




T-BAS Version 2.1: Tree-Based Alignment Selector Toolkit for Evolutionary Placement of DNA Sequences and Viewing Alignments and Specimen Metadata on Curated and Custom Trees

Ignazio Carbone,^a James B. White,^a Jolanta Miadlikowska,^b A. Elizabeth Arnold,^{c,d} Mark A. Miller,^e Nicolas Magain,^{b,f}  Jana M. U'Ren,^g François Lutzoni^b

^aCenter for Integrated Fungal Research, Department of Entomology and Plant Pathology, North Carolina State University, Raleigh, North Carolina, USA

^bDepartment of Biology, Duke University, Durham, North Carolina, USA

^cSchool of Plant Sciences, The University of Arizona, Tucson, Arizona, USA

^dDepartment of Ecology and Evolutionary Biology, The University of Arizona, Tucson, Arizona, USA

^eSan Diego Supercomputer Center, University of California, San Diego, La Jolla, California, USA

^fInstitut de Botanique B22, Biologie de l'Évolution et de la Conservation, UR InBios, Université de Liège, Liège, Belgium

^gDepartment of Biosystems Engineering, The University of Arizona, Tucson, Arizona, USA

ABSTRACT The Tree-Based Alignment Selector (T-BAS) toolkit combines phylogenetic-based placement of DNA sequences with alignment and specimen metadata visualization tools in an integrative pipeline for analyzing microbial biodiversity. The release of T-BAS version 2.1 makes available reference phylogenies, supports multilocus sequence placements and permits uploading and downloading trees, alignments, and specimen metadata.

The Tree-Based Alignment Selector (T-BAS) toolkit offers an integrated phylogenetic analysis and visualization framework that combines evolutionary placement of DNA sequences with multilocus sequence alignments and metadata (e.g., countries, states, sites, hosts, sequence data, taxonomic classifications, and phenotypic traits), enabling biodiversity discovery and facilitating rapid species description (1–3). Here, we announce the public release of T-BAS version 2.1, which builds on the functionality of T-BAS version 1.0 and allows users to upload any reference tree, multilocus DNA sequence alignments, and specimen metadata and to perform multilocus placements of DNA sequences with user-customizable metadata (Fig. 1). Phylogeny-based placement of DNA sequences can be performed on trees and multilocus alignments (e.g., those available in TreeBase [www.treebase.org] [4, 5] or custom trees uploaded by users), greatly expanding the utility of T-BAS across biological disciplines. Trees, alignments, and specimen metadata also can be uploaded and viewed without performing placements (6, 7), making it a versatile phylogenetic tool across diverse taxonomic groups.

In T-BAS version 2.1, phylogenetic placement and specimen metadata output are standardized via the extended PhyloXML format (8) so that placements can be readily viewed by programs such as PhyD3 (9), IcyTree (10), TreeGraph 2 (11), and Archaeopteryx.js (version 0.9928 beta-2018-07-05) (12). Metadata can be displayed as colored outer rings in phylogenies, as network graphs with node attributes, or as pie charts via the JavaScript library D3.js (<https://d3js.org/>). Users can select specific clades on tree outputs from T-BAS analyses and then perform phylogenetic inference across single or multiple alignments with RAXML (13) or network inference with the genealogical method of Templeton et al. (14) implemented in TCS version 1.21 (15) and NetworkX

Citation Carbone I, White JB, Miadlikowska J, Arnold AE, Miller MA, Magain N, U'Ren JM, Lutzoni F. 2019. T-BAS version 2.1: Tree-Based Alignment Selector toolkit for evolutionary placement of DNA sequences and viewing alignments and specimen metadata on curated and custom trees. *Microbiol Resour Announc* 8:e00328-19. <https://doi.org/10.1128/MRA.00328-19>.

Editor Christina A. Cuomo, Broad Institute

Copyright © 2019 Carbone et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Ignazio Carbone, ignazio_carbone@ncsu.edu.

