



Original article

DPTedb, an integrative database of transposable elements in dioecious plants

Shu-Fen Li¹, Guo-Jun Zhang², Xue-Jin Zhang¹, Jin-Hong Yuan¹,
Chuan-Liang Deng¹, Lian-Feng Gu^{3,*} and Wu-Jun Gao^{1,*}

¹College of Life Sciences, Henan Normal University, Xinxiang 453007, China, ²School of Basic Medical Sciences, Xinxiang Medical University, Xinxiang 453003, China and ³Basic Forestry and Proteomics Center, Haixia Institute of Science and Technology (HIST), Fujian Agriculture and Forestry University, Fuzhou 350002, China

*Corresponding author: Tel: +86-0373-3326340; Fax: +86-0373-3329102; Email: gaowujun1@163.com and Correspondence may also be addressed to: Lian-Feng Gu, Tel: +86-591-83590305; Fax: +86-591-86392267. Email: lfgu@fafu.edu.cn

Citation details: Li,S.-F., Zhang,G.-J., Zhang,X.-Z. *et al.* DPTedb, an integrative database of transposable elements in dioecious plants. *Database* (2016) Vol. 2016: article ID baw078; doi:10.1093/database/baw078

Received 29 December 2015; Revised 4 April 2016; Accepted 22 April 2016

Abstract

Dioecious plants usually harbor ‘young’ sex chromosomes, providing an opportunity to study the early stages of sex chromosome evolution. Transposable elements (TEs) are mobile DNA elements frequently found in plants and are suggested to play important roles in plant sex chromosome evolution. The genomes of several dioecious plants have been sequenced, offering an opportunity to annotate and mine the TE data. However, comprehensive and unified annotation of TEs in these dioecious plants is still lacking. In this study, we constructed a dioecious plant transposable element database (DPTedb). DPTedb is a specific, comprehensive and unified relational database and web interface. We used a combination of *de novo*, structure-based and homology-based approaches to identify TEs from the genome assemblies of previously published data, as well as our own. The database currently integrates eight dioecious plant species and a total of 31 340 TEs along with classification information. DPTedb provides user-friendly web interfaces to browse, search and download the TE sequences in the database. Users can also use tools, including BLAST, GetORF, HMMER, Cut sequence and JBrowse, to analyze TE data. Given the role of TEs in plant sex chromosome evolution, the database will contribute to the investigation of TEs in structural, functional and evolutionary dynamics of the genome of dioecious plants. In addition, the database will supplement the research of sex diversification and sex chromosome evolution of dioecious plants.

Database URL: <http://genedenovoweb.ticp.net:81/DPTedb/index.php>

Introduction

Transposable elements (TEs) are DNA elements that are capable of moving from one place in the genome to another. They contribute greatly to eukaryotic genomes, particularly plant genomes, and can account for up to 90% of the genome size in several plant species (1, 2). TEs are classified into two classes based on their mode of transposition. Class I elements are also known as retrotransposons that propagate via an RNA intermediate. These elements use a ‘copy and paste’ mechanism to insert themselves into a new location in the genome. Class II elements include DNA transposons, which move through a direct ‘cut-and-paste’ mechanism (3). The two classes can be further subdivided into orders, superfamilies and then families based on structural features. Retrotransposons are further divided into two groups, namely, long terminal repeat (LTR) elements, such as members of the Ty1/Gypsy and Ty3/Copia superfamilies, and the non-LTR elements that lack LTR sequences (e.g. long interspersed nuclear elements or LINES, short interspersed nuclear elements or SINES) (3). DNA transposons are further classified into three main subclasses, namely, terminal inverted repeats (TIRs), Helitrons and Mavericks (4).

