Mobile DNA

COMMENTARY

TE Hub: A community-oriented space for sharing and connecting tools, data, resources, and methods for transposable element annotation

The TE Hub Consortium, Tyler A. Elliott[1], Tony Heitkam[2], Robert Hubley[3], Hadi Quesneville[4], Alexander Suh[5,6] and Travis J. Wheeler[7*]

Abstract

Transposable elements (TEs) play powerful and varied evolutionary and functional roles, and are widespread in most eukaryotic genomes. Research into their unique biology has driven the creation of a large collection of databases, software, classification systems, and annotation guidelines. The diversity of available TE-related methods and resources raises compatibility concerns and can be overwhelming to researchers and communicators seeking straightforward guidance or materials. To address these challenges, we have initiated a new resource, TE Hub, that provides a space where members of the TE community can collaborate to document and create resources and methods. The space consists of (1) a website organized with an open wiki framework, https://tehub.org, (2) a conversation framework via a Twitter account and a Slack channel, and (3) bi-monthly Hub Update video chats on the platform's development. In addition to serving as a centralized repository and communication platform, TE Hub lays the foundation for improved integration, standardization, and effectiveness of diverse tools and protocols. We invite the TE community, both novices and experts in TE identification and analysis, to join us in expanding our community-oriented resource.

Keywords: Transposable elements, Classification, Annotation, Database, Software, Collaboration, Community

Introduction

Transposable elements (TEs) are mobile and often-replicating genetic elements that make up a significant fraction of most eukaryotic genomes (for reviews, see [1, 2]). Their study is important in genome research [3] as they can be viewed as motors of evolution [4], regulators of gene control [5], and as genomic building blocks [6]. Over the years, the field has accumulated a plethora of

databases, software, classification systems, and annotation guidelines [7, 8]. These options provide researchers with the tools to discover and investigate TEs in existing and new genome sequences, and to update and revisit already characterized TEs. However, in-depth TE detection and analysis is laborious, and largely requires significant expertise in TE biology.

The expansive and diverse collection of tools and methods often leads to at least two significant problems for even the most experienced bioinformatician. First, the set of available choices can be overwhelming, leaving the researcher unsure of preferred methods for analyzing particular data types. Second, the multitude of databases and tools often suffers from compatibility concerns in

* Correspondence: Travis.Wheeler@umontana.edu
Tyler A. Elliott, Tony Heitkam, Robert Hubley, Hadi Quesneville, Alexander Suh, and Travis J. Wheeler wish to be considered as equal co-authors, and note that author order is due to the alphabetical order of last names. Consortium members are listed after Acknowledgements.
[7]Department of Computer Science, University of Montana, Missoula, MT, USA
Full list of author information is available at the end of the article

both nomenclature and output format [9, 10]; this is especially true for databases focusing on different TE types or host organisms.

Yearly conferences and workshops [11–13] provide some relief from these pressures – tool/database developers can meet and find common ground, while users of these resources can gain valuable exposure to new methods and best practices. Even so, the transient and punctuated nature of these meetings, combined with rapid developments in the TE field, leave much to be desired in terms of collaboration, interactivity, and persistent documentation of existing methods and best practices.

We have initiated TE Hub as an answer to these challenges. TE Hub is envisioned as a community-oriented framework that will serve as a resource for novice and expert TE researchers. For novices, TE Hub gives practical insight into available TE resources and methods, and for experts and developers, it provides a platform for increased communication and improved integration of methods and databases (see Fig. 1). Specifically, TE Hub is designed to support the TE community in three ways:

1. We have developed a website (https://tehub.org) that serves as an up-to-date compendium of information about TE research; the site is managed by an open wiki framework, so that all members of the TE community can contribute in an open, nimble, and transparent manner.
2. We have established a framework for focused communication among and with TE Hub contributors, via a messaging channel (#te-hub) housed in the larger TransposonsWorldwide Slack workspace [14], and a dedicated Twitter account (@hub_te).
3. The website, supplemented by open bi-monthly meetings, lays the foundation for development of a federated mechanism for integrating tools, databases, and resources in a way that will, over the long term, improve and standardize their value to the TE research community.

In the following sections, we provide further details about these components of TE Hub, describing the current state and establishing a vision for its future. TE Hub is a community-oriented resource, and we wrap up by describing how interested TE experts and novices can get involved.

