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

Quentin Geissmann[1*], Luis Garcia Rodriguez[2], Esteban J Beckwith[1], Giorgio F Gilestro[1*]

1 Department of Life Sciences, Imperial College London, London, United Kingdom
2 Affiliation Dept/Program/Center, Institution Name, City, State, Country

* qgeissmann@gmail.com, giorgio@gilest.ro

Abstract

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. The `rethomics` framework is available and documented at https://github.com/rethomics/.

Introduction

Animal behaviours are complex phenotypical manifestations of the interaction between nervous systems and their 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 address 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 (*e.g.* sex, treatment and genotype). Efficiently combining and manipulating these information, on datasets of hundreds of individuals, each recorded for weeks, is not trivial. The availability of a unified ethomics toolbox would help promoting the analysis of behaviour as a data science.

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

Design and Implementation

`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 original input platform. Analyses results and plots integrate seamlessly within the `R` ecosystem, hence providing users with state visualisation and statistics tools.

0.1 Data import

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

0.2 Internal data structure

The baha

0.3 Circadian analysis

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

0.4 Sleep scoring

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

0.5 Visualisation

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

Results

Availability and Future Directions

All packages in the `rethomics` framework are available under the terms of the GPLv3 license and listed at https://github.com/rethomics/. Extensive installation instructions as well as reproducible demos and tutorials are available at https://rethomics.github.io/.

Acknowledgements

References

1. Conant GC, Wolfe KH. Turning a hobby into a job: how duplicated genes find new functions. Nat Rev Genet. 2008 Dec;9(12):938–950.

2. Ohno S. Evolution by gene duplication. London: George Alien & Unwin Ltd. Berlin, Heidelberg and New York: Springer-Verlag.; 1970.

3. Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a Duplication. PLoS Genet. 2011 Oct;7(10):e1002337.