linus: Conveniently explore, share and present large-scale biological trajectory data from a web browser.

Authors:

Johannes Waschke [1,2], Mario Hlawitschka [2], Kerim Anlas [3], Vikas Trivedi [3,4], Ingo Roeder [5,6], Jan Huisken [7], and Nico Scherf [1,6*]

[1] Max Planck Institute for Human Cognitive and Brain Sciences, Stephanstr. 1a, 04103 Leipzig, Germany

[2] Faculty of Computer Science and Media, Leipzig University of Applied Sciences, 04277 Leipzig, Germany

[3] EMBL Barcelona, C/ Dr. Aiguader 88, 08003 Barcelona, Spain.

[4] EMBL Heidelberg, Developmental Biology Unit, 69117 Heidelberg, Germany.

[5] National Center of Tumor Diseases (NCT), Partner Site Dresden, 01307 Dresden, Germany

[6] Institute for Medical Informatics and Biometry, Carl Gustav Carus Faculty of Medicine, School of Medicine, TU Dresden, 01307 Dresden, Germany

[7] Morgridge Institute for Research, Madison, Wisconsin 53715, USA

* Correspondence: to nico.scherf@tu-dresden.de

Abstract

In biology we encounter large-scale trajectory data, but exploring and communicating patterns in such data is often a cumbersome task. Ideally, such data should be provided with an interactive visualisation in one concise package so that we can create and test hypotheses collaboratively. To address these challenges, we have developed a Python-based tool, linus, which makes the process of exploring and sharing 3D trajectories as easy as browsing a website. linus creates intriguing and interactive visualisations that are shareable, both offline and online, to facilitate the collaborative discovery of patterns in large and complex trajectory data.

In biology, we often face large-scale trajectory data such as dense spatial pathways (e.g. brain connectivity from diffusion MRI imaging) or tracking data (e.g. cell trajectories or animal trails). Although this type of data is becoming more and more prominent in biomedical research [1–5], exploring, sharing and communicating patterns in such data is often a cumbersome task requiring a set of software that is complex to use or even to install. Recently, new tools have become available for visualising 3D volumetric data[6–8], and some allow the user to overlay tracking results on images to cross-check the quality of the results or to visualise defined patterns. However, such multi-purpose software generally lacks the means to explore, present or share the visualisations interactively.

Exploratory analysis and interactive visualisation [9] of complex and large-scale trajectory data are needed to create and test hypotheses collaboratively. An interactive, scriptable and shareable visualisation of findings would open up novel ways of communicating and discussing experimental results and findings. Importantly, since such bioinformatics tools exist at the interface of computational and life sciences, they need to be accessible and usable for scientists with little or no background in programming. Ideally, the data should be bundled with its guided, interactive visualisation in one concise package. To address these challenges, we have developed a python-based tool, *linus*, which renders the complicated process of exploring and sharing 3D trajectories nearly as easy as opening a website, without any local installation. *linus* creates interactive visualisations, while keeping data presentation simple and shareable, both offline and online (Fig 1a). We have exemplarily demonstrated this concept using a preliminary version in [10].

Creating a visualisation package with *linus* is a simple workflow (Fig. 1a): The user imports trajectory data from a generic, plain CSV format (described in the Supplement) or from a variety of established trajectory formats such as SVF [2], TGMM XML [11] and the community standard biotracks [12] (which itself supports import from a wide variety of cell tracking tools like CellProfiler [13] or TrackMate [14]). During the conversion, *linus* can enrich the trajectory data with additional attributes or spatial context. For example, we can declutter dense trajectories by highlighting the major "highways" (Fig.1 b,c). The result is a ready-to-use software package that can be opened in a web browser on almost any device with WebGL support. The amount of data that can be fluently visualised depends on the underlying hardware (smartphones: 2,000+ trajectories, notebooks and desktop computers: 10,000+ trajectories).

The browser renders an interactive visualisation of the trajectories, as well as an interface for the user to update and adapt the visualisation to their needs (e.g. colour scales, projections, clipping planes). (Fig. 1b). One major feature is the ability to blend between different data transformations seamlessly (e.g. original vs bundled to focus on major "highways", original vs 2D projections to provide a global perspective on the trajectories, see Fig. 1c). The user can interactively filter the displayed data for generic properties such as space and time or and for user-specified numerical properties such as marker expression in cell tracking. The user can define spatial regions of interest (ROIs) and iteratively refine them in order to highlight trajectories of interest and to download them as CSV files for subsequent analysis.