Contrary to their initial portrayal as ‘junk’ DNA, increasing investigations suggested that TEs play vital roles in the genomes they occupy, such as impacting the structure and function of the host genome (5), influencing the evolutionary trajectories of their host (6), regulating gene expression of host genes (7) and donating new genes to the host genome (8). Recent studies have suggested that TEs may also be main contributors in diversification and evolution of sex chromosomes (9–11). In contrast to those of mammals and birds, the sex chromosomes of plants have evolved much more recently. Such species with ‘young’ sex chromosomes offer a unique opportunity to elucidate the mechanism of the very early stages of sex chromosome evolution (12, 13). TEs have been observed to accumulate in the sex chromosomes of a number of dioecious plant species, such as *Carica papaya* (14–16), *Silene latifolia* (17–20), *Rumex acetosa* (21), *Cannabis sativa* (22, 23) and *Bryonia dioica* (24). The identification of TEs in the genome of dioecious plants will facilitate the investigation of the influence of TEs on the structural, functional and evolutionary dynamics of the sequenced genomes, particularly on the origin and evolution of sex chromosomes.

More plant genomes, including several dioecious plants, such as *C. papaya* (25), *C. sativa* (26), *Populus trichocarpa* (27), *Morus notabilis* (28) and *Phoenix dactylifera* (29, 30), have been sequenced with the development of next-generation sequencing. In addition, the genomes of the dioecious plant *Asparagus officinalis* (31) and *S. latifolia*

(32) have been sequenced and partially assembled. However, the TE annotation of these genomes is incomplete and is based on different methods. For better usage and comparison of TEs in dioecious plants, comprehensive and unified annotation of TEs in these plants is necessary. A number of databases concerning TEs are currently available. These databases can be divided into two main types. One emphasizes the analysis and classification of TEs in the context of the tree of life, such as GyDB (33) and Repbase (34). The other focuses on the identification and characterization of TEs in a specific species, such as BmTEdb (35) and MnTEdb (36). However, no databases concerning the TEs of dioecious plants exist at present.

In this study, we used a combined method to identify, classify and annotate TEs in the genomes of sequenced dioecious plants. All identified TEs were organized and deposited in a dioecious plant transposable element database (DPTEdb). Useful tools were also integrated for the analysis of TEs. We believe that DPTEdb can be used to study the origin, amplification and evolutionary dynamics of TEs in dioecious plants. DPTEdb establishes a foundation for further comparative analysis among different dioecious species to decipher the roles of TEs on plant sex chromosome evolution.

Construction and Content of the Database

System implementation

The server of DPTEdb was constructed using Linux openSUSE 13.1, Apache 2.4.10, MySQL Server 5.5.33-MariaDB and Perl v5.18.1/PHP 5.4.20. All TE data and information were stored in MySQL tables for efficient management, search and display. Common Gateway Interface programs were mainly developed using Perl, JavaScript and PHP programming languages. The JBrowse Genome Browser that was built with HTML5 and JavaScript was used to manipulate and display the genome coordinates of TEs in the assemblies of eight dioecious plants in the DPTEdb (37).

Data sources

The DPTEdb houses the information on TEs from eight dioecious plants, including *C. papaya*, *C. sativa*, *P. trichocarpa*, *M. notabilis*, *P. dactylifera*, *A. officinalis*, *Humulus scandens* and *S. latifolia*. The downloaded hyperlink for the assembly genome sequences of *C. papaya*, *C. sativa*, *P. trichocarpa*, *M. notabilis*, *P. dactylifera* and *S. latifolia* are listed in Table 1. The partial genomes of *A. officinalis* and *H. scandens* were assembled in our lab from Illumina sequencing data.

