. [PubMed: 23028271]
8. Costa, M.; Ostrovsky, AD.; Manton, JD.; Prohaska, S.; Jefferis, GS. NBLAST: Rapid, sensitive comparison of neuronal structure and construction of neuron family databases. *BioRxiv.* 2014. DOI: <http://dx.doi.org/10.1101/006346>

9. Scorcioni R, Polavaram S, Ascoli GA. L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Prot.* 2008; 3:866–76.
10. St. Pierre SE, Ponting L, Stefancsik R, McQuilton P, the FlyBase Consortium. FlyBase 102 - advanced approaches to interrogating FlyBase. *Nucleic Acids Res.* 2014; 42(D1):D780–D788. [PubMed: 24234449]
11. Xiao H, Peng H. APP2: automatic tracing of 3D neuron morphology based on hierarchical pruning of a gray-weighted image distance-tree. *Bioinformatics.* 2013; 29:1448–54. [PubMed: 23603332]
12. Rivera-Alba M, Peng H, de Polavieja GG, Chklovskii DB. Wiring economy can account for cell body placement across species and brain areas. *Curr Biol.* 2014; 24:R109–10. [PubMed: 24502781]
13. Milyaev N, Osumi-Sutherland D, Reeve S, Burton N, Baldock RA, Armstrong JD. The Virtual Fly Brain browser and query interface. *Bioinformatics.* 2012; 28:411–5. [PubMed: 22180411]
14. Parekh R, Ascoli GA. Neuronal morphology goes digital: a research hub for cellular and system neuroscience. *Neuron.* 2013; 77:1017–38. [PubMed: 23522039]
15. Parekh R, Ascoli GA. Quantitative investigations of axonal and dendritic arbors: development, structure, function, and pathology. *Neuroscientist.* 2014 pii: 1073858414540216. [Epub ahead of print] PMID: 24972604.
16. Hamad MI, Ma-Högemeier ZL, Riedel C, Conrads C, Veitinger T, Habijan T, Schulz JN, Krause M, Wirth MJ, Hollmann M, Wahle P. Cell class-specific regulation of neocortical dendrite and spine growth by AMPA receptor splice and editing variants. *Development.* 2011; 138:4301–13. [PubMed: 21865324]
17. Nogueira-Campos AA, Finamore DM, Imbiriba LA, Houzel JC, Franca JG. Distribution and morphology of nitrergic neurons across functional domains of the rat primary somatosensory cortex. *Front Neural Circuits.* 2012; 6:57. doi: 10.3389/fncir.2012.00057. [PubMed: 23133407]