OXFORD

Subject Section

Rethomics: an R framework to analyse high-throughput behavioural data

Quentin Geissmann [1,*]**, Luis Garcia** [2]**, Esteban Beckwith** [1]**, and Giorgio Gilestro** [1,*]

[1] Department, Institution, City, Post Code, Country and
[2] Department, Institution, City, Post Code, Country.

[*] To whom correspondence should be addressed.

Associate Editor: XXXXXXX

Abstract

Motivation: Ethomics, a quantitative and high-throughput approach to ethology, is a novel and exciting field. The recent development of methods that automatically score variables in multiple animals provides an unprecedented insight into the study of behaviours. The analysis of ethomics data presents many challenges that are conceptually independent of the acquisition platform. However, there is, little effort in providing a general framework to analyse multiple and long behavioural time series. We developed the `rethomics` framework, a suite of R packages that altogether provide utilities to: import, store, visualise and analyse behavioural data.

Availability and Implementation: All packages in the `rethomics` framework are available under the terms of the GPLv3 license. Exaustive installation instructions and tutorials are available at https://github.com/rethomics/ https://rethomics.github.io/.

Contact: qgeissmann@gmail.com

1 Introduction

Animal behaviours are complex phenotypes that are manifestations of the interaction between the nervous system and the external or internal environment. In the last few decades, our ability to record large amounts of various phenotypical data has tremendously increased. Behaviour scoring is certainly not an exception to this trend. Indeed, many platforms (todo citations) have been developed in order to allow biologists to continuously record behaviours such as activity, position and feeding of multiple animals over long durations (days or weeks).

The availability of ever so large amounts of data is very exciting as it paves the way for refined analyses. Clearly, the multiplicity of model organisms, hypotheses and paradigms makes the diversity of recording tools crucial. However, regarding subsequent data analysis, there is no unified, programmatic, framework that could be used as a set of building blocks in a pipeline. Instead, tools tend to consist of graphical interfaces with strictly defined functionalities that only import data from a single platform. There are three issues with this approach. First of all, state of the art analysis and visualisation requires a flexibility than only a programmatic interface can provide. Secondly, it favours replicated work as developers need to provide their own implementation to very similar problems. Lastly, it links analysis and visualisation to the target platform, which makes it very difficult to share cross-platform tools.

Thankfully, behavioural data is conceptually largely agnostic of the acquisition platform and paradigm. Typically, the behaviour of each individual is a long time series (possibly multivariate and heterogeneous). In addition, each individual has to be unambiguously identified and associated with arbitrary metadata defined by the experimenter (sex, treatment, genotype and so on). Efficiently combining and manipulating these information, on datasets of hundreds of individuals, each recorded for weeks, is not trivial. The availability of such a tool help promoting the analysis of behaviour as a data science.

Here, we describe `rethomics`, a framework that unifies analysis of large behavioural dataset in an efficient and manner. It offers an elegant solution to store, manipulate and visualise a large amount of behavioural data. Rethomics comes with a extensive documentation and a set of both practical and theoretical tutorials.

2 Approach

`rethomics` is implemented as a collection of small packages linked to one another. This development model follows modern frameworks such as the `tidyverse`, which results in increases testability and maintainability. The different tasks of the analysis workflow (*i.e.* data import, manipulation and visualisation) are explicitly handled by different packages (Fig. 1). At the core of `rethomics`, the `behavr` package offers an very flexible and efficient solution to store both large amounts data (*e.g.* position and activity) and metadata (*e.g.* treatment, genotype and so on) in a single `data.table` derived object. Any input package will import experimental data as a `behavr` table which can, in turn, be manipulated and visualised regardless of the input platform. Analyses results and plots integrate seamlessly within the `R` ecosystem, hence providing users with state visualisation and statistics tools.

3 Framework features

3.1 Data import

xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl

3.2 Internal data structure

xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl

3.3 Circadian analysis

xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl

3.4 Sleep scoring

xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl

3.5 Visualisation

xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl

4 Conclusion

xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl xxx xxxxx xxxx dwjdkl

References

Bofelli,F., Name2, Name3 (2003) Article title, *Journal Name*, **199**, 133-154.

Bag,M., Name2, Name3 (2001) Article title, *Journal Name*, **99**, 33-54.

Yoo,M.S. *et al*. (2003) Oxidative stress regulated genes in nigral dopaminergic neurnol cell: correlation with the known pathology in Parkinson's disease. *Brain Res. Mol. Brain Res.*, **110**(Suppl. 1), 76–84.

Lehmann,E.L. (1986) Chapter title. *Book Title*. Vol. 1, 2nd edn. Springer-Verlag, New York.

Crenshaw, B.,III, and Jones, W.B.,Jr (2003) The future of clinical cancer management: one tumor, one chip. *Bioinformatics*, doi:10.1093/bioinformatics/btn000.

Auhtor,A.B. *et al*. (2000) Chapter title. In Smith, A.C. (ed.), *Book Title*, 2nd edn. Publisher, Location, Vol. 1, pp. ???–???.

Bardet, G. (1920) Sur un syndrome d'obesite infantile avec polydactylie et retinite pigmentaire (contribution a l'etude des formes cliniques de l'obesite hypophysaire). PhD Thesis, name of institution, Paris, France.