Table 1. List of dioecious plant species analyzed in this study

Plant species	URL	References
<i>C. papaya</i>	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000150535.1_Papaya1.0	(25)
<i>C. sativa</i>	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000230575.1_canSat3	(26)
<i>P. trichocarpa</i>	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000002775.2_Poptr2_0	(27)
<i>M. notabilis</i>	http://morus.swu.edu.cn/morusdb/	(28)
<i>P. dactylifera</i>	ftp://ftp.ncbi.nlm.nih.gov/genomes/all/GCA_000413155.1_DPv01	(29, 30)
<i>A. officinalis</i>	http://datadryad.org/resource/doi:10.5061/dryad.s71t4	(31)
<i>H. scandens</i>	Our own data	unpublished
<i>S. latifolia</i>	http://w3lamc.umbr.cas.cz/lamc/?page_id=8 , respectively	(32)

Identification of TEs in the eight dioecious plant genomes

TE libraries of the eight dioecious plants were generated using three different approaches. (1) Signature-based identification of TEs. For retrotransposons, LTR_FINDER (v1.05) (38) and MGEScan-nonLTR (v2) (39) programs were used with default parameters to search against the assemblies of the eight dioecious plant genomes to identify the LTR and non-LTR retrotransposons, respectively. HelitronScanner (40), which was developed based on the local combinational variable algorithm (41), was used to detect Helitron transposons with default parameters. For MITE transposons, MITEHunter (42) was used to search the assemblies with default parameters. (2) Similarity-based identification of TEs. The assemblies of the eight dioecious plant genomes were searched for further similarity-based identification of TEs against repeat databases. The TE sequences were detected using RepeatMasker (www.repeatmasker.org) to search against the Repbase database (34). Results with scores <250 or those with target coverage <40% were discarded. (3) *De novo* identification of TEs. *De novo* identification of TEs was performed using PILER (v1.0) (43), RepeatScout (v1.0.5) (44) and RepeatModeler (<http://www.repeatmasker.org/RepeatModeler.html>, version 1.0.7). First, the assemblies of the eight dioecious plants were analyzed to identify putative TEs using the three software programs. The putative TEs that have >90% sequence similarity to each other were then discarded. Finally, the putative TEs with >90% sequence similarity to another prediction (signature-based identification and similarity-based identification) were removed to reduce the redundancy of similar predictions.

Definition of superfamily and families of putative TEs

The putative TEs identified by the methods mentioned above were classified and annotated by comparing with Repbase using RepeatMasker, and the best hit target TE was selected as the superfamily of the analyzed TEs.

Furthermore, the putative TEs were categorized into families based on the 80–80–80 rule; that is, two TE elements were classified into the same family if they shared at least 80% of sequence identity in at least 80% of their coding or internal domain, within their terminal repeat region, or in both regions on segments longer than 80 bp (3).

Results

Identification of TEs in eight dioecious plants


Using the methods described earlier, a total of 31 340 TEs belonging to 3446 TE families were identified in the eight dioecious plant genomes. These TEs and families were organized into an easy-to-use web-based database, DPTedB; the composition of which is presented in Table 2. A total of 2458, 5622, 11 673, 7071 and 4038 TEs were detected from the entire genome sequences of *C. papaya*, *C. sativa*, *P. trichocarpa*, *M. notabilis* and *P. dactylifera*, respectively. Apparently, retrotransposons were more abundant than DNA transposons in these five species (Table 1). For the three plants with assemblies representing a minority of the genome, only 408, 35 and 35 TEs were identified from *A. officinalis*, *H. scandens* and *S. latifolia*, respectively. In the future, more TEs will be identified when the comprehensive genome information for these species is available.

Web interface and usage

DPTedB is an integrated dioecious plant TE database that provides a comprehensive platform to study TEs in dioecious plants, as well as materials for further studies on the genome evolution of dioecious plants. An efficient and user-friendly web-based user interface was built based on different classifications of TEs from eight dioecious plants. Users could easily browse and search for the TE information and perform various analyses using the analysis tools. Data download and export are also available. A top menu (Figure 1A) and a side menu (Figure 1B) were designed to