Received 19 March 2019

Accepted 23 June 2019

Published 18 July 2019

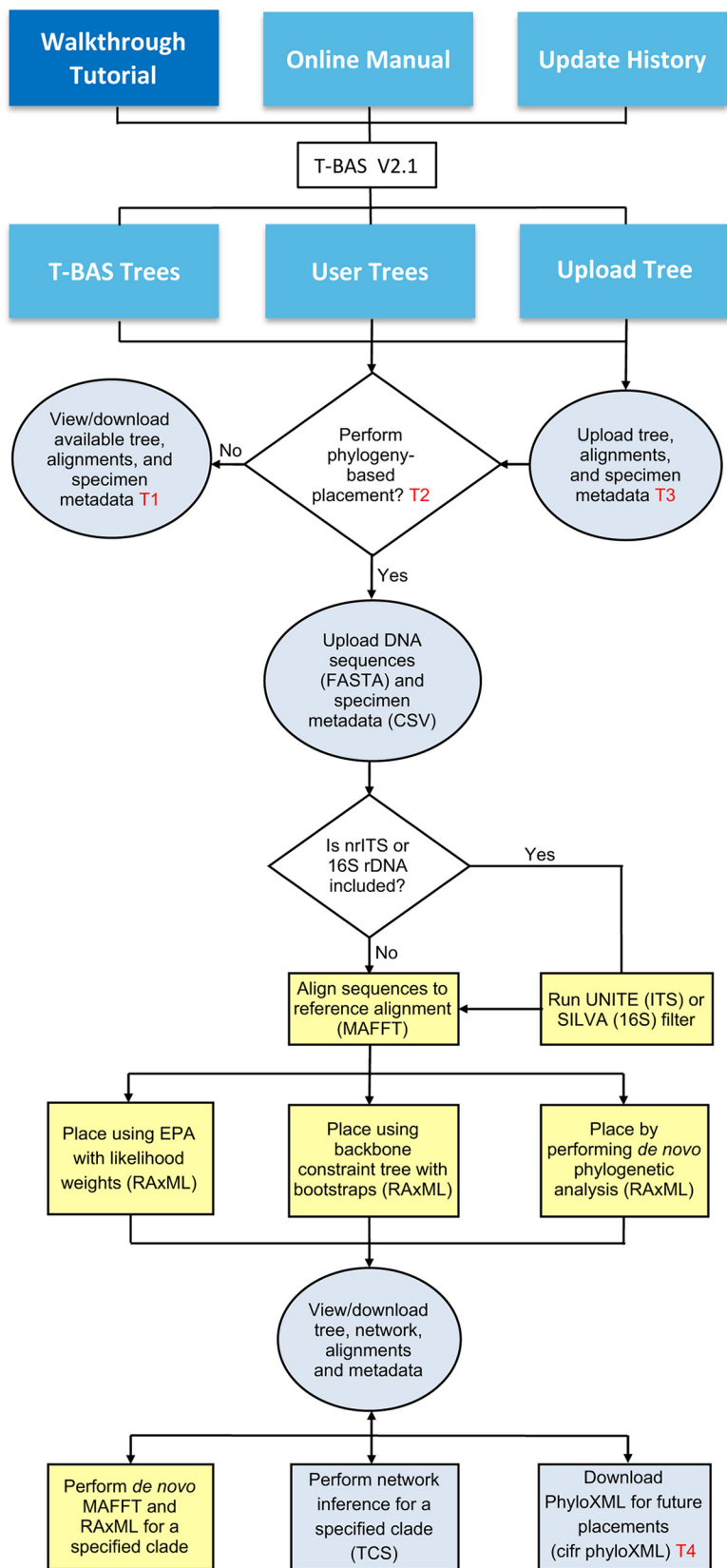


FIG 1 A schematic flowchart showing the three major features in T-BAS version 2.1 and their partitioning between servers at North Carolina (NC) State University (blue) and CIPRES (yellow). Reference trees included in this release are fungi (22), bacteria (23), Pezizomycotina version 1 (1) and extended version 2, (Continued on next page)