The user can conveniently record all these properties of the visualisation and create interactive tours as well as pre-rendered videos (e.g. such as Supplementary Video 1) and screenshots. In order to communicate and distribute new findings, these tours can easily be shared online or offline with the community (colleagues, readers of a manuscript, audience of a real or virtual presentation), see Fig. 1d-f for examples. This interactive data sharing is particularly helpful when collaborations, presentations or teaching are taking place remotely. The target audience can explore the data instantly on their computers, tablets, or smartphones. Touch screens or virtual reality goggles (e.g. Oculus Go), increase the immersion with more natural controls and with true 3D-rendering, helping to grasp the spatial relation of the trajectories. With these features, *linus* will improve the way we collectively explore, communicate and teach the spatio-temporal patterns from experimental data.

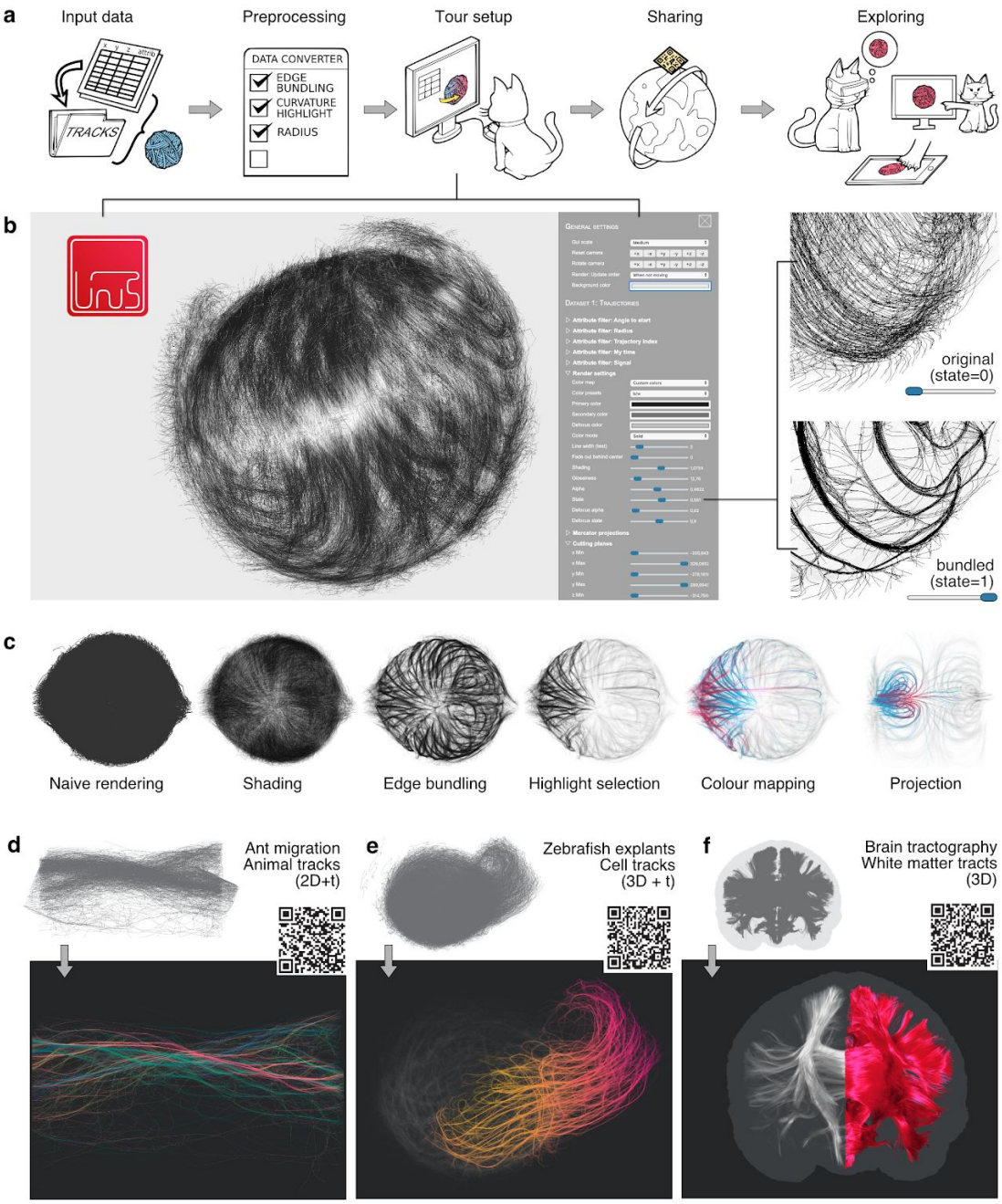

Figure 1 Browser-based exploration and sharing of trajectory visualizations with *linus*. (a) shows the control workflow of our tool. Starting with the data, a Python-converter is used to enrich the data with further features (e.g. numeric metrics, an edge-bundled version of the data, visual context) and to prepare the visualisation package. (b) Within minutes, the data can be visualised and explored in the browser, and different aspects of the data can be interactively highlighted (example shows the effect of changing the degree of trajectory bundling). (c) The user can choose from a range of several visualisation methods and - together with annotations and camera motion paths - combine them in a scheduled tour that can be shared by a custom URL or QR code (as in (d)-(f)) generated directly in the browser interface (b). (d)-(f) demonstrate use cases for real-world datasets with different characteristics and dimensionality. (d) Ant trails (2D+t) from [15]. Bundling and colour-coding (spatial orientation by mapping (x,y,z) to (R,G,B) values) indicate the major trails running in opposing directions. (e) Cell movements during the elongation process of zebrafish blastoderm explants (3D+t) [16]. Bundling, colour coding, and spatial selection highlight collective cell movements as the explant starts elongating, focusing on a subpopulation of cells driving this process. Colour code shows time from early (yellow) to late (red) for selected tracks. (f) Brain tractography data showing major white matter connectivity from diffusion MRI (3D). The spatial selection highlights the left hemisphere while anatomical context is provided by the outline of the entire brain (from mesh data) and the defocused tracts of the right hemisphere.

Data availability

You can try out exemplary visualizations by scanning the QR codes in Fig.1. directly or by visiting https://imb-dev.gitlab.io/linus-manuscript/

Code availability