Table 2. Summary of identified TEs in eight dioecious plant assemblies

Class	Order	Superfamily	<i>C. papaya</i> Members/ families	<i>C. sativa</i> Members/ families	<i>P. trichocarpa</i> Members/ families	<i>M. notabilis</i> Members/ families	<i>P. dactylifera</i> Members/ families	<i>A. officinalis</i> Members/ families	<i>H. scandens</i> Members/ families	<i>S. latifolia</i> Members/ families	
Retrotransposons	LTR	<i>Caulimovirus</i>									
		<i>Copia</i>	184/43	1/1	6/5	3/1	24/2				
		<i>DIRS</i>	2/2	941/36	1557/60	1832/48	882/75		3/3		
		<i>ERV1</i>	6/5	5/5	52/18	9/8	14/9				
		<i>ERV4</i>		184/13		35/17	70/36				
		<i>ERVK</i>	2/1	7/7	10/8	1/1	8/3				
		<i>ERVL</i>			2/2	263/8	27/17				
		<i>Gypsy</i>	617/20	1230/31	5587/39	4/1	3/3		1/1		
		<i>Ngaro</i>		1/1	6/4	6/5	6/3				
		<i>Pao</i>	1/1	72/8	140/8	121/6	24/13				
		Unknown	1110/42	1110/202		1514/92	605/234			7/2	
		CR1						1/1			
		CRE								1/1	
		DRE		2/2							
		<i>I</i>					2/2	1/1	2/2		
		<i>Jockey</i>					19/19	145/17	58/52	12/12	
		L1	69/49	992/111	87/19						
L2											
<i>Penelope</i>		1/1			1/1						
R2											
Rex					29/1		1/1				
RTE											
<i>RTE-BovB</i>							4/4	1/1			
<i>Tad1</i>							1/1				
Subtotal DNA transposons	TIR	<i>Academ</i>	1991/163	4547/419	7452/168	6021/260	2816/483	70/64	12/12	15/10	
		CMC	2/2	20/17	5/2	2/2	7/7		13/13		
		<i>bAT</i>	1/1	34/27	17/14	6/5	19/17		26/26	9/9	
		MULE	1/1	73/66	5/5	5/5	22/14		34/33		
		<i>Novosib</i>									
		<i>p</i>				1/1	1/1				
		<i>PIF-Harginger</i>	3/2	26/26	3/3	7/7	47/43		7/7	1/1	
		<i>TcMar</i>	1/1	3/3	3/2	1/1	31/30		1/1	8/8	
		Unknown			2770/136				2/2		
		MITE	2/2	122/70	78/41	150/60	103/65		76/26		
		<i>Helitron</i>	457/152	797/158	1340/177	878/77	991/299		179/175	5/5	
			467/161	1075/367	4221/380	1050/158	1222/477		338/283	23/23	
			2458/324	5622/786	11 673/548	7071/418	4038/960		408/347	35/35	
		Subtotal									
		Total									



DPTedb: Dioecious Plants Transposable Elements Database

[Home](#)
[Browse](#)
[Tools](#)
[JBrowse](#)
[Search](#)
[Information](#)
[Contact](#)
[Announcement](#)
[Helps](#)

A

Species Database

Carica papaya

Cannabis sativa

Populus trichocarpa

Morus notabilis

Phoenix dactylifera

Asparagus officinalis

Humulus scandens

Silene latifolia

Tools

Blast

Hmm

GetORF

JBrowse

Cut sequence

Links

Rebase

Plant Repeat DB

RepeatMasker

TEfam

TE tools

GyDB

BmTEdb

Soytedb

SINEBase

P-MITE

LTR_FINDER

LTR_STRUC

RepeatModeler

RepeatScout

TE Displayer

MaizeTEdb

MnTEdb

LTR_STRUC

C Browse

Class	Order	Superfamily	Members	families
Retrotransposons	LTR	Copia	184	43
		DIRS	2	2
		ERV1	6	5
		ERVK	2	1
		Gypsy	617	20
		Pao	1	1
DNA transposons	LINE	Unknown	1110	42
		L1	69	49
		CMC	2	2
		hAT	1	1
		MULE	1	1
		PIF-Harbinger	3	2
DNA transposons	TIR	TcMar	1	1
		MITE	2	2
		Helitron	457	152
Total			2458	324

