Neuroinform
DOI 10.1007/s12021-015-9272-7

ORIGINAL ARTICLE

BlastNeuron for Automated Comparison, Retrieval and Clustering of 3D Neuron Morphologies

Yinan Wan[1] · **Fuhui Long**[1,2] · **Lei Qu**[1,4] · **Hang Xiao**[1] · **Michael Hawrylycz**[2] · **Eugene W. Myers**[1,3] · **Hanchuan Peng**[1,2]

Abstract Characterizing the identity and types of neurons in the brain, as well as their associated function, requires a means of quantifying and comparing 3D neuron morphology. Presently, neuron comparison methods are based on statistics from neuronal morphology such as size and number of branches, which are not fully suitable for detecting local similarities and differences in the detailed structure. We developed *BlastNeuron* to compare neurons in terms of their global appearance, detailed arborization patterns, and topological similarity. *BlastNeuron* first compares and clusters 3D neuron reconstructions based on global morphology features and moment invariants, independent of their orientations, sizes, level of reconstruction and other variations. Subsequently, *BlastNeuron* performs local alignment between any pair of retrieved neurons via a tree-topology driven dynamic programming method. A 3D correspondence map can thus be generated at the resolution of single reconstruction nodes. We applied *BlastNeuron* to three datasets: (1) 10,000+ neuron reconstructions from a public morphology database, (2) 681 newly and manually reconstructed neurons, and (3) neurons reconstructions produced using several independent reconstruction methods. Our approach was able to accurately and efficiently retrieve morphologically and functionally similar neuron structures from large morphology database, identify the local common structures, and find clusters of neurons that share similarities in both morphology and molecular profiles.

Keywords Neuron comparison · Neuron morphology · Tree matching · Neuron reconstruction

✉ Hanchuan Peng
hanchuanp@alleninstitute.org

[1] Janelia Research Campus, Howard Hughes Medical Institute, Ashburn, VA, USA

[2] Allen Institute for Brain Science, Seattle, WA, USA

[3] Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany

[4] Key Laboratory of Intelligent Computation and Signal Processing, Ministry of Education, Anhui University, Hefei, China

Introduction

Neurons and glial cells are fundamental building blocks of the brain. These cells usually have distinct, tree-like arborization patterns, which distinguish themselves clearly from most other types of cells, and also discriminate one neuron (cell) type from another. The 3D morphology of a neuron (or a glial cell), which is typically reconstructed from microscopic images, is also believed to correlate well with the neuron's physiological properties and functions (Dumitriu et al. 2007). Increasingly available 3D neuronal morphology databases, such as NeuroMorpho (neuronmorpho.org) (Ascoli et al. 2007) and FlyCircuit (flycircuit.org) (Chiang et al. 2011), demand high performance computational methods to understand questions such as: how to compare neuronal structures, how to retrieve neurons sharing similar 3D morphology, and how to organize them into biologically meaningful clusters. 3D neuron comparison techniques are crucial to characterizing neurons and studying their networks (Peng et al. 2013).

Prior to the present work, the comparison of two or more neuron structures has been mainly studied in two scenarios. In the first case, neuron structures with significant geometrical variation are compared using morphological descriptors of the entire neuron (Scorcioni et al. 2008), or secondly, decomposed into hierarchy of neurite segments and then compared using morphological statistics (Basu et al. 2011).

Published online: 03 June 2015

However, these approaches typically account for the global similarities between neurons. Neither is designed to directly compare neuron morphologies at the single compartment (i.e., reconstruction-node) level. Pinpointing local differences in morphology has only been applied in the second method, where multiple reconstructions are produced in the same coordinate system. One can compute the spatial divergence of neurons in a point-to-point or segment-by-segment manner and thus determine the correspondence of all reconstruction nodes in different reconstructions (Peng et al. 2010, 2014; Costa et al. 2014; Ganglberger et al. 2014), or reconstruction with ground truth (Mayerich et al. 2012). These approaches, however, use little or limited topological difference observed between neurons. They may also fail when the reconstructions cannot be pre-aligned in the same coordinate system. Other methods, such as Cardona et al. (2010) decomposed unbranched neurites into 3D curves and aligned them using dynamic programming, but unbranched neuronal segments do not naturally represent tree structures and thus those methods failed to deal with complex structures with many segments. Independently to our work, Gillette et al. developed an algorithm that decomposed the neuron topology into sequences of branching patterns, and then compared neuron structures using sequence alignment (Gillette and Ascoli 2015; Gillette et al. 2015). The method was able to identify the difference in branching patterns in dendritic and axonic arbors and extracting common topological "motifs" in the structure. Yet, such an analysis uses only topology but not geometry of neurons.

The limitation of the previous methods for 3D neuron comparison is partly due to the diversity in neuronal structures themselves: comparison of neuron morphologies at the compartment level can be achieved only when handling very similar morphologies. On the other hand, even if the neurons are globally similar, the anatomical and topological complexity in their morphology prohibits us from using conventional point cloud or tree comparison algorithms, which cannot handle neurons that have complex articulation and branching patterns (Tschirren et al. 2005). Therefore, a reasonable neuron comparison strategy should simultaneously combine searching for globally similar neurons and comparing their detailed structure, and in this context it is necessary to develop novel algorithms particularly suited to local morphology comparison.

We designed the *BlastNeuron* (Basic Local Alignment Search Tool for Neurons) system as a framework for rapidly searching morphologically similar neurons in a large database of 3D digital neuron reconstructions, as well as locally aligning similar neurons to pinpoint their common substructures or difference. Specifically, *BlastNeuron* includes a package of computer algorithms to analyze both the global and local structures of neurons. *BlastNeuron* integrates morphological features, invariant moments features, shape context and deformable mapping to compare and align 3D neuron reconstructions. Our software is able to automatically search large databases of neurons with 10,000+ neurons and retrieve similar neuron structures, independent of their orientations, sizes, level of reconstruction or other variations. We demonstrate that *BlastNeuron* is able to pinpoint differences at the single reconstruction-node resolution. We tested *BlastNeuron* using neuron morphologies from the publicly available database Neuromorpho.org, as well as 681 *Drosophila* neurons newly reconstructed from microscopic images. *BlastNeuron* was able to provide interesting clusters of neurons that are both visually and biologically meaningful. Finally we show the potential of *BlastNeuron* to improve the accuracy of neuron reconstruction. However, we remark that *BlastNeuron* is only a starting point for more comprehensive analyses on complex neuron morphologies. Components of *BlastNeuron* can be further improved especially when applied to incompletely reconstructed neurons, or neurons from different brain regions or even different species.

Methods

Overview of BlastNeuron

The motivation of *BlastNeuron* is similar to searching and aligning DNA sequences, especially the widely used tool BLAST (Basic Local Alignment Search Tool) (Altschul et al. 1990), and is designed to provide a framework that enables a rapid search of morphologically similar neurons in a large database population of 3D digital neuron reconstructions, as well as provide local alignment between similar neurons to pinpoint the common (i.e., "neuron motifs" or "structure motifs") or different sub-structures in their morphology. We take 3D neuron structures in the SWC format (Ascoli et al. 2001; Cannon et al. 1998) as they are readily reconstructed from light or electron microscopy data, commonly available in public neuron morphology databases such as NeuroMorpho.org (Ascoli et al. 2007), or widely supported by popular network analysis programs such as L-measure (Scorcioni et al. 2008) or NETMORPH (Koene et al. 2009). In the SWC format, a digital neuron structure is described as a tree that consists of a number of reconstruction nodes, or "compartments". Each compartment has properties including its 3D spatial coordinates, its radius in description of the thickness of the neurite at a specific 3D location, a node type indicating whether it is soma, axon or dendrite, and at most one parent compartment to which it directly connects via neuronal arbors.