The TE Hub website

The focal point of TE Hub is the website: https://tehub.org, which is intended to serve as a compendium of tools, databases, and other features of value to TE researchers, both novice and expert. The site content is managed via a wiki system, so that researchers can contribute to the content in an open, timely, and transparent fashion. TE Hub data is roughly organized along the following facets of TE-related information:

1. Classification. This section captures a collection of established classification schemes, both overarching and specific for certain hosts and TE-types. At the time of this writing, the five most commonly used overarching TE classification systems [15–19] are represented, along with four specialized classification systems [20–23]. Furthermore, a collection of 519 TE lineages is captured, each with at least one relevant reference in the literature. These will be particularly useful to TE novices, aiming to understand common nomenclature and the relationships between alternative systematic hierarchies.
2. Databases. This section compiles a list of databases for the storage of sequences and metadata associated with TEs, with links to each database and corresponding publication, along with a description of the represented repeat types and taxonomic groups. At the time of this writing, 150 databases are represented.
3. Tools. This section compiles a list of software for the detection, annotation, analysis, simulation, and visualization of TEs. Websites, preprints, and journal articles are linked, and associated with
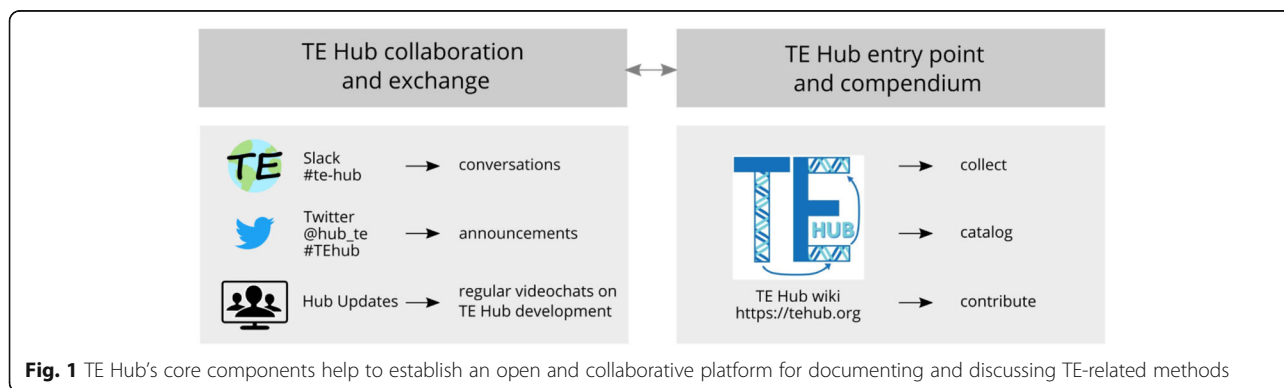
Fig. 1 TE Hub's core components help to establish an open and collaborative platform for documenting and discussing TE-related methods

Elliott *et al. Mobile DNA*   (2021) 12:16

Page 3 of 5

keywords. At the time of this writing, 505 tools are represented.

4. Protocols. Over time, this section will hold a collection of suggested protocols for use by researchers engaged in TE identification and annotation. The lack of carefully-crafted, discoverable, open-access protocols is an impediment to novice TE annotators. At the time of this writing, two protocols are listed; we expect this section to be substantially expanded in the coming months, and invite experienced annotators to contribute their mature and open access protocols.

5. Journals and Conferences. These sections capture a collection of journals that often publish TE-relevant articles, and a (community-maintained) listing of upcoming TE-related conferences.

6. Outreach and Teaching Resources. These sections hold a collection of educational resources that are intended to provide background on TEs, course materials for TE-related classes and workshops, and links to public talks on TEs intended for a general audience.

Contribution to the TE Hub is strongly encouraged and requires ORCID authentication. Dependency on ORCID ensures that content can be credited to each contributor, and represents a small barrier to contribution, as creation of an ORCID account takes only a few minutes. All TE Hub content is made available under the CC-BY license (https://creativecommons.org/licenses/by/4.0).

TE Hub communication channels

As a complement to the frequently updated but relatively static content of the TE Hub website, we have established mechanisms for scheduled and ad hoc communication about TE annotation resources and methods. These include:

1. The #te-hub channel, housed in the broader TransposonsWorldwide Slack workspace (https://transposonsworldwide.slack.com; currently with over 500 members). The #te-hub messaging channel is focused on the databases, software, and annotation methods central to TE Hub, leaving broader matters of TE biology to other TransposonsWorldwide channels. To insure against a records loss, conversations on the #te-hub channel will be regularly archived.