Superfamily	Family	Number	Sequence	Download
Gypsy	RLGypsy_1	592	View	↓
	RLGypsy_2	6	View	↓
	RLGypsy_3	2	View	↓
	RLGypsy_4	1	View	↓
	RLGypsy_5	1	View	↓
	RLGypsy_6	1	View	↓
	RLGypsy_7	1	View	↓
	RLGypsy_8	1	View	↓
	RLGypsy_9	1	View	↓
	RLGypsy_10	1	View	↓
	RLGypsy_11	1	View	↓

Download the whole sequences of *Carica papaya L.*

Tools

Blast

Hmm

GetORF

JBrowse

Cut sequence

Links

Rebase

Plant Repeat DB

RepeatMasker

TEfam

TE tools

GyDB

BmTEdb

Soytedb

SINEBase

P-MITE

LTR_FINDER

LTR_STRUC

RepeatModeler

RepeatScout

TE Displayer

MaizeTEdb

MnTEdb

LTR_STRUC

D Search

Please enter sequence ID:
Search TE sequences by inputting single or multiple sequence ID(separated by ; or -)
(e.g. DTCMC_1_1_aos,DTCMC_2_1_cpa).

RLCopia_1_1_aos,RLCopia_2_1_aos

Submit Reset

A total of 2 records

ID	Class	Order	Superfamily	family	length	chr	start	end	strand	view	download
RLCopia_1_1_aos	Retrotransposons	LTR	Copia	RLCopia_1	4929	scaffold203675	4891	9819	-	view	↓
RLCopia_2_1_aos	Retrotransposons	LTR	Copia	RLCopia_2	5076	scaffold65751	791	5866	+	view	↓
Total					10005						all

Please select the species:
species Asparagus officinalis

By family

Enter a key name:
Copia

searched in
superfamily

Search TE sequences by inputting a order, superfamily or family name e.g LTR.

Submit Reset

A total of 3 records

ID	Class	Order	superfamily	family	length	chr	start	end	strand	view	download
RLCopia_1_1_aos	Retrotransposons	LTR	Copia	RLCopia_1	4929	scaffold203675	4891	9819	-	view	↓
RLCopia_2_1_aos	Retrotransposons	LTR	Copia	RLCopia_2	5076	scaffold65751	791	5866	+	view	↓
RLCopia_3_1_aos	Retrotransposons	LTR	Copia	RLCopia_3	3253	scaffold213610	7	3259	+	view	↓
Total					13258						all

Tools

Blast

Hmm

GetORF

JBrowse

Cut sequence

Links

Rebase

Plant Repeat DB

RepeatMasker

TEfam

TE tools

GyDB

BmTEdb

Soytedb

SINEBase

P-MITE

LTR_FINDER

LTR_STRUC

RepeatModeler

RepeatScout

TE Displayer

MaizeTEdb

MnTEdb

LTR_STRUC

E The information of the common dioecious plants

No.	Family	Species	Sex chromosome	Sex determination
1	Liliaceae	<i>Asparagus officinalis</i>	Homomorphic, pair 5	Dominant Y

Tools

Blast

Hmm

GetORF

JBrowse

Cut sequence

Links

Rebase

Plant Repeat DB

RepeatMasker

TEfam

TE tools

GyDB

BmTEdb

Soytedb

SINEBase

P-MITE

LTR_FINDER

LTR_STRUC

RepeatModeler

RepeatScout

TE Displayer

MaizeTEdb

MnTEdb

LTR_STRUC

F Related papers and news

Hobza R, Kubat Z, Cegan R, et al. Impact of repetitive DNA on sex chromosome evolution in plants. *Chromosome Res*. 2015, 23: 561-570. 2015-12-29