Downloaded from <http://mra.asm.org/> on August 22, 2019 at UNIVERSITY OF ARIZONA LIBRARY

FIG 1 Legend (Continued)

Lecanoromycetes (24), Laboulbeniomyces (25), Sebaciales (26), Xylariaceae (27), *Peltigera* (28), *Ramularia eucalypti* species complex (29), *Aspergillus* section *Flavi* (7), and the *Fusarium* species complex (30). The three primary user actions (gray) are (i) select a T-BAS reference tree, which includes DNA sequence alignments and specimen metadata information for viewing and downloading; (ii) perform phylogeny-based placement of DNA sequences for up to seven loci on a T-BAS tree, directly or after sequence data are compiled into operational taxonomic units (OTUs); or (iii) upload a new tree with alignments and metadata and place specimen sequences, with the option of submitting it for inclusion in the T-BAS framework. Documentation includes tutorials, a user manual, and an update history which lists the new features in T-BAS version 2.1 compared with those in version 1.0. Abbreviations in the flowchart are as follows: T1, tutorial 1; T2, tutorial 2; T3, tutorial 3; T4, tutorial 4; EPA, Evolutionary Placement Algorithm (17); UNITE, unified system for the DNA-based fungal species linked to the classification (19); SILVA, database for rRNA gene sequences (16S rDNA) (20); BLAST, Basic Local Alignment Search Tool (31); RAxML, Randomized Axelerated Maximum Likelihood (13); MAFFT, Multiple Alignment using Fast Fourier Transform (32); PhyloXML, standard using extensible markup language (8) to encode phylogenetic tree, sequence alignments, and specimen metadata information.

(16). Three phylogeny-based placement options are available, including Evolutionary Placement Algorithm (EPA) (17), backbone constraint tree, and *de novo* phylogenetic reconstruction via RAxML (13) through the Representational State Transfer Application Program Interface (REST API) service at Cyberinfrastructure for Phylogenetic Research (CIPRES) (18). Synchronization of color-coded attributes across subtrees and networks is possible using a new specimen metadata color editor (see online user manual).

A new guide tree framework in T-BAS version 2.1 allows users to select the appropriate reference tree for multilocus placements of taxa via sequence data for up to seven loci. Phylogenomic trees based on hundreds of loci also can serve as reference trees for multilocus placement of DNA sequences. Prefiltering of barcode sequences is included as an option via the UNITE fungal nrITS database (version 7.2 release date, 28 June 2017) (19) and SILVA bacterial database (release version 132) (20). This is important when placing sequences from DNA metabarcoding of nrITS for organisms such as fungi to ensure that only taxon sequences within the taxonomic breadth of the reference trees are aligned.

Fine resolution of phylogenetic relationships is possible in T-BAS version 2.1 via an expanded phylogenetic framework that can accommodate more reference trees and increased taxonomic sampling. For example, the Pezizomycotina reference tree released in T-BAS version 1.0 has been extended from 979 to 1,625 taxa based on NCBI RefSeq records (21) and provides access to alignments for up to six loci (nr5.8S, nrLSU, nrSSU, mtSSU, *RPB1*, and *RPB2*). Additional reference trees, multilocus alignments, and specimen metadata are made available (Fig. 1). We anticipate the number of trees available for placement to grow as users contribute new reference trees and metadata. T-BAS version 2.1 allows the research community to build on phylogenetic resources by submitting reference trees for inclusion in the T-BAS framework and sharing these trees, alignments, and specimen metadata with other worldwide experts to enhance taxonomic discovery, description of new species, and the tracking of biodiversity for ecological studies.

Data availability. The T-BAS toolkit is available at <https://tbas.hpc.ncsu.edu/>.

