1
2
3
4 **Title**

5 ethoscopy & ethoscope-lab: a framework for behavioural analysis

6 to lower entrance barrier and aid reproducibility

7
8
9
10

11 **Authors and affiliations**

12 Laurence Blackhurst[1#] and Giorgio F. Gilestro[1#]

13

14 *# To whom correspondence should be addressed.*

15 *E-mail:* l.blackhurst19@imperial.ac.uk, *giorgio@gilest.ro*

16

17 [1]Department of Life Sciences, Imperial College London, London, UK

Abstract

Summary

High-throughput analysis of behaviour is a pivotal instrument in modern neuroscience, allowing researchers to combine modern genetics breakthrough to unbiased, objective, reproducible experimental approaches. To this extent, we recently created an open-source hardware platform (ethoscope[1]) that allows for inexpensive, accessible, high-throughput analysis of behaviour in *Drosophila* or other animal models. Here we equip ethoscopes with a Python framework for data analysis, ethoscopy, designed to be a user-friendly yet powerful platform, meeting the requirements of researchers with limited coding expertise as well as experienced datascientists. Ethoscopy is best consumed in a prebaked Jupyter-based docker container, ethoscope-lab, to improve accessibility and to encourage the use of notebooks as a natural platform to share post-publication data analysis.

Availability and implementation

Ethoscopy is a Python package available on GitHub and PyPi. Ethoscope-lab is a docker container available on DockerHub. A landing page aggregating all the code and documentation is available at https://lab.gilest.ro/ethoscopy.

Introduction

The use of open-source software in science is commended by researchers and funders alike, for it frees academic science from ties with possibly ephemeral third-parties; it creates a bottom-up environment in which researchers actively collaborate to build their own solutions based on their own demands; it maximises the value of public funding; it provides the transparency that is essential for reproducibility. At the same time, however, academic open-source software still encounters barriers in terms of ease of use, compatibility with multiple operating systems and constantly evolving software ecosystems, as well as often unmet good practices in post-publication data sharing. We recently created an open-source, well documented platform for high-throughput behavioural analysis in *Drosophila* that has been widely adopted by the community (ethoscope[1], https://lab.gilest.ro/ethoscope) and we initially equipped it with a powerful and versatile selection of R packages (rethomics[2]). However, while R remains a language of first-choice for data scientists, its structure, syntax and philosophy can prove unwelcoming

46   to newcomers and ultimately discourage adoption in the field. Here we describe ethoscopy and

47   ethoscope-lab, two instruments whose genesis was motivated by our very experience in the past years.

48   These tools are not simply the porting of rethomics into a more accessible programming language: in

49   fact, they offer new functions and tools, integrating with the modern sleep and circadian literature and

50   at the same time adopt the best practice in code accessibility and data sharing. Here below we detail

51   how.

Improved accessibility for the community at large

53   Software accessibility is a key aspect of bioinformatics, as even the best tools are useless if they are not

54   adopted by the community. One way to improve accessibility is for authors to distribute Docker

55   containers[3,4]. Docker containers increase adoption because they guarantee the tool will always ship

56   with the needed dependencies; they are truly multiplatform as they offer the same experience to any

57   user irrespective of their operating system; they guarantee that all researchers will operate on the same

58   version of code and underlying libraries, ultimately increasing reproducibility. Ethoscope-lab is a

59   docker container distributed via the official DockerHub platform and providing at any time access to

60   the latest versions of ethoscopy and rethomics in a Jupyter instance.

Improved accessibility for users within the laboratory

62   Ethoscope-lab ships with a stable release of JupyterHub, the multiuser web-based notebook that rapidly

63   became the tool of choice for most data scientists[5]. The typical scenario will be for a laboratory to

64   deploy ethoscope-lab on a shared powerful workstation, accessible to all laboratory members via the

65   intranet or the internet. Ideally, the same workstation will also store the raw data of the behavioural

66   analysis, as extracted directly by the ethoscopes, so that ethoscopy can have local, read-only access to

67   them. Users will connect to Jupyter-hub using their favourite web-browser on a computer or tablet,

68   login using their own credentials, and work in the cloud, saving their notebooks, stats and figures on a

69   remote folder on the workstation that can be backup-ed up and shared backed with themselves or other

70   users (for instance, via Samba or the open source software Owncloud). This scenario offers multiple

71   advantages: all the users in the laboratory access ethoscopy as a SaaS (Software as a Service) which is

72   the best way to improve accessibility to end users; the docker container and the workstation can be

73   setup and maintained by one tech-savvy user or by the IT department, allowing all other end users to

74   concentrate on the data analysis, not the tool maintenance; the only machine requiring powerful

75 computing specifications will be the workstation hosting the docker container so that users will be able
76 to perform big-data analysis from any device and any location, saving time and resources.