The *linus* software including source code and documentation is freely available at our repository at https://gitlab.com/imb-dev/linus.

Acknowledgments

The authors are grateful to Gopi Shah and Konstantin Thierbach for sharing data and contributing useful feedback. J.W. received funding from the International Max Planck Research School on Neuroscience of Communication: Function, Structure, and Plasticity (Leipzig, Germany; https://imprs-neurocom.mpg.de).

Author contributions

N.S., J.H., and I.R. conceived the project. J.W. wrote the software code. M.H. and N.S. supervised the project. N.S. and J.W. wrote the manuscript. K.A. and V.T. generated the dataset on zebrafish blastoderm explants. All authors read, edited, and approved the manuscript.

References

1. Wallingford, J. B. The 200-year effort to see the embryo. *Science* **365**, 758–759 (2019).
2. McDole, K. *et al.* In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. *Cell* **0**, (2018).
3. Kwok, R. Deep learning powers a motion-tracking revolution. *Nature* **574**, 137–138 (2019).
4. Romero-Ferrero, F., Bergomi, M. G., Hinz, R., Heras, F. J. H. & de Polavieja, G. G. idtracker.ai: Tracking all individuals in large collectives of unmarked animals. *arXiv [cs.CV]* (2018).
5. Liu, C. *et al.* A resource for the detailed 3D mapping of white matter pathways in the marmoset brain. *Nat. Neurosci.* **23**, 271–280 (2020).
6. Pietzsch, T., Saalfeld, S., Preibisch, S. & Tomancak, P. BigDataViewer: visualization and processing for large image data sets. *Nat. Methods* **12**, 481–483 (2015).
7. Royer, L. A. *et al.* ClearVolume: open-source live 3D visualization for light-sheet microscopy. *Nat. Methods* **12**, 480–481 (2015).
8. Schmid, B. *et al.* 3Dscript: animating 3D/4D microscopy data using a natural-language-based syntax. *Nat. Methods* **16**, 278–280 (2019).
9. Shneiderman, B. The eyes have it: a task by data type taxonomy for information visualizations. in *Proceedings 1996 IEEE Symposium on Visual Languages* 336–343 (1996).
10. Shah, G. *et al.* Multi-scale imaging and analysis identify pan-embryo cell dynamics of germlayer formation in zebrafish. *Nat. Commun.* **10**, 5753 (2019).
11. Amat, F. *et al.* Fast, accurate reconstruction of cell lineages from large-scale fluorescence microscopy data. *Nat. Methods* **11**, 951–958 (2014).
12. Gonzalez-Beltran, A. N. *et al.* Community Standards for Open Cell Migration Data. *bioRxiv* 803064 (2019) doi:10.1101/803064.
13. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, e2005970 (2018).
14. Tinevez, J.-Y. *et al.* TrackMate: An open and extensible platform for single-particle tracking. *Methods* (2016) doi:10.1016/j.ymeth.2016.09.016.
15. Imirzian, N. *et al.* Automated tracking and analysis of ant trajectories shows variation in forager exploration. *Sci. Rep.* **9**, 13246 (2019).
16. Trivedi, V. *et al.* Self-organised symmetry breaking in zebrafish reveals feedback from morphogenesis to pattern formation. *bioRxiv* 769257 (2019) doi:10.1101/769257.