Figure 1. Interface and some functional sections of DPTedb. (A) Top menu of DPTedb. (B) Side menu of DPTedb. (C) Browsing interface of database DPTedb. (D) Searching interface of DPTedb. (E) A table embedded in the DPTedb, containing the information of the commonly studied dioecious plants. (F) Related publications and news.

navigate the DPTEdb content. The top menu includes four main sections, namely, Browse, Tools, Search and Information. The side menu contains three sections, namely, Species Database, Tools and Links. In the homepage, we also provide a summary of this database and display news and publications related to dioecious plants (Figure 1F).

Browse

Users can browse in the top menu or species database in the side menu to browse and visualize the basic information of the TEs in a selected plant species through hyperlinks of the species name. For each species, the TEs identified were grouped into different superfamilies, which can be browsed using the hyperlinks provided in the website (Figure 1C). In the 'TE superfamily statistical information' page, clicking the corresponding entry of each family allows users to obtain the detailed information of this family, including classification, length, location and nucleotide sequences.

Search and download

In the search interface of DPTEdb, we provide two pathways for users, namely, sequence ID(s) and search keyword(s). Users can use a specific sequence ID to search the DPTEdb and find corresponding entries. When using a keyword to search the DPTEdb, users should select one species first and enter a keyword afterward (TE order, superfamily or family name) to locate the information of interest quickly. After searching, a new webpage will open to display all the matched results, which can be printed out as a tabular format output or downloaded by clicking the corresponding hyperlinks (Figure 1D). Users can download all TE sequences or TE sequences of interest by order, superfamily or family in a given species.

Tools

DPTEdb offers five sequence analysis tools (BLAST, GetORF, HMMER, Gbrower and Cut sequence) to facilitate users to analyze the TE data. Users can submit any query sequences to do BLASTN or tBLASTN against the DPTEdb for homology search (Figure 2A). Users can also analyze the potential open reading frame (ORF) of a query sequences using the GetORF tool (Figure 2B). HMMER is provided to facilitate the identification and classification of TEs. In this page, Hidden Markov model (HMM) profile of coding domains of retrotransposons were collected from previous studies. Users can use the ORFs obtained by GetORF to search against these TE HMM models using

HMMER package for the identification and classification of the query sequences (Figure 2C). Furthermore, users can obtain a sequence or sequences in a defined position using the Cut sequence tool (Figure 2D). The genome browser based on Jbrowse was also used to display the coordinates of TEs in the genomes of the eight dioecious plants in the DPTEdb. Users can click the Jbrowse tool in the top menu and select one species to browse on a large scale in a graph visualization interface (Figure 3A). Users can also conveniently view the detailed information of TEs by simply clicking the name of the TE in the graphic interface (Figure 3B). The information on reference sequences and TEs could be selectively shown on Jbrowse by clicking 'Select Tracks' button to set output items.

Information and links

In the information section, we provided a table that shows the sex chromosome and sex determination information of 65 commonly studied dioecious plants (Figure 1E). The information could help users understand these dioecious plants quickly. In the side menu, a number of links to other database and software websites relevant to DPTEdb were provided. Users can reach these websites easily by clicking the corresponding links (Figure 1B).

Discussion

An increasing number of studies have indicated that TEs may play important roles in the diversification and evolution of sex chromosomes (12, 45, 46). Identification of TEs in the genome of dioecious plants could establish a foundation for study of the roles of TEs played on the genome of the sequenced dioecious plant. Comparison of TEs and other repetitive sequences among different dioecious plants will facilitate the investigation of the contributions of TEs to the sex diversification and sex chromosome evolution (21, 47). We constructed a database called DPTEdb for better usage and comparative analysis of TEs in dioecious plants. DPTEdb is manually curated and dedicated to TE identification and classification in the genomes of dioecious plants using comprehensive and unified annotation approaches, serving as the initial TