EvAM-Tools: tools for evolutionary accumulation and cancer progression models

Ramon Diaz-Uriarte*, Pablo Herrera-Nieto*

2022-07-05

Abstract

EvAM-Tools is an R package and web application that provides a unified interface to state-of-the-art cancer progression models (CPMs) and, more generally, evolutionary models of event accumulation. The output includes, in addition to the fitted models, the transition (and transition rate) matrices between genotypes and the probabilities of evolutionary paths. Generation of random cancer progression models is also available. Using the GUI in the web application users can easily construct models (modifying Directed Acyclic Graphs —DAGs— of restrictions, matrices of mutual hazards, or specifying genotype composition), generate data from them (with user-specified observational/genotyping error), and analyze the data.

Availability and Implementation: Implemented in R and C; open source code available under the GNU Affero General Public License v3.0 at https://github.com/rdiaz02/EvAM-tools. Docker images freely available from https://hub.docker.com/u/rdiaz02. Web app freely accessible at https://iib.uam.es/evamtools.

Contact: r.diaz@uam.es

1 Introduction

Cancer progression models (CPMs) try to identify dependencies between genetic events (e.g., mutations) from cross-sectional data (Beerenwinkel *et al.*, 2015; Diaz-Uriarte, 2018). These dependencies can be deterministic, as in Oncogenetic Trees (OT) (Desper *et al.*, 1999; Szabo and Boucher, 2008), Conjunctive Bayesian Networks (CBN) (Gerstung *et al.*, 2009; Montazeri *et al.*, 2016), Disjunctive Bayesian Networks (DBN) (Nicol *et al.*, 2021), and Hidden Extended Suppes-Bayes Causal Networks (H-ESBCNs) (Angaroni *et al.*, 2021), or stochastic as in Mutual Hazard Networks (MHN) (Schill *et al.*, 2020). These models also implicitly encode the possible mutational trajectories with predictions about their probability, and have been used for predicting cancer evolution, both long-term (Diaz-Uriarte and Vasallo, 2019; Hosseini *et al.*, 2019) and short-term (Diaz-Colunga and Diaz-Uriarte, 2021). Although developed in the field of computational oncology, these models are not limited to cancer: they can be applied to other questions involving the (irreversible) accumulation of discrete items (Gotovos *et al.*, 2021), and have been used to examine tool use in animal taxa (Johnston and Røyrvik, 2020).

However, everyday use of these methods, and comparison of output from different ones, is hindered because they are implemented in different languages and with dependencies

*Department of Biochemistry, Universidad Autónoma de Madrid, Instituto de Investigaciones Biomédicas "Alberto Sols" (UAM-CSIC), Madrid, Spain. Author for correspondence: `r.diaz@uam.es`

that make them difficult to install under some operating systems, and they use heterogeneous formats for input and output. In addition, output is often textual (not graphical) and most implementations do not provide the output users might be interested in, such as probabilities of evolutionary paths and genotypes. Finally, it is difficult to explore how changes in genotype composition affect the fitted models and to analyze with one model data generated under another. Here we present EvAM-Tools, an R package and web app for evolutionary and event accumulation models, that facilitates the use and comparison of CPMs, and their extension to other problems.

2 Functionality

EvAM-tools is available as an R package, `evamtools`, that can also launch an interactive GUI as a Shiny (Chang *et al.*, 2021) web app. A distinctive feature is the unified input and usage of five CPMs, the available state-of-the-art procedures with existing implementations: OT (Desper *et al.*, 1999; Szabo and Boucher, 2008), CBN (both H-CBN — Gerstung *et al.*, 2009 — and MC-CBN — Montazeri *et al.*, 2016), DBN (Nicol *et al.*, 2021), H-ESBCNs (Angaroni *et al.*, 2021), MHN (Schill *et al.*, 2020). Output includes:

- DAGs of restrictions annotated with parameter values (OT, CBN, OncoBN, H-ESBCN) and log-Θ matrix (MHN).
- Plots of the transition probabilities between genotypes and transition rate matrices and observed genotype transitions from the continuous-time Markov Chain (for models where this exists: CBN, H-ESBCN, MHN). As visualizing the acquisition of mutations in a complex network can be challenging, we use the hypergraph transition graph from HyperTraPS (Greenbury *et al.*, 2020) as our standard and, to enhance interpretability, also allow the user to modify the number of paths to show and the type of annotation.
- Predicted probabilities of genotypes and finite samples from the predicted probabilities with user-specified noise (observational, genotyping error) levels; probabilities of evolutionary paths.

The Shiny web app provides a **GUI** that allows to interactively modify: (a) genotype frequencies; (b) the log-Θ matrix for MHN models; (c) the DAG of restrictions and the type of dependencies (AND, OR, XOR) and their rates (CBN, H-ESBCN) or conditional probabilities (OT, OncoBN). We provide examples of each to ease initial exploration. The web app allows users to import data sets and download created data sets and analysis output. The R package also provides a function for the random generation of CPMs/evam models.

3 Tests, code coverage, and documentation

The package includes more than 1400 tests that run at every build/check cycle. These tests provide a code coverage of more than 90%. The Shiny web app includes usage documentation in the landing page, as well as in the application itself as short documentation strings next to each option and, for longer help, via clickable "Help" buttons. The `evamtools` R package includes standard R function documentation. The repository README and additional documentation include: installation instructions; details about the web app; details about predicted genotype frequencies and random evam models; a FAQ; explanation on using OncoSimulR (Diaz-Uriarte, 2017) to obtain transition probabilities between genotypes.