ACKNOWLEDGMENTS

Funding for the development of T-BAS version 2.1 was provided by the National Science Foundation (NSF) Genealogy of Life (GoLife) program to I.C. (DEB-1541418), A.E.A. (DEB-1541496), and F.L. and J.M. (DEB-1541548). The CIPRES RESTful API is supported by National Institutes of Health (NIH) 5 R01 GM1264635 and NSF DBI-1759844 and a generous award (TG-DEB090011) of computer time and development support from the XSEDE project (also sponsored by the U.S. National Science Foundation).

We thank Conrad Schoch, Barbara Robbertse, and Pooja Strope for their help in retrieving and validating GenBank sequences for version 2 of the Pezizomycotina tree. We also thank the following for contributing published trees: Tim James (fungi), Laura Hug (bacteria), Danny Haelewaters (Laboulbeniomyces), Sigisfredo Garnica (Sebaci-

nales), Pedro Crous (*Ramularia eucalypti* species complex), Geromy Moore (*Aspergillus* section *Flavi*), and Kerry O'Donnell (*Fusarium* species complex). We appreciate the help of Vicki Cornish for assistance with T-BAS tutorials.

REFERENCES

- Carbone I, White JB, Miadlikowska J, Arnold AE, Miller MA, Kauff F, U'Ren JM, May G, Lutzoni F. 2017. T-BAS: Tree-Based Alignment Selector toolkit for phylogenetic-based placement, alignment downloads and metadata visualization: an example with the Pezizomycotina tree of life. *Bioinformatics* 33:1160–1168. <https://doi.org/10.1093/bioinformatics/btw808>.
- Daru BH, Bowman EA, Pfister DH, Arnold AE. 2018. A novel proof of concept for capturing the diversity of endophytic fungi preserved in herbarium specimens. *Philos Trans R Soc Lond B Biol Sci* 374:20170395. <https://doi.org/10.1098/rstb.2017.0395>.
- Huang Y-L, Bowman EA, Massimo NC, Garber NP, U'Ren JM, Sandberg DC, Arnold AE. 2018. Using collections data to infer biogeographic, environmental, and host structure in communities of endophytic fungi. *Mycologia* 110:47–62.
- Cullen M, Jacob ME, Cornish V, VanderSchel IQ, Cotter HVT, Cubeta MA, Carbone I, Gilger BC. 2019. Multi-locus DNA sequence analysis, antifungal agent susceptibility, and fungal keratitis outcome in horses from Southeastern United States. *PLoS One* 14:e0214214. <https://doi.org/10.1371/journal.pone.0214214>.
- Koehler AM, Larkin MT, Rogers LW, Carbone I, Cubeta MA, Shew HD. 2019. Identification and characterization of *Septoria steviae* as the causal agent of Septoria leaf spot disease of stevia in North Carolina. *Mycologia* 111:456–465. <https://doi.org/10.1080/00275514.2019.1584503>.
- Blanco-Meneses M, Carbone I, Ristaino JB. 2018. Population structure and migration of the tobacco blue mold pathogen, *Peronospora tabacina*, into North America and Europe. *Mol Ecol* 27:737–751. <https://doi.org/10.1111/mec.14453>.
- Moore GG, Olarte RA, Horn BW, Elliott JL, Singh R, O'Neal CJ, Carbone I. 2017. Global population structure and adaptive evolution of aflatoxin-producing fungi. *Ecol Evol* 7:9179–9191. <https://doi.org/10.1002/ece3.3464>.
- Han MV, Zmasek CM. 2009. phyloXML: XML for evolutionary biology and comparative genomics. *BMC Bioinformatics* 10:356. <https://doi.org/10.1186/1471-2105-10-356>.
- Kreft L, Botzki A, Coppens F, Vandepoel K, Van Bel M. 2017. PhyD3: a phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics* 33:2946–2947. <https://doi.org/10.1093/bioinformatics/btx324>.