BlastNeuron consists of two steps shown in Fig. 1: (i) feature-based global search for neurons with similar morphology in a neuron structure database (Section 2.2); (ii) local alignment of the neuron pairs using dynamic programming (Section 2.3).

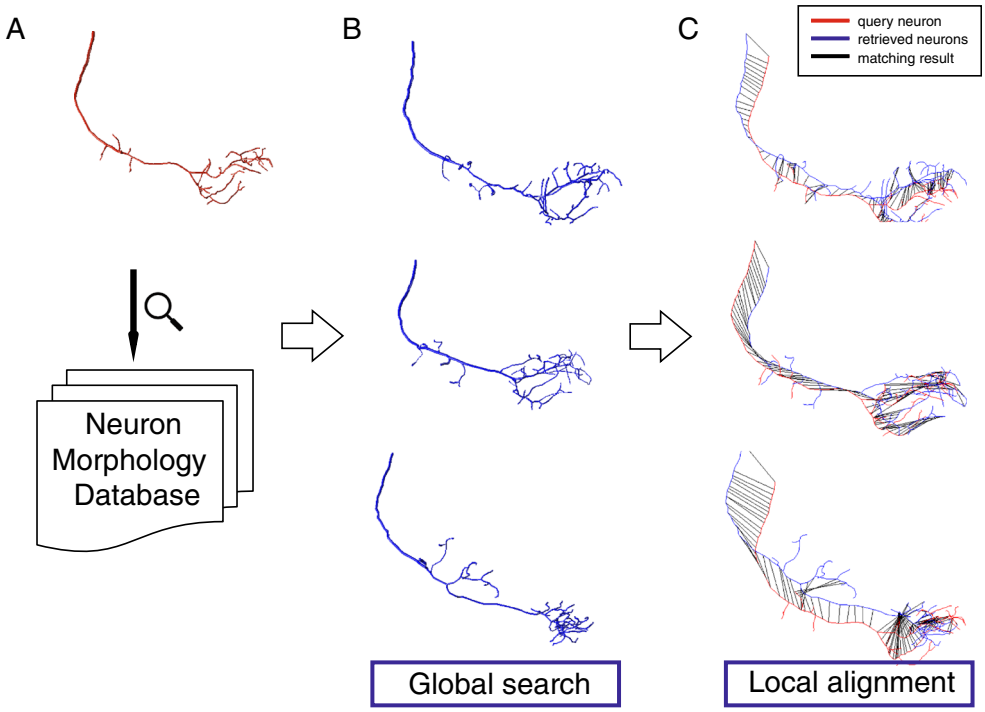

Fig. 1 Overview of *BlastNeuron*, which compares and searches a query neuron against a database of neuron morphologies. *BlastNeuron* has two steps, global search and local alignment. The global search step detects neurons with similar macro structure using global morphological features and moment invariants. The local alignment step finds correspondence of individual compartments of two neurons (or similar tree-shaped cells) by matching their topology and then geometry of the matched segments

Global Search Based on Morphological Features and Moment Invariants

The usage of feature vectors (FVs) is a standard approach for object retrieval (Bustos et al. 2005). Feature extraction from neuron structures is challenging due to the morphological heterogeneity in sample preparation, differences in imaging approaches and reconstruction methods. In order to compare neurons without reconstruction bias, the extracted features should be invariant to translation, rotation and the reconstruction's spatial sampling resolution. Importantly, it should be intuitive in revealing the overall morphology and be consistent with human vision.

The global search module is designed specifically to address the above issues. As a pre-processing step, we use three normalization procedures for the neurons before calculation of the global features. First, prune the neurons so that any small branches shorter than 5 % of the longest path in the tree structure are deleted. Second, all neuron reconstructions are resampled along the neuronal segments to achieve a consistent reconstruction rate (typically 0.5 μm). Third, all the neurons are rotated so that its dimension with the greatest span is aligned to the X-axis in the 3D space, the second-longest dimension aligned to Y-axis, and the shortest dimension aligned to Z-axis. In this way the pre-processed neurons are normalized with minimal reconstruction artifacts and remain to be consistent with human visual inspection.

In the second stage, we calculate two types of feature vectors for a neuron: (i) 21 global morphological features (Table 1), which are selected from the function list of the L-measure software (Scorcioni et al. 2008, http://cng.gmu.edu:8080/Lm/help/index.htm) and are invariant to translation and rotation in the pre-processed neurons; and (ii) 13 geometrical moments invariant (GMI) features, including four real (Hu 1962) and nine complex moments up to the third order (Lo and Don 1989). When the global search is applied to a database of neurons, the feature vectors are pre-calculated and stored in the feature database to enhance the searching speed. The similarity of two given neurons is defined as the closeness of the two neurons' feature vectors in the feature space. We use rank score normalization, in which the feature values in each dimension are sorted according to the closeness to the query and then rankings are summed up across all the dimensions. This approach

Table 1 List of global morphological features

Number of nodes	Soma surface area
Number of stems (branches on cell body)	Number of bifurcations
Number of branches	Number of tips
Neuronal height	Neuronal width
Neuronal depth	Average diameter (thickness)
Total length	Total surface area
Total volume	Maximum Euclidean distance to root
Maximum path distance to root	Maximum branch order
Average contraction	Average fragmentation
Average parent-daughter ratio	Average local amplitude angle
Average remote amplitude angle	

accounts for the scale of variability across different types of features by keeping the neurons with close feature values in every dimension. Given the number of neurons to be retrieved, *N*, *BlastNeuron* automatically returns the *N* candidate neurons that are closest to the query neuron in the morphology database.

Local Neuron Alignment Using Dynamic Programming

For globally similar neurons, we further compare their morphology locally at the compartment level. This compartment-level comparison provides a detailed view of the corresponding relationship of the building blocks (i.e., compartments or reconstruction nodes) of neurons, similar to the nucleotide or amino acid level alignment in gene or protein sequence comparison. Local neuron comparison is implemented in a two-step manner. First, a point-cloud registration algorithm is performed on the two structures regardless of topology to find the affine transform that projects one neuron (called "subject") to best align to the other (called "target"). Then, a tree-matching algorithm is designed to match the spatially registered neuron trees, taking into account both tree topology and distance in 3D space.

In the point-cloud matching step, we register the subject and the target neurons roughly in the same spatial location and orientation. To do so, the two neurons are first normalized to the same center location and scale in 3D space. Second, for each point in the two point clouds (called *A* and *B*, correspondingly) we calculate the shape context (Belongie and Malik 2000), which is defined as the Euclidean distance matrix of each point to all the other points. The similarities of shape context among points in the two neurons are then evaluated so that for each point in *A*, an optimal matching point in set *B* is found. A RANSAC sampling process (Schnabel et al. 2007) is used to estimate the optimal affine transform from *A* to *B*. The target neuron is then inversely projected to the same space as the query neuron, serving as a starting point of 3D topology matching in the next step.

In the local tree-matching step, we match the topology of two neuron trees while considering the shape and location of neuronal arbors. Tree-graph matching has been extensively studied both in theory and with various applications in evolutionary and structural biology (e.g., Bille 2005). A classical approach is based on the edit distance (Zhang and Shasha 1989) defined using "insertion", "deletion" and "substitution" operations to transform one tree to another at the minimum cost. However, most existing tree-graph matching algorithms consider only abstract tree topology, whereas in matching neuron reconstructions we need to also consider several more aspects. This approach has been adopted in neuron comparison of pyramidal cells, e.g., Heumann and Wittum (2009), which also takes into account the geometrical information as well as other neuronal features. These tree comparison algorithms would work well in classifying several neuron types based on a distance score, but few were designed or applied to the scenarios to pinpoint the local similarity of discrepancy between neuron structures. Extending the tree comparison frameworks for local comparison is thus critical.