Good practice in post-publication data sharing

78 At the time of publication, researchers will share information on which version of the docker container
79 they used for their analysis and upload fully annotated Jupyter notebooks as supplementary data. We
80 provide several example notebooks that can serve as step-by-step tutorials on the most common
81 functions of ethoscopy and, at the same time, offer direct evidence of why a Jupyter notebook is
82 considered the best instrument to share post-publication data processing in research. A reader that has
83 access to: 1) the raw behavioural data, 2) the metadata describing the experimental conditions, 3) a well
84 documented Jupyter notebook and 4) the matching version of ethoscopelab, will be able to reproduce
85 and re-analyse and figure of any publication, with no risk of obsolescence even after decades. All these
86 can be shared on the journal platform or via third party scholarly services, such as Zenodo.

Ethoscopy features

88 Similarly to rethomics, ethoscopy will work out-of-the-box on ethoscope and DAMs data[6] but can in
89 principle process any behavioural data as long as they are appropriately formatted. Ethoscopy
90 capabilities are mostly geared towards animals' activity, with a particular focus on circadian biology
91 and sleep, including a newly introduced analysis of sleep-depth that uses a hidden Markov model to
92 identify sleep stages in *Drosophila*[7]. Figures are created using the open-source graphing library plotly[8].
93 Plotly output is interactive by default: users can select and zoom-in on smaller windows of data and can
94 hide or show individual datasets with a click of the mouse, for instance hiding genotypes from a
95 complex comparison. Ethoscopy has a focus on sleep and circadian analysis and methods for the most
96 common operations are embedded in the code, so that novice users will be able to perform the most
97 common analysis of sleep using only a single line of code. However, any bespoke analysis will still be
98 easy to adopt given that ehoscopy objects build on the Pandas dataframe, so experienced users will
99 have access to the entire rich toolbox available to all Python datascientists.

Proof of principle

101 As proof of principle of ethoscopy capabilities, we show analyses of sleep (Figure 1a,b) and sleep-
102 depth (Figure 1c) for a group of wild-type Canton S flies whose activity was recorded with ethoscopes
103 for several days. Wiggin *et al* first proposed an interesting computational paradigm to identify *bona*

104   *fide* covert sleep stages using Hidden Markov Chains (HMC)[7]. Given that their initial analyses were

105   mostly based on data collected using *Drosophila* activity monitors (thus offering only limited

106   resolution of activity[9]), we ought to explore how HMC would perform on activity data collected at

107   higher resolution, employing ethoscope video tracking. While we found that the initial sleep

108   characterisation in four hidden sleep stages generally holds true (Figure 1c), we also identified a

109   potentially interesting difference: using ethoscopes, we could detect deep-sleep not throughout the

110   night as originally reported[7], but almost exclusively in the first half of the night, with a clear prominent

111   peak at ZT15 (Figure 1c). Interestingly, this specific window of time was previously shown to be the

112   one associated to the highest arousal threshold using an essay of odour perception[10], nicely matching

113   ethoscopy's description of sleep stages.

114

115   **Conclusions**

116   We built ethoscopy and ethoscope-lab to further improve accessibility and reproducibility in the field of

117   *Drosophila* sleep. Besides offering state-of-the-art data analysis, we envision these tools will open new

118   doors to behavioural scientists and introduce them to good sharing practice in terms of code

119   accessibility and reproducibility.

120

126

127   **Code and documentation.**

128   Code, documentation, and tutorials are available at https://lab.gilest.ro/ethoscopy

129

130   **Reference**

131

1. Geissmann, Q. *et al.* Ethoscopes: An Open Platform For High-Throughput Ethomics. *PLOS Biol.* **15**, e2003026 (2017).

2. Geissmann, Q., Rodriguez, L. G., Beckwith, E. J. & Gilestro, G. F. Rethomics: An R framework to analyse high-throughput behavioural data. *PLOS ONE* **14**, e0209331 (2019).

3. Boettiger, C. An introduction to Docker for reproducible research. *ACM SIGOPS Oper. Syst. Rev.* **49**, 71–79 (2015).

4. Nüst, D. *et al.* Ten simple rules for writing Dockerfiles for reproducible data science. *PLOS Comput. Biol.* **16**, e1008316 (2020).

5. Why Jupyter is data scientists' computational notebook of choice. https://www.nature.com/articles/d41586-018-07196-1.

6. Rosato, E. & Kyriacou, C. P. Analysis of locomotor activity rhythms in Drosophila. *Nat. Protoc.* **1**, 559–568 (2006).

7. Wiggin, T. D. *et al.* Covert sleep-related biological processes are revealed by probabilistic analysis in Drosophila. *Proc. Natl. Acad. Sci.* (2020) doi:10.1073/pnas.1917573117.

8. Sievert, C. *Interactive Web-Based Data Visualization with R, plotly, and shiny*. (Chapman and Hall/CRC, 2020). doi:10.1201/9780429447273.