- Vaughan TG. 2017. IcyTree: rapid browser-based visualization for phylogenetic trees and networks. *Bioinformatics* 33:2392–2394. <https://doi.org/10.1093/bioinformatics/btx155>.
- Stover BC, Muller KF. 2010. TreeGraph 2: combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics* 11:7. <https://doi.org/10.1186/1471-2105-11-7>.
- Zmasek CM, Eddy SR. 2001. ATV: display and manipulation of annotated phylogenetic trees. *Bioinformatics* 17:383–384. <https://doi.org/10.1093/bioinformatics/17.4.383>.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>.
- Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619–633.
- Clement M, Posada D, Crandall KA. 2000. TCS: a computer program to estimate gene genealogies. *Mol Ecol* 9:1657–1659. <https://doi.org/10.1046/j.1365-294x.2000.01020.x>.
- Hagberg A, Schult D, Swart P. 2008. Exploring network structure, dynamics, and function using NetworkX, p 11–15. *In Proceedings of the 7th Python in Science Conference*, Pasadena, CA.
- Berger SA, Stamatakis A. 2011. Aligning short reads to reference alignments and trees. *Bioinformatics* 27:2068–2075. <https://doi.org/10.1093/bioinformatics/btr320>.
- Miller MA, Schwartz T, Pickett BE, He S, Klem EB, Scheuermann RH, Passarotti M, Kaufman S, O'Leary MA. 2015. A RESTful API for access to phylogenetic tools via the CIPRES science gateway. *Evol Bioinform Online* 11:43–48. <https://doi.org/10.4137/EBO.S21501>.
- Köljalg U, Nilsson RH, Abarenkov K, Tedersoo L, Taylor AFS, Bahram M, Bates ST, Bruns TD, Bengtsson-Palme J, Callaghan TM, Douglas B, Drenkhan T, Eberhardt U, Dueñas M, Grebenc T, Griffith GW, Hartmann M, Kirk PM, Kohout P, Larsson E, Lindahl BD, Lücking R, Martín MP, Matheny PB, Nguyen NH, Niskanen T, Oja J, Peay KG, Peintner U, Peterson M, Pöldmaa K, Saag L, Saar I, Schüßler A, Scott JA, Senés C, Smith ME, Suija A, Taylor DL, Telleria MT, Weiss M, Larsson K-H. 2013. Towards a unified paradigm for sequence-based identification of fungi. *Mol Ecol* 22:5271–5277. <https://doi.org/10.1111/mec.12481>.
- Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glockner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and Web-based tools. *Nucleic Acids Res* 41:D590–D596. <https://doi.org/10.1093/nar/gks1219>.
- O'Leary NA, Wright MW, Brister JR, Ciuffo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, Astashyn A, Badretdin A, Bao Y, Blinkova O, Brover V, Chetvernin V, Choi J, Cox E, Ermolaeva O, Farrell CM, Goldfarb T, Gupta T, Haft D, Hatcher E, Hlavina W, Joardar VS, Kodali VK, Li W, Maglott D, Masterson P, McGarvey KM, Murphy MR, O'Neill K, Pujar S, Rangwala SH, Rausch D, Riddick LD, Schoch C, Shkeda A, Storz SS, Sun H, Thibaud-Nissen F, Tolstoy I, Tully RE, Vatsan AR, Wallin C, Webb D, Wu W, Landrum MJ, Kimchi A. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Res* 44:D733–D745. <https://doi.org/10.1093/nar/gkv1189>.