There are several issues that must be addressed in the development of an effective local alignment algorithm. Although the neurons are pre-processed to obtain homogenous sampling resolution, the tree topology in the neurons' SWC representation does not represent the real topology of the neurons, and can vary dramatically with the level of details presented in the reconstructions. Thus, we first group the reconstruction nodes into a group "segments", each of which is a simple line-graph bounded by branching nodes, tip nodes, or the root node (a special tip point) of a neuron (Fig. 2A). A two-step matching scheme is implemented: we first treat each neuron-segment as a single component for topology matching, and then apply a more detailed alignment of the inner nodes of neuron-segments at the second step.

Secondly, due to errors in neuron reconstruction and limited ability of light-microscopic resolution to resolve fine-trained arbors, extra, broken or misplaced branches are frequently observed. For example, suppose neuron *A* has an additional branch on its main branch compared to neuron *B*. The additional branch will break *B*'s continuous segment into two parts, causing huge difference in the topological structures between the segment trees (Fig. 2A). The newly developed algorithm should account for extra/missing branch with low cost and be able to re-combine segments fragmented by the extra branch.

Finally, as we try to develop an algorithm that is consistent with human vision and brain geometry, neuron-matching has to consider the relative spatial locations of the branches in addition to abstract tree topology, with the latter being the only factor considered by existing tree-graph matching methods. Thus our definition of similarity between topologies also takes into consideration the corresponding geometry of two neurons in the standardized space.

To address these issues we design a specialized tree-matching algorithm. Let $T(a)$ be a segment-tree with the root node of a, which represents a segment in the original neuron reconstruction. Given segments a and b, in the segment-tree T, we define a as the parent of b if in the original neuron reconstruction a's tip is the parent of the root in b (see Fig. 2A). Since neurons have been mapped to the same "standard" space after the overall spatial registration, we define the weight $w(a, b)$ between two segments in two different neurons as the shortest Euclidean distance among all possible pairs of reconstruction nodes in these two segments in the normalized standard space. For two segment trees $T(a)$ and $T(b)$, which have root nodes a and b respectively, we allow multiple consecutive nodes in $T(a)$ be able to match one single node in

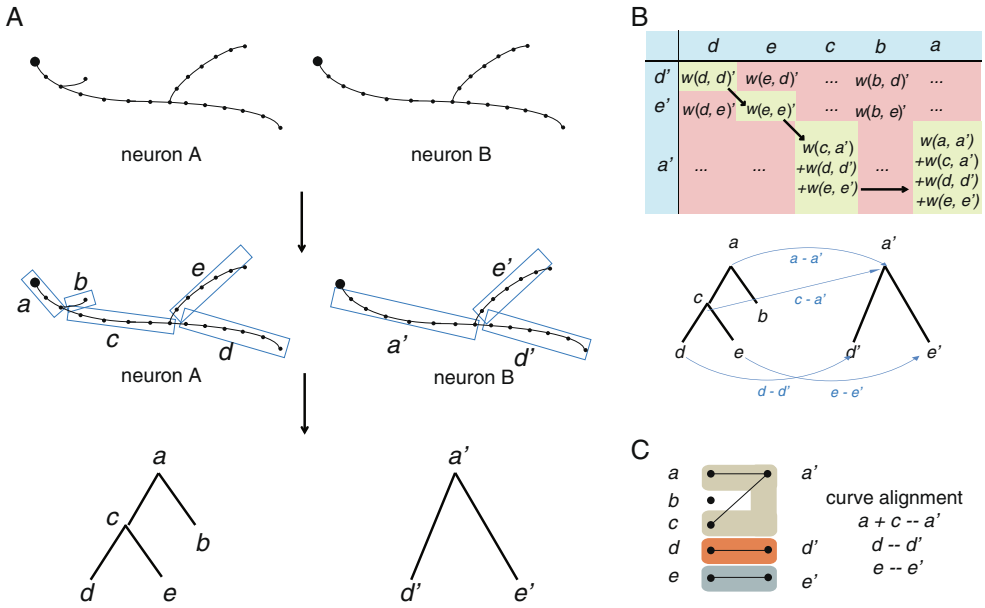

Fig. 2 Schematic illustration of local alignment algorithm. **A** Neuron reconstructions are decomposed into segment trees bounded by the branch and tip reconstruction nodes. **B** Topology matching is realized through dynamic programming. Upper table shows the distance matrix between every sub-tree, with *green* indicating optimal matching solution. The *lower panel* shows the matching result of segments. **C** Adjacent graph of matched segments: segments matched to the same segments or connected through matching relationships in the same connected component in the adjacent graph are merged into one piece for curve alignment

$T(b)$, and vice versa. In this way, the broken segments caused by additional branches in one neuron are able to match to the right corresponding part in another neuron.

We further define the distance score $W(T(a), T(b))$ of two segment-trees, with roots a and b respectively, in an iterative way, where $F(x)$ stands for the set of direct children of node x:

$$W(T(a),T(b)) = w(a,b) + \min\begin{cases} \min_{a_i \in F(a)} W(T(a_i),T(b)) \\ \min_{b_j \in F(b)} W(T(a),T(b_j)) \\ W\left(T(a)\backslash a, T(b)\backslash b\right) \end{cases} \quad (1)$$

Let R(X) to be the set of root nodes of disjoint trees in a forest X. The distance score $W(A, B)$ between two forests $A=T(a)\backslash a$ and $B=T(b)\backslash b$ in Eq. (1) can be defined as

$$W(A,B) = \min_{a_i \in R(A), b_j \in R(B)} W\left(T(a_i),T(b_j)\right) + W\left(A\backslash T(a_i), B\backslash T(b_j)\right) \quad (2)$$

Iterating Eq. (2) will eventually convert the distance between forests into the distance between two trees. In the case of binary trees, $T(a)\backslash a$ has only 2 root nodes a_1 and a_2, and $T(b)\backslash b$ has only 2 root nodes b_1 and b_2. Eq (2) can thus be written as:

$$W(\{T(a_i),T(a_2)\},\{T(b_1),T(b_2)\}) = \min\begin{cases} W(T(a_1),T(b_1)) + W(T(a_2),T(b_2)) \\ W(T(a_1),T(b_2)) + W(T(a_2),T(b_1)) \end{cases} \quad (3)$$

In our implementation, the distance score between trees is calculated in a bottom-up manner. At the leaf level, there is no subtree and thus Eqs. (1) and (3) can be computed easily. Once the distance scores are calculated at the leaf-level, the distance score between the trees at the upper level can be calculated iteratively from distance scores of their child trees (Fig. 2B).

We then build an adjacent graph from the one-to-multiple topology matching so that the corresponding long and continuous branches can map to each other. All the segments matched to the same segments or connected through matching relationships in the same connected component in the adjacent graph are merged into one piece (Fig. 2C). The merged piece does not contain any branch so it can be treated as curves. To match curves, we use the dynamic programming method in Sebastian et al. (2003).

Results

To evaluate the validity and accuracy of *BlastNeuron*, we performed three distinct tests as follows: First we applied both the global and local modules of *BlastNeuron* to annotated neuron structures in a public neuron morphology database to inspect the visual similarity as well as the biological annotations of the results. Then we used *BlastNeuron* to detect biologically meaningful clusters in a set of neurons newly reconstructed with high quality from microscopic images. Finally we showed the potential application of *BlastNeuron* to improve neuron reconstruction.