2. The @hub_te Twitter account (https://twitter.com/hub_te) will be used for TE Hub announcements, and the #TEhub hashtag will be adopted as a mechanism for highlighting Hub-relevant tweets.

3. 'Hub Updates' are video calls that serve as a regular medium for communication among database/methods developers and users of these methods. Meetings run for one hour, are held on a bi-monthly basis (organized transparently via the above Slack channel and Twitter account), and are open to all. These meetings have been ongoing since June 2020.

A foundation for the future of TE annotation

Creation of the TE Hub wiki resource and communication channels are the first step in a larger plan to develop a framework for improved integration of disparate TE datasets, tools, and resources. TE Hub is not, and is not intended to become, a replacement for individual TE databases (e.g. Repbase Update [15], DFAM [18], RepetDB [24], GyDB [20]) or annotation methods (e.g. RepeatModeler2 [25], REPET [26], RepeatExplorer2 [27]). Rather, the vision is that these first TE Hub developments will lay the foundation for future efforts to build a common language around diverse databases, establish a system for improving interoperability of independent TE identification and annotation software that capitalizes on each tool's individual strengths, and develop an increasingly robust catalog of annotation protocols, all with the goal of improving the ease and effectiveness of annotation for a maximally-broad diversity of organisms. In the meantime, the current compendium of methods and data will serve as a bridge to the future for TE annotators.

Call for engagement and contribution

TE Hub has grown out of a grassroots effort to expand international collaboration in the development of TE identification and annotation methods, to broaden and unify their applicability to non-model organisms, and to establish a comprehensive catalog of TE resources that can be easily updated by members of the community. The regular Hub Update meetings grew out of discussions in the Slack channel, and led to the content and vision described here. While this effort has been driven by a small steering committee rising out of these Hub Update meetings, the future of TE Hub depends on engagement and contribution from others in the community.

We invite the TE community, both novice and expert TE researchers, to join us in expanding our community-oriented resource. Please follow us on Twitter: @hub_te (https://twitter.com/hub_te), and visit https://tehub.org/volunteer for more information about contributing to the future of TE Hub. To fully engage, two registrations are recommended:

Elliott *et al. Mobile DNA*   (2021) 12:16

Page 4 of 5

1. Join the TransposonsWorldwide Slack workspace (https://transposonsworldwide.slack.com) and find the #te-hub channel under "Browse channels". This will allow you to track and contribute to ongoing conversations related to TE Hub content development, and to receive notification of upcoming 'Hub Update' discussions and notes.

2. Register on the TE Hub wiki, using your ORCID iD (https://orcid.org/). Though this is not required in order to view TE Hub content, it will enable your future contribution of content by editing appropriate individual wiki pages.

Authors' contributions

TAE: Vision; collected resources and generated content on TE Hub; edited the manuscript. TH: Vision; contributed content to the TE Hub wiki; wrote first draft of manuscript and managed all future drafts. RH: Vision; managed the wiki; edited the manuscript. HQ: Vision; initiated, organized, and animated regular meetings of the TE Hub; edited the manuscript. AS: Vision; organized collaboration; edited the manuscript. TJW: Vision; organized collaboration; established and hosted web site and wiki server; wrote draft text of the manuscript.

Availability of data and materials

Not applicable.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Author details

[1]Centre for Biodiversity Genomics, University of Guelph, Guelph, ON, Canada. [2]Faculty of Biology, Technische Universität Dresden, 01069 Dresden, Germany. [3]Institute for Systems Biology, Seattle, WA, USA. [4]Université Paris-Saclay, INRAE, URGI, 78026 Versailles, France. [5]School of Biological Sciences, University of East Anglia, Norwich, UK. [6]Department of Organismal Biology, Uppsala University, Uppsala, Sweden. [7]Department of Computer Science, University of Montana, Missoula, MT, USA.