- James TY, Kauff F, Schoch CL, Matheny PB, Hofstetter V, Cox CJ, Celio G, Gueidan C, Fraker E, Miadlikowska J, Lumbsch HT, Rauhut A, Reeb V, Arnold AE, Amtoft A, Stajich JE, Hosaka K, Sung G-H, Johnson D, O'Rourke B, Crockett M, Binder M, Curtis JM, Slot JC, Wang Z, Wilson AW, Schüßler A, Longcore JE, O'Donnell K, Mozley-Standridge S, Porter D, Letcher PM, Powell MJ, Taylor JW, White MM, Griffith GW, Davies DR, Humber RA, Morton JB, Sugiyama J, Rossman AY, Rogers JD, Pfister DH, Hewitt D, Hansen K, Hambleton S, Shoemaker RA, Kohlmeyer J, Volkmann-Kohlmeier B, Spotts RA, Serdani M, Crous PW, Hughes KW, Matsuura K, Langer E, Langer G, Untereiner WA, Lücking R, Büdel B, Geiser DM, Aptroot A, Diederich P, Schmitt I, Schultz M, Yahr R, Hibbett DS, Lutzoni F, McLaughlin DJ, Spatafora JW, Vilgalys R. 2006. Reconstructing the early evolution of fungi using a six-gene phylogeny. *Nature* 443:818–822. <https://doi.org/10.1038/nature05110>.
- Hug LA, Baker BJ, Anantharaman K, Brown CT, Probst AJ, Castelle CJ, Butterfield CN, Hemsdorf AW, Amano Y, Ise K, Suzuki Y, Dudek N, Relman DA, Finstad KM, Amundson R, Thomas BC, Banfield JF. 2016. A new view of the tree of life. *Nat Microbiol* 1:16048. <https://doi.org/10.1038/nmicrobiol.2016.48>.
- Miadlikowska J, Kauff F, Hognabba F, Oliver JC, Molnar K, Fraker E, Gaya E, Hafellner J, Hofstetter V, Gueidan C, Otolara MA, Hodkinson B, Kukwa M, Lücking R, Bjork C, Sipman HJ, Burgaz AR, Thell A, Passo A, Myllys L, Goward T, Fernandez-Brime S, Hestmark G, Lendemer J, Lumbsch HT, Schumler M, Schoch CL, Serusiaux E, Maddison DR, Arnold AE, Lutzoni F, Stenroos S. 2014. A multigene phylogenetic synthesis for the class Lecanoromycetes (Ascomycota): 1307 fungi representing 1139 infrageneric taxa, 317 genera and 66 families. *Mol Phylogenet Evol* 79:132–168. <https://doi.org/10.1016/j.ympev.2014.04.003>.
- Haelewaters D, Pfliegler WP, Gorczak M, Pfister DH. 2019. Birth of an order: comprehensive molecular phylogenetic study excludes *Herpomyces* (Fungi, Laboulbeniomycetes) from Laboulbeniales. *Mol Phylogenet Evol* 133:286–301. <https://doi.org/10.1016/j.ympev.2019.01.007>.
- Garnica S, Riess K, Schon ME, Oberwinkler F, Setaro SD. 2016. Divergence times and phylogenetic patterns of Sebaciales, a highly diverse and widespread fungal lineage. *PLoS One* 11:e0149531. <https://doi.org/10.1371/journal.pone.0149531>.
- U'Ren JM, Miadlikowska J, Zimmerman NB, Lutzoni F, Stajich JE, Arnold AE. 2016. Contributions of North American endophytes to the phylogeny, ecology, and taxonomy of Xylariaceae (Sordariomycetes, Ascomy-

- cota). *Mol Phylogenet Evol* 98:210–232. <https://doi.org/10.1016/j.ympev.2016.02.010>.
28. Chagnon P-L, Magain N, Miadlikowska J, Lutzoni F. 2019. Species diversification and phylogenetically constrained symbiont switching generated high modularity in the lichen genus *Peltigera*. *J Ecol*. <https://doi.org/10.1111/1365-2745.13207>.
29. Videira SI, Groenewald JZ, Kolecka A, van Haren L, Boekhout T, Crous PW. 2015. Elucidating the *Ramularia eucalypti* species complex. *Persoonia* 34:50–64. <https://doi.org/10.3767/003158515X685670>.
30. O'Donnell K, Sutton DA, Wiederhold N, Robert VA, Crous PW, Geiser DM. 2016. Veterinary fusarioses within the United States. *J Clin Microbiol* 54:2813–2819. <https://doi.org/10.1128/JCM.01607-16>.
31. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
32. Katoh K, Toh H. 2010. Parallelization of the MAFFT multiple sequence alignment program. *Bioinformatics* 26:1899–1900. <https://doi.org/10.1093/bioinformatics/btq224>.