Global Search of *BlastNeuron*

We considered the entire database of NeuroMorpho.org, containing 10,004 completely or partially reconstructed neurons of different annotated "cell types" from various species. To evaluate the accuracy of *BlastNeuron*, we used 252 *Drosophila* antennal lobe and protocerebrum neurons (Jefferis et al. 2007) as queries. In such a dataset, for each

query neuron we retrieved four closest neurons (called "target") other than the query itself in the entire morphology database. Each set of four retrieval results, together with the query, were visually inspected and compared using the Vaa3D software (Peng et al. 2010). We found that for all 252 queries, most of the query neurons retrieved 3–4 morphologically similar neurons (Fig. 3). Visual comparison done by two independent people showed that 84.9 % (214/252) of the first target was correctly retrieved in terms of major branches in the anatomical structures. In addition, the 252 query neurons consist of 233 projection neurons (PNs) in the antennal lobe and 19 lateral horn neurons (LHs) in the protocerebrum. Out of all the target neurons retrieved from the 233 PN queries, 870/932 (93.3 %, p-value<1e-15) were also *Drosophila* PNs. The percentage of the same "cell type" (as indicated by the NeuroMorpho datasets themselves) for LHs was 38/76 (50.0 %, p-value<1e-15). These numbers were significantly larger than randomly choosing four candidates from 100,000+ neurons in the entire database.

We further evaluated the accuracy of *BlastNeuron* by plotting the retrieve accuracy against the number of candidates. When the 233 *Drosophila* PNs were used as query neurons, the accuracy was defined as the average percentage of PNs out of all the retrieved neurons. The accuracy decreased as more candidates were included, yet it was consistently above 39 % even if we used *BlastNeuron* to retrieve 233 candidates (Fig. 4). The rate was significantly higher than randomly selecting 233 neurons from the entire NeuroMorpho.org database (p-value<1e-15). This indicates that the *BlastNeuron* is capable of searching for globally similar neuron morphology effectively.

To quantify the robustness of the global search algorithm we also did the perturbation test using this dataset of 252 *Drosophila* neurons. We showed in Table 2 that even with a combination of 20 % spatial disturbance, and deletion of 5 % terminal branches, the global search algorithm was still robust in producing over 85 % accurate retrieval results.

Local Alignment of *BlastNeuron*

We further evaluated the local alignment module of *BlastNeuron*. In a first test, we used 30 pairs of neurons that were similar both visually and mutually supported by the

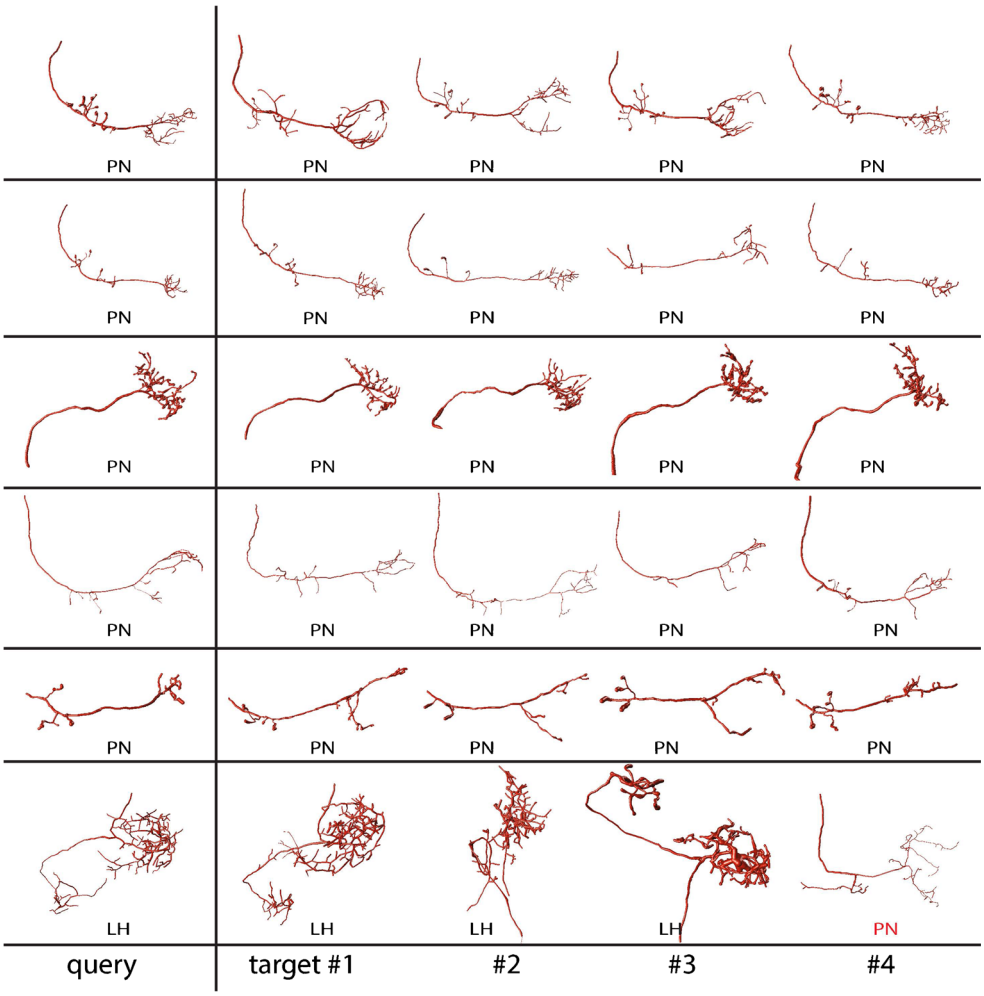

Fig. 3 Query and retrieved neurons with similar morphology using *BlastNeuron* global search. Six queries are shown on the *left*. It is apparent that retrieved neurons (called 'targets') have similar 3D global appearance despite the different orientations and sizes from the queries. The "cell types" as annotated in the database are shown under the neuron illustration (PN: projection neurons, LH: lateral horn neurons)

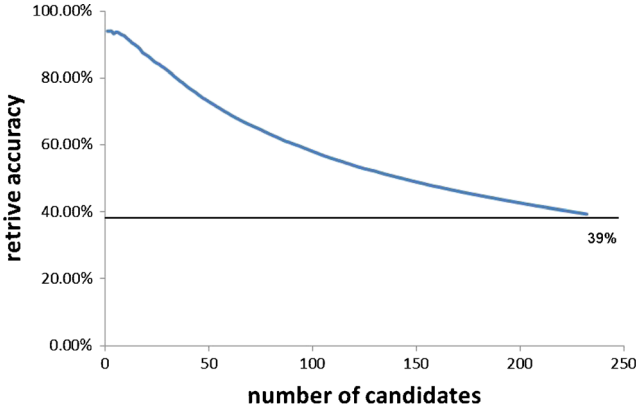

Fig. 4 Retrieval accuracy of 233 *Drosophila* projection neurons against the entire NeuroMorpho.org database

global search algorithm. We tested the local alignment algorithm on these 30 pairs (Fig. 5). The alignment between the two structures was visually meaningful to a large extend, in the sense that the major skeletons were consistently matched to each other. Indeed, the major branches of 27 out of the 30 pairs, i.e., 90 % of all the cases, were successfully matched. It was clear that some of the extra or missing branches (e.g., the top left example of Fig. 5) that clearly should be skipped in the local alignment were correctly filtered out.

To quantify the robustness of local alignment we further perturbed the neuron structures. We randomly bended a portion of the tested neuron structure by 15~60°, and locally aligned it to the original neuron structure. As we see from the matching results (Fig. 6), the majority of branches were mapped correctly even after the strong articulation of the branches.

BlastNeuron Detects Biologically Meaningful Clusters and Motifs of Neurons

With the development of high-throughput imaging techniques and automatic neuron reconstruction methods, newly created neuron morphology databases have emerged with a large number of reconstructions but limited annotations. *BlastNeuron* can be used in these situations to identify novel clusters of neurons with similar morphologies that are potentially meaningful in biology.