9. Gilestro, G. F. Video tracking and analysis of sleep in Drosophila melanogaster. *Nat. Protoc.* **7**, 995–1007 (2012).

10. French, A. S., Geissmann, Q., Beckwith, E. J. & Gilestro, G. F. Sensory processing during sleep in Drosophila melanogaster. *Nature* **598**, 479–482 (2021).

132

Figure legend.

Figure 1 – Ethoscopy analysis of sleep and sleep depth. a) Overview of sleep and activity using the heatmap plotting method of ethoscopy. Here we show the overall activity of a group of wild type flies in control conditions (upper panel) or during a sleep deprivation experiment (lower panel). The blue to yellow color gradient quantifies activity (blue) and sleep (yellow). **b)** 24h profile of baseline sleep of

Blackhurst *et al.*   Page 6

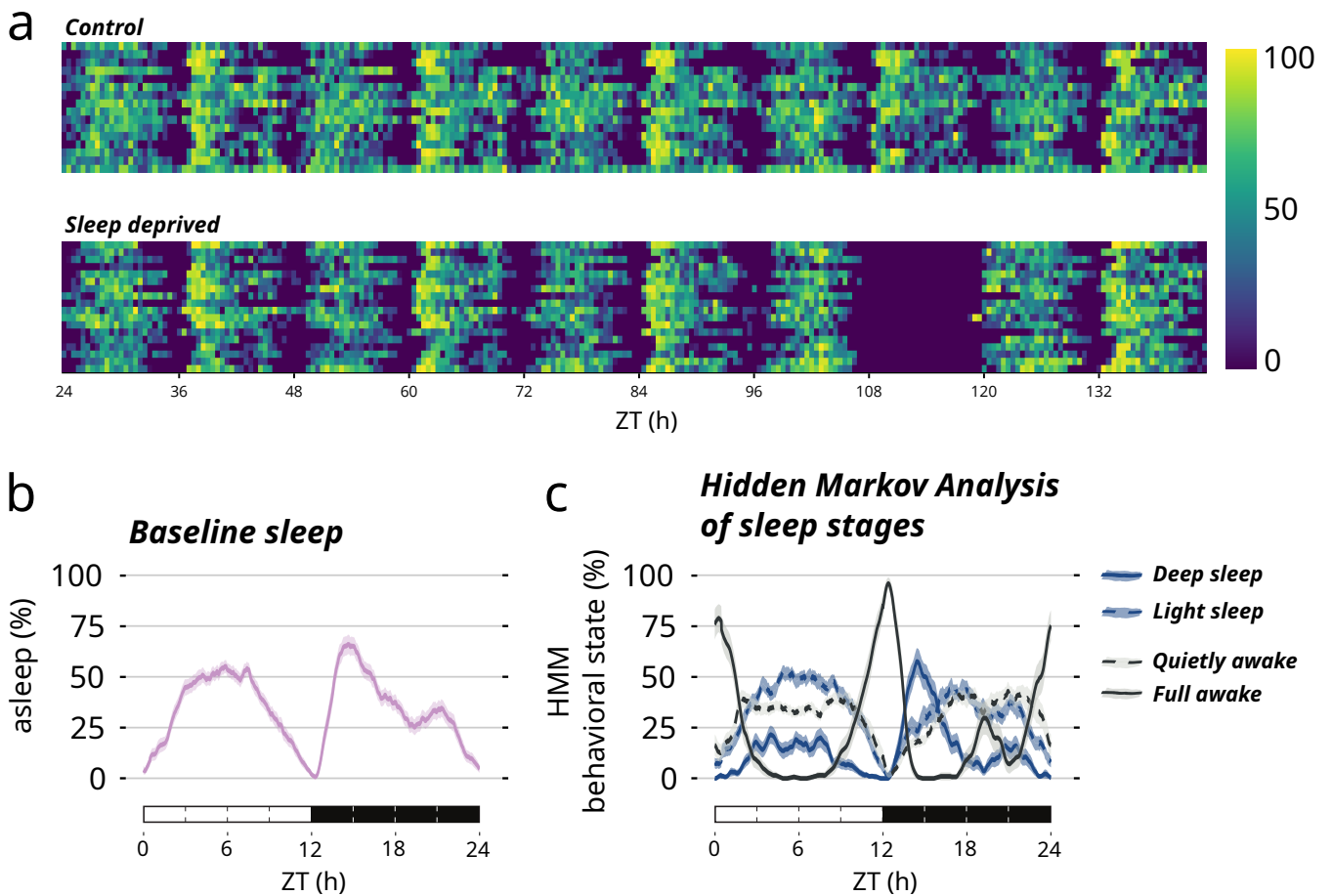137   the control flies shown in b. **c)** Hidden Markov Chains modelling of four covert sleep stages. ZT:

138   Zeitgeber.

Figure 1

Blackhurst *et al.* 2022