4 Use cases and availability

EvAM-Tools can be used, among others, as:

- A teaching tool to understand CPMs and how different inputs (which can be easily modified interactively by the user) affect the fitted models.
- A research tool to analyze cross-sectional data with state-of-the-art CPMs.
- A research tool in methodological work to compare the performance of different methods under data generated by other models or under specific evolutionary scenarios, and to evaluate new methods.

The supplementary file "Supplementary examples" illustrates some of the above uses with both cancer data and simulated data sets.

The web app can be run on our servers or locally. In the latter case after installing the package from source or using one of the Docker images we provide, which include all dependencies and will run both the package and the web app.

5 Conclusion

EvAM-Tools provides a unified interface to state-of-the-art CPMs, a user-friendly GUI, output (graphical and tabular) that extends the usual output derived from these methods, and functionality that facilitates methodological work. EvAM-Tools should result in better understanding and increased use of these methods and encourage further methodological work and their extension to other scenarios, making it a tool of interest for computational oncologists and more broadly to evolutionary biologists.

Acknowledgements

C. Lazaro-Perea for comments on the ms.

Funding

Work supported by grant PID2019-111256RB-I00 funded by MCIN/AEI/10.13039/501100011033 to RDU. PHN supported by Comunidad de Madrid's PEJ-2019-AI/BMD-13961 to RDU.

References

Angaroni, F., Chen, K., Damiani, C., Caravagna, G., Graudenzi, A., and Ramazzotti, D. (2021). PMCE: efficient inference of expressive models of cancer evolution with high prognostic power. *Bioinformatics*, **38**(3), 754—762.

Beerenwinkel, N., Schwarz, R. F., Gerstung, M., and Markowetz, F. (2015). Cancer evolution: Mathematical models and computational inference. *Systematic Biology*, **64**(1), e1–e25.

Chang, W., Cheng, J., Allaire, J., Sievert, C., Schloerke, B., Xie, Y., Allen, J., McPherson, J., Dipert, A., and Borges, B. (2021). *shiny: Web Application Framework for R*. R package version 1.7.1.

Desper, R., Jiang, F., Kallioniemi, O.-P., Moch, H., Papadimitriou, C. H., and Schäffer, A. A. (1999). Inferring tree models for oncogenesis from comparative genome hybridization data. *Journal of computational biology*, **6**(1), 37–51.

Diaz-Colunga, J. and Diaz-Uriarte, R. (2021). Conditional prediction of consecutive tumor evolution using cancer progression models: What genotype comes next? *PLOS Computational Biology*, **17**(12), e1009055.

Diaz-Uriarte, R. (2017). OncoSimulR: Genetic simulation with arbitrary epistasis and mutator genes in asexual populations. *Bioinformatics*, **33**(12), 1898–1899.

Diaz-Uriarte, R. (2018). Cancer progression models and fitness landscapes: A many-to-many relationship. *Bioinformatics*, **34**(5), 836–844.

Diaz-Uriarte, R. and Vasallo, C. (2019). Every which way? On predicting tumor evolution using cancer progression models. *PLoS computational biology*, **15**(8), e1007246.

Gerstung, M., Baudis, M., Moch, H., and Beerenwinkel, N. (2009). Quantifying cancer progression with conjunctive bayesian networks. *Bioinformatics*, **25**(21), 2809–2815.

Gotovos, A., Burkholz, R., Quackenbush, J., and Jegelka, S. (2021). Scaling up Continuous-Time Markov Chains Helps Resolve Underspecification. *arXiv:2107.02911 [cs, stat]*.

Greenbury, S. F., Barahona, M., and Johnston, I. G. (2020). Hypertraps: Inferring probabilistic patterns of trait acquisition in evolutionary and disease progression pathways. *Cell systems*, **10**(1), 39–51.

Hosseini, S.-R., Diaz-Uriarte, R., Markowetz, F., and Beerenwinkel, N. (2019). Estimating the predictability of cancer evolution. *Bioinformatics*, **35**(14), i389–i397.

Johnston, I. G. and Røyrvik, E. C. (2020). Data-Driven Inference Reveals Distinct and Conserved Dynamic Pathways of Tool Use Emergence across Animal Taxa. *iScience*, **23**(6), 101245.

Montazeri, H., Kuipers, J., Kouyos, R., Böni, J., Yerly, S., Klimkait, T., Aubert, V., Günthard, H. F., Beerenwinkel, N., and Study, S. H. C. (2016). Large-scale inference of conjunctive bayesian networks. *Bioinformatics*, **32**(17), i727–i735.

Nicol, P. B., Coombes, K. R., Deaver, C., Chkrebtii, O., Paul, S., Toland, A. E., and Asiaee, A. (2021). Oncogenetic network estimation with disjunctive bayesian networks. *Computational and Systems Oncology*, **1**(2), e1027.

Schill, R., Solbrig, S., Wettig, T., and Spang, R. (2020). Modelling cancer progression using mutual hazard networks. *Bioinformatics*, **36**(1), 241–249.

Szabo, A. and Boucher, K. M. (2008). Oncogenetic trees. In W.-Y. Tan and L. Hanin, editors, *Handbook of Cancer Models with Applications*, pages 1–24. World Scientific.