We applied *BlastNeuron* on a carefully reconstructed neuron dataset (Xiao and Peng 2013) that contained 681 *Drosophila* neurons traced from randomly selected raw images from the FlyCircuit database (Chiang et al. 2011). First, the 35 global features (see section 2.2) were extracted from each neuron. Then, as the number of clusters in the dataset was unknown, we hierarchically clustered this dataset based on the normalized feature space using principal component analysis to keep only 95 % variation of the data. Next, we asked two people to manually identified and reported only the tight clusters in which the neurons were morphologically very similar to each other after analyzing the dendrogram (Fig. 7A). Visual inspection of the corresponding neuron morphology, such as those shown in Fig. 7B, confirmed the validity of these clusters. Despite the complexity and variability in orientations of these neurons (e.g., cluster 1, Fig. 7B), the detected clusters shared similar morphology. Interestingly, we also found that the neurons in these tight clusters tend to share the same molecular profiles (Fig. 7A and B), as manifested by the color coded genetic (GAL4) lines Cha-Gal4, Fru-Gal4, etc. As the fly lines are generated using the mosaic analysis with a repressible cell marker (MARCM) technique at specific times during development, each Gal4 line represents a group of single neurons expressing the same gene (Chiang et al. 2011). The result seems to be consistent with neuroanatomists' expectation that the instances of the same "cell type" would share both similar morphology and molecular profiles. A detailed analysis of the relationship between such molecular profiles and morphology similarity is beyond this paper but we plan to address it at a different venue.

We then performed local alignment for the pairs of neurons that appeared in the same tight cluster. The local alignment of *BlastNeuron* was able to find the corresponding branches or sub-structure of neuron morphologies for the tightly connected neurons. Such consistent sub-structures, e.g., major branches of the matching neurons with similar projection patterns, actually can be viewed as signatures of these clusters (Fig. 7C). In computational biology, such indicative structures are often called "motifs". *BlastNeuron* as shown in this example can evidently detect the neuron morphology motifs, which could be used in future classification of neuron types and production of neuron morphology ontology.

Table 2 Robustness of the global search algorithm in *BlastNeuron* with respect to the noise level and deletion of branches

Noise Deletion	0	0.1	0.2	0.4
0	100.0 %	99.6 %	95.2 %	73.4 %
0.05	90.1 %	88.5 %	85.7 %	66.7 %
0.1	72.2 %	73.4 %	73.8 %	59.9 %
0.2	43.3 %	46.8 %	48.4 %	46.0 %

The 'noise level' is defined as the standard deviation of Gaussian noise added to the x, y, z location of each reconstruction node divided by the average adjacent reconstruction nodes' distances. The 'deletion level' is defined as the number of branches removed divided by the total number of terminal branches

Application of *BlastNeuron* in Neuron Reconstruction

We next show that *BlastNeuron* can also be used to assist automated neuron reconstruction. The key idea is that due to the various assumptions underlying different neuron reconstruction algorithms, neuron morphologies reconstructed from

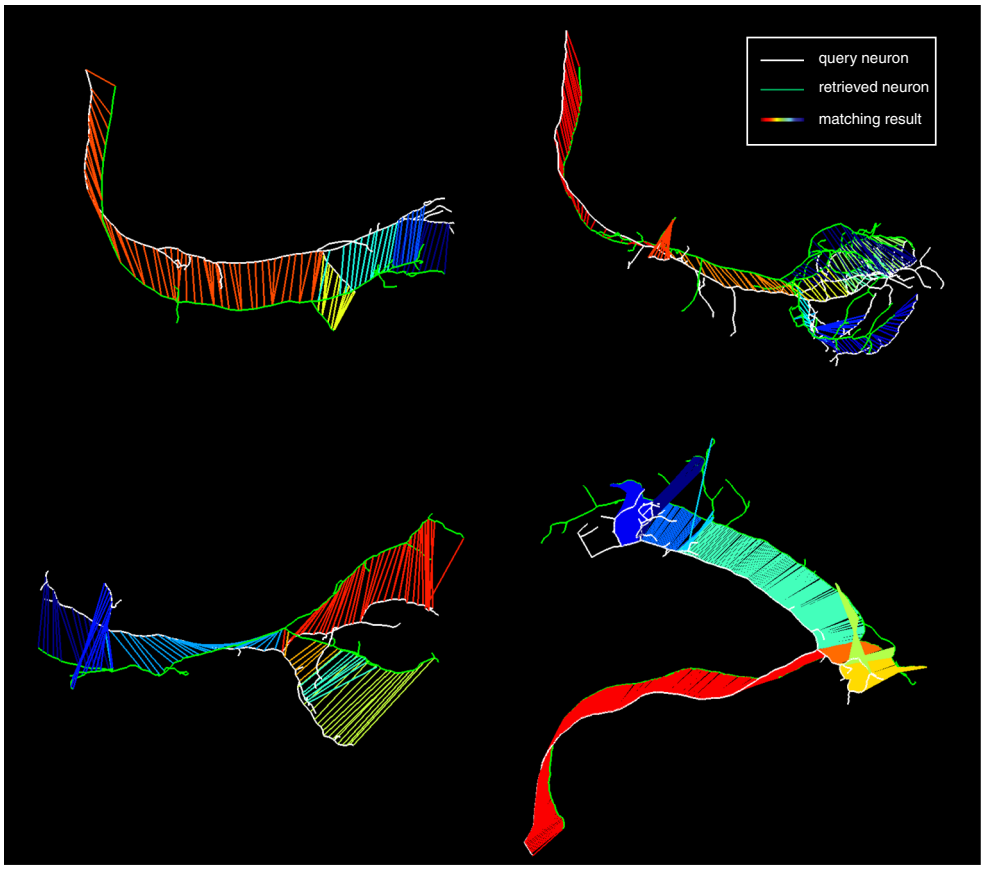

Fig. 5 Four pairs of local alignment (with each matched segment indicated by the same color in the *color coded lines*) between query (*white*) and retrieved (*green*) neurons

the same image using different algorithms could vary substantially in topology even if the overall spatial distribution of these reconstructions are similar. However, mapping out the topological difference or similarity in different neuron reconstructions is usually challenging, especially in highly complicated neurons. Here the local alignment module in

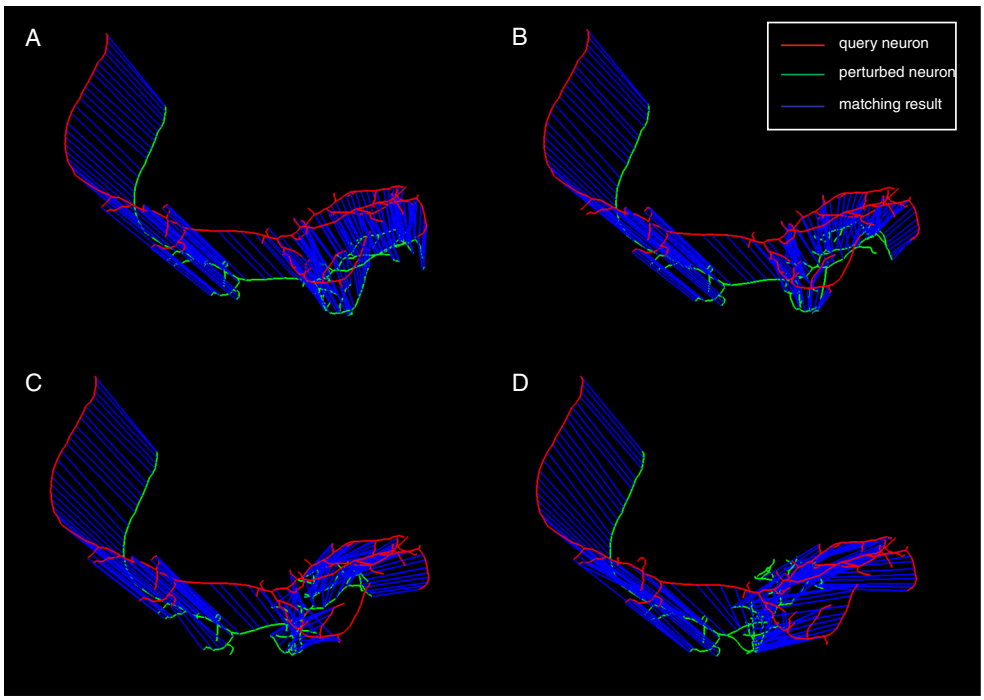

Fig. 6 Robustness test of local alignment in *BlastNeuron* by matching a query neuron against a perturbed (articulated) neuron at varying degrees. **A**~**D** for 15, 30, 45 and 60°, respectively