References

1. Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, et al. Ten things you should know about transposable elements. Genome Biol. 2018;19:199.
2. Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. Annu Rev Genet. 2020;54:539–61.
3. Keith Slotkin R. The case for not masking away repetitive DNA. Mob DNA. 2018;9:1–4.
4. Oliver KR, Greene WK. Transposable elements: powerful facilitators of evolution. Bioessays. 2009;31:703–14.
5. Rebollo R, Romanish MT, Mager DL. Transposable elements: an abundant and natural source of regulatory sequences for host genes. Annu Rev Genet. 2012;46:21–42.
6. Chang C-H, Chavan A, Palladino J, Wei X, Martins NMC, Santinello B, et al. Islands of retroelements are major components of *Drosophila* centromeres. PLoS Biol. 2019;17:e3000241.
7. O'Neill K, Brocks D, Hammell MG. Mobile genomics: tools and techniques for tackling transposons. Philos Trans R Soc Lond Ser B Biol Sci. 2020;375: 20190345.
8. Goerner-Potvin P, Bourque G. Computational tools to unmask transposable elements. Nat Rev Genet. 2018;19:688–704.
9. Hoen DR, Hickey G, Bourque G, Casacuberta J, Cordaux R, Feschotte C, et al. A call for benchmarking transposable element annotation methods. Mob DNA. 2015;6:13.
10. Bennetzen JL, Park M. Distinguishing friends, foes, and freeloaders in giant genomes. Curr Opin Genet Dev. 2018;49:49–55.
11. Abrams JM, Arkhipova IR, Belfort M, Boeke JD, Joan Curcio M, Faulkner GJ, et al. Meeting report: mobile genetic elements and genome plasticity 2018. Mob DNA. 2018;9:1–10.
12. Lesage P, Bétermier M, Bridier-Nahmias A, Chandler M, Chambeyron S, Cristofari G, et al. International Congress on Transposable Elements (ICTE 2016) in Saint Malo: mobile elements under the sun of Brittany. Mob DNA. 2016;7:1–8.
13. Ray DA, Paulat N, An W, Boissinot S, Cordaux R, Kaul T, et al. The 2019 FASEB science research conference on The Mobile DNA Conference: 25 years of discussion and research, June 23–28, Palm Springs, California, USA. FASEB J. 2019;33:11625–8.
14. Berrens R. Transposons Worldwide Slack Workspace [Internet]. 2018 [cited 2021 Apr 28]. Available from: https://transposonsworldwide.slack.com.
15. Bao W, Kojima KK, Kohany O. Repbase Update, a database of repetitive elements in eukaryotic genomes. Mob DNA. 2015;6:11.
16. Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–82.
17. Piégu B, Bire S, Arensburger P, Bigot Y. A survey of transposable element classification systems—a call for a fundamental update to meet the challenge of their diversity and complexity. Mol Phylogenet Evol. 2015;86: 90–109.
18. Storer J, Hubley R, Rosen J, Wheeler TJ, Smit AF. The Dfam community resource of transposable element families, sequence models, and genome annotations. Mob DNA. 2021;12:2.
19. Arkhipova IR. Using bioinformatic and phylogenetic approaches to classify transposable elements and understand their complex evolutionary histories. Mob DNA. 2017;8:19.
20. Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, et al. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. Nucleic Acids Res. 2011;39:D70–4.
21. Walker PJ, Siddell SG, Lefkowitz EJ, Mushegian AR, Dempsey DM, Dutilh BE, et al. Changes to virus taxonomy and the International Code of Virus Classification and Nomenclature ratified by the International Committee on Taxonomy of Viruses (2019). Arch Virol. 2019;164:2417–29.
22. Zhou Y, Lu C, Wu Q-J, Wang Y, Sun Z-T, Deng J-C, et al. GISSD: Group I Intron Sequence and Structure Database. Nucleic Acids Res. 2007;36: D31–7.
23. Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. ISfinder: the reference centre for bacterial insertion sequences. Nucleic Acids Res. 2006; 34:D32–6.

24. Amselem J, Cornut G, Choisne N, Alaux M, Alfama-Depauw F, Jamilloux V, et al. RepetDB: a unified resource for transposable element references. Mob DNA. 2019;10:6.
25. Flynn JM, Hubley R, Goubert C, Rosen J, Clark AG, Feschotte C, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020;117:9451–7.
26. Quesneville H, Bergman CM, Andrieu O, Autard D, Nouaud D, Ashburner M, et al. Combined evidence annotation of transposable elements in genome sequences. PLoS Comput Biol. 2005;1:166–75.
27. Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. Nat Protoc. 2020;15: 3745–76.

Publisher's Note