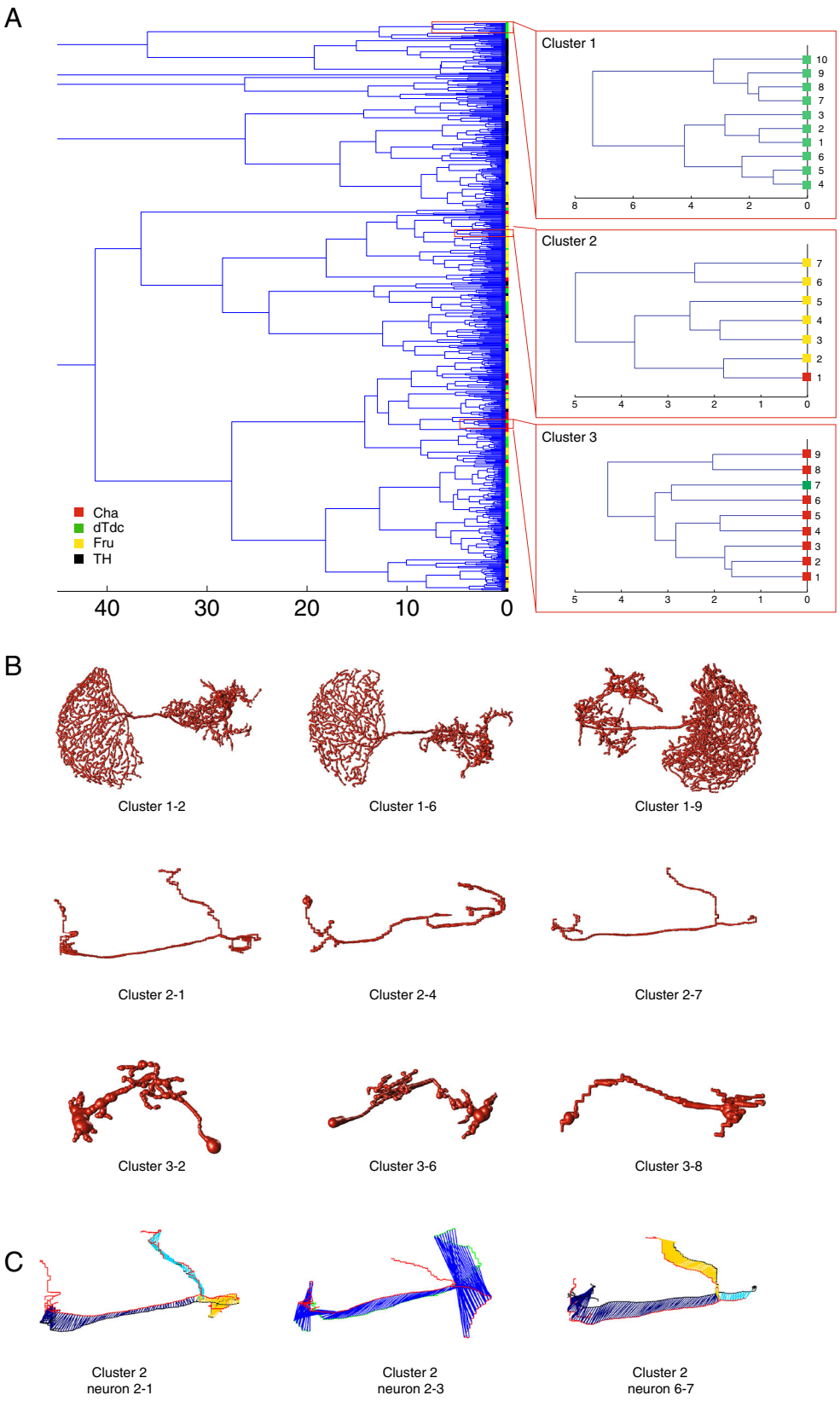

Fig. 7 Identification of neuron clusters and motifs using 681 accurately reconstructed Drosophila neurons. **A** Generation of neuron clusters. **B** Examples of the three highlighted clusters. **C** Local alignment of neurons with similar appearance shows the common neuron sub-structures, i.e., neuron motifs, illustrated as the corresponding matched segments in the figures. The corresponding branches in the common structures are marked using the same *colors*

BlastNeuron offers a powerful way to pinpoint the difference in topology of structures of same neuron reconstructed using different tracing algorithms. Such different substructures could be used for potential refinement or proofreading of these reconstructions.

As a concrete example, we considered four different neuron reconstruction algorithms developed recently: all-path-pruning 2 (APP2) (Xiao and Peng 2013), MOST ray-bursting (Ming et al. 2013), NeuTube (Zhao et al. 2011), and FARSIGHT-snake (snake) (Wang et al. 2011). These methods have different designing principles and thus capture different characteristics of a neuron. We applied these methods to the same image of a *Drosophila* neuron (Fig. 8A) and produced four different reconstructions that were globally similar to each other. Then we applied the local tree-matching alignment method in *BlastNeuron* to these four reconstructions to compare their topologies (Fig. 8B). Our algorithm was insensitive to extra branches while keeping the matching result in a consistent topological order. In the results, we found that the matching neuron-segments that also had small spatial distance indicated the common sub-structures produced by both algorithms. On the other hand, the unmatched neuron-segments or matching neuron-segments that are spatially far from each other indicate the topological difference between the two neuron reconstructions.

For simplicity of visualization, in Fig. 8C we showed the first three levels of tree branches of neuron reconstructions produced by different algorithms. When the local alignment method was used, it was evident in Fig. 8C that the major branches of the neuron structure reconstructed by all four algorithms were consistent. However, *BlastNeuron* successfully mapped out the additional branches reconstructed by MOST but not by APP2. In the case of NeuTube and snake, the topology of the reconstructed neuron structures was consistent, while the location of small branches on the main projection differed from each other, which was indicated by the match between distant points in the substructures. This utility of *BlastNeuron* is not only demonstrated for simple neuron

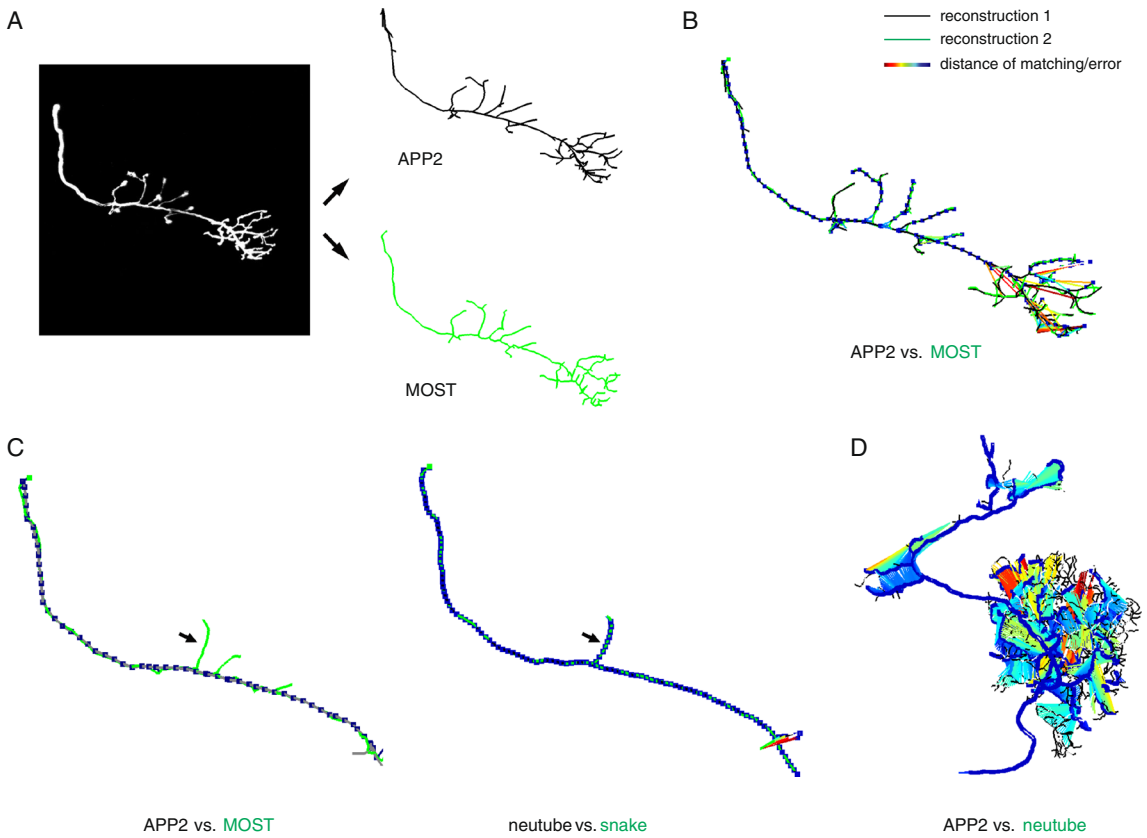

Fig. 8 Application of *BlastNeuron* in neuron reconstruction. **A** Raw image of a *Drosophila* projection neuron for reconstruction (*left*) and neuron structures reconstructed with APP2 (*upper right*) and MOST (*lower right*) algorithms. **B** Application of the local matching algorithm to the two reconstructions. The distance of the corresponding matched components was indicated by *color*, with the consistent part shown in *blue* and suspicious substructure (higher disagreement between algorithms) shown in *red*. Each matched node in the query neuron structure is indicated with a point thicker than the skeleton **C** A scheme showing the *BlastNeuron* comparing simple neuron structures, with the *arrows* indicating the mismatched compartment in APP2 vs. MOST reconstructions (*left panel*), and the same compartment consistently matched in neutube vs. snake reconstructions (*right panel*). **D** Example of complicated structures aligned with *BlastNeuron*. Substructures with a large distance need visual inspection. The distance between matched points is color-coded, with *blue* indicating zero (correct reconstruction), and *red* indicating biggest topological error

morphology like Fig. 8A and C, but also for complex neuron morphology such as the *Drosophila* neuron in Fig. 8D, which has over 300 segments and 10,000 reconstruction nodes. With *BlastNeuron*, it is efficient to map both consistent and varying sub-structures in the complex neuron morphologies. The consistent part is likely to be correct in representing the neuron's true morphology, while the inconsistent sub-structures need to be further inspected by human experts.

Discussion and Conclusions

We have designed *BlastNeuron* as a software pipeline for automatic retrieval and comparison of neuron morphology in 3D neuron reconstruction database. Using this pipeline, we show that global morphological features together with geometric moment invariants are sufficient to identify similar patterns in large databases. BlastNeuron is fast. Typically for the entire NeuroMorpho database we could run the global search within hours using a desktop computer. The local search will only be performed for neurons of very similar shapes. We have also performed a clustering analysis on the feature space to detect biologically meaningful neuron classes, regardless of the position, orientation or level of details in the reconstructions. By aligning morphologically similar neuron structures using an algorithm that considers both topology and geometry relationships, we successfully pinpointed the similar sub-structures (structure-motifs) in morphologically similar structures. Such alignment can be used in comparing neuron reconstructions.

The effectiveness of global search indicates the distinguishing power of morphological feature selection. Often, different cells tend to have different shape, level of complexity as well as projection patterns. In the shown example of *Drosophila* olfactory system, projection neurons naturally have long-range projections from the antennal lobe to mushroom body or lateral horn (Jefferis et al. 2007). Insect neurons from different species may share similar structures as compared to other mammalian neurons. By searching against the entire database with the *Drosophila* uniglomerular projection neurons, we actually found Cercal sensory interneurons in crickets also share such global features (unreported results) (Jacobs and Theunissen 2000). We plan to report such results in detail at another venue.

We also observed that in the local alignment of similar neurons, the conserved motifs or sub-structures tend to be the main branches or longest paths in the neuron structure, especially for the projection neurons (Fig. 5), which correspond to the tracts the neurons follow between the antennal lobe and higher brain centers. Because of the spatial closeness of these neurons' projection patterns, our algorithm is able to search for conserved common structures across levels of hierarchy. Complementary to our study, Gillete et al. developed an algorithm to convert neuron trees to sequences of bifurcation patterns, and thus using sequence comparison techniques to compare neuron trees (Gillette and Ascoli 2015; Gillette et al. 2015). In such a way, aligning neuron structures could be based on tree topology only. One interesting observation made by Gillete et al. was that the same type of neurite showed similar branching topology at the level of sub-trees with depth<3. These two orthogonal and independent studies provide a guideline to compare general and local topologies of neuron morphologies.

One advantage of *BlastNeuron* is that it searches for neuron structures in large databases regardless of their original context, i.e., 3D location, orientation, methods of sample preparation and imaging solutions. Thus *BlastNeuron* can be applied to correlating neuron reconstructions from different sources. In Section 3.4 we showed an application of *BlastNeuron*'s local alignment module for identifying the topological difference in neurons reconstructed with different automated tracing algorithms. The same approach may be used for searching and aligning the same neuron type from different animals, or the same neurons acquired and reconstructed with different imaging and analysis methods including both manual and automated reconstructions. For example, electron microscopy provides us reconstructions with high precision in describing the detailed neuron morphology and localizing synaptic connections with other neurons (Chklovskii et al. 2010; Denk and Horstmann 2004). By searching for the same structure in databases reconstructed using fluorescent microscopy, one may be able to quickly put such a neuron in the context of a whole brain, correct errors in a low resolution reconstruction, and develop some further experimental approaches to study the neuron's functions.

Due to the complexity in neuronal structures and ambiguity in defining "similarity" in neuron morphology, we propose *BlastNeuron* as a robust algorithm for retrieving and aligning neuron structures from large databases. However there are still enhancements that can be made in future work. For instance, in local alignment, an optimal strategy should be formulated to automatically generate a consensus result based on multiple reconstruction results. This could be done by aligning all reconstructions to the median reconstruction (Peng et al. 2011). Alternatively, a multiple-tree alignment method might be needed. Good reference algorithms can be found in the field of computer vision for generating consensus skeletons from objects of variant geometry (Zheng et al. 2010).

As one reviewer pointed out, nearly all neurons labeled and reconstructed from slice recordings have artifacts of amputated dendritic and/or axonal branches that affect the overall morphology as well as local branching. In this context, BlastNeuron could play an important role to compare and align reconstructions obtained using many different methods. Indeed we have begun to study this application in a large-scale neuroinformatics initiative BigNeuron (http://bigneuron.org, Peng et al. 2015), where we plan to use BlastNeuron as one

of the tools to compare neuron reconstructions automatically produced using 20 or so algorithms and thus identify potential regions for further correction.

It is also interesting to further investigate which global features would play a greater role in distinguishing different neuron shapes. We have found using either L-measure global features or moments features, or their variants (e.g., the ranked scores) will have comparable performance (unpublished results). However, these feature captured different properties in the data, and a thus a more robust way is to use them together as reported in this paper. In addition, for specific single neuron database such as the mouse cortical neuron database to be released by Allen Institute, we have designed a machine learning approach to select the most distinguishing features that correlate well with human annotations of these neurons. Such a piece of work can be viewed as an extension of the BlastNeuron framework.

Information Sharing Statement

The method is freely available as C/C++ code at http://home.penglab.com/proj/blastneuron.zip. To test the algorithm, SWC files of 3D neuronal morphologies are also available upon request.

Acknowledgments This work was supported primarily by the Janelia Research Campus of HHMI and the Allen Institute for Brain Science. Lei Qu was also partially supported by Chinese Natural Science Foundation Project (61201396, 61301296, 61377006, U1201255); Scientific Research Foundation for the Returned Overseas Chinese Scholars, State Education Ministry; Technology Foundation for Selected Overseas Chinese Scholar, Ministry of Personnel of China. We thank Zhi Zhou for providing some neuron reconstructions for testing in Fig. 8.

References

Altschul, S. F., et al. (1990). Basic local alignment search tool. *Journal of Molecular Biology, 215*, 403–410.

Ascoli, G. A., et al. (2001). Generation, description and storage of dendritic morphology data. *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 356*, 1131–1145.

Ascoli, G. A., Donohue, D. E., & Halavi, M. (2007). NeuroMorpho.Org: a central resource for neuronal morphologies, The. *Journal of Neuroscience, 27*, 9247–9251.

Basu, S., Condron, B., & Acton, S. T. (2011). Path2Path: Hierarchical path-based analysis for neuron matching. In *Biomedical imaging: From nano to macro, 2011 I.E. international symposium on* (pp. 996–999). IEEE.

Belongie, S., & Malik, J. (2000). Matching with shape contexts. IEEE Workshop on Content-Based Access of Image and Video Libraries. Proceedings, 20–26.

Bille, P. (2005). A survey on tree edit distance and related problems. *Theoretical Computer Science, 337*, 217–239.

Bustos, B., et al. (2005). Feature-based similarity search in 3D object databases. *ACM Computing Surveys, 37*, 345–387.

Cannon, R. C., et al. (1998). An on-line archive of reconstructed hippocampal neurons. *Journal of Neuroscience Methods, 84*, 49–54.

Cardona, A., et al. (2010). Identifying neuronal lineages of Drosophila by sequence analysis of axon tracts. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 30*, 7538–7553.

Chiang, A. S., Lin, C. Y., Chuang, C. C., Chang, H. M., Hsieh, C. H., Yeh, C. W., & Hwang, J. K. (2011). Three-dimensional reconstruction of brain-wide wiring networks in Drosophila at single-cell resolution. *Current Biology, 21*(1), 1–11.

Chklovskii, D. B., Vitaladevuni, S., & Scheffer, L. K. (2010). Semi-automated reconstruction of neural circuits using electron microscopy. *Current Opinion in Neurobiology, 20*, 667–675.

Costa, M., et al. (2014). NBLAST: Rapid, sensitive comparison of neuronal structure and construction of neuron family databases. bioRxiv, 006346.

Denk, W., & Horstmann, H. (2004). Serial block-face scanning electron microscopy to reconstruct three-dimensional tissue nanostructure. *PLoS Biology, 2*, e329.

Dumitriu, D., Cossart, R., Huang, J., & Yuste, R. (2007). Correlation between axonal morphologies and synaptic input kinetics of interneurons from mouse visual cortex. *Cerebral Cortex, 17*(1), 81–91.

Ganglberger, F., et al. (2014). Structure-based neuron retrieval across Drosophila brains. *Neuroinformatics, 12*, 423–434.

Gillette, T.A., & Ascoli, G.A. (2015). Topological characterization of neuronal arbor morphology via sequence representation. I. Motif analysis (in press).

Gillette, T.A., Hosseini, P., & Ascoli, G.A. (2015). Topological characterization of neuronal arbor morphology via sequence representation. II. Global alignment (in press).

Heumann, H., & Wittum, G. (2009). The tree-edit-distance, a measure for quantifying neuronal morphology. *Neuroinformatics, 7*(3), 179–190.

Hu, M. (1962). Visual-pattern recognition by moment invariants. *IRE Transactions on Information Theory 8*, 179-&.

Jacobs, G. A., & Theunissen, F. E. (2000). Extraction of sensory parameters from a neural map by primary sensory interneurons. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 20*, 2934–2943.

Jefferis, G. S., et al. (2007). Comprehensive maps of Drosophila higher olfactory centers: spatially segregated fruit and pheromone representation. *Cell, 128*, 1187–1203.

Koene, R. A., et al. (2009). NETMORPH: a framework for the stochastic generation of large scale neuronal networks with realistic neuron morphologies. *Neuroinformatics, 7*, 195–210.

Lo, C. H., & Don, H. S. (1989). 3-D moment forms - their construction and application to object identification and positioning. *IEEE Transactions on Pattern Analysis, 11*, 1053–1064.

Mayerich, D., et al. (2012). NetMets: software for quantifying and visualizing errors in biological network segmentation. *BMC Bioinformatics, 13*(Suppl 8), S7.

Ming, X., et al. (2013). Rapid reconstruction of 3D neuronal morphology from light microscopy images with augmented rayburst sampling. *PLoS One, 8*, e84557.

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

Peng, H., Chung, P., Long, F., Qu, L., Jenett, A., Seeds, A. M., & Simpson, J. H. (2011). BrainAligner: 3D registration atlases of Drosophila brains. *Nature Methods, 8*(6), 493–498.

Peng, H., Roysam, B., & Ascoli, G. A. (2013). Automated image computing reshapes computational neuroscience. *BMC Bioinformatics, 14*(1), 293.

Peng, H., Bria, A., Zhou, Z., Iannello, G., & Long, F. (2014). Extensible visualization and analysis for multidimensional images using Vaa3D. *Nature Protocols, 9*(1), 193–208.

Peng, H., Meijering, E., & Ascoli, G. (2015) From DIADEM to BigNeuron. *NeuroInformatics*. doi:10.1007/s12021-015-9270-9.

Schnabel, R., Wahl, R., & Klein, R. (2007). Efficient RANSAC for point-cloud shape detection. In *Computer graphics forum* (26(2), 214–226). Blackwell Publishing Ltd.

Scorcioni, R., Polavaram, S., & Ascoli, G. A. (2008). L-Measure: a web-accessible tool for the analysis, comparison and search of digital reconstructions of neuronal morphologies. *Nature Protocols, 3*(5), 866–876.

Sebastian, T. B., Klein, P. N., & Kimia, B. B. (2003). On aligning curves. *IEEE Transactions on Pattern Analysis, 25*, 116–125.

Tschirren, J., et al. (2005). Matching and anatomical labeling of human airway tree. *IEEE Transactions on Medical Imaging, 24*, 1540–1547.

Wang, Y., et al. (2011). A broadly applicable 3-D neuron tracing method based on open-curve snake. *Neuroinformatics, 9*, 193–217.

Xiao, H., & Peng, H. (2013). APP2: automatic tracing of 3D neuron morphology based on hierarchical pruning of a gray-weighted image distance-tree. *Bioinformatics, 29*(11), 1448–1454.

Zhang, K., & Shasha, D. (1989). Simple fast algorithms for the editing distance between trees and related problems. *SIAM Journal on Computing, 18*(6), 1245–1262.

Zhao, T., Xie, J., Amat, F., Clack, N., Ahammad, P., Peng, H., & Myers, E. (2011). Automated reconstruction of neuronal morphology based on local geometrical and global structural models. *Neuroinformatics, 9*(2–3), 247–261.

Zheng, Q., et al. (2010). Consensus skeleton for non-rigid space-time registration. *Computer Graphics Forum, 29*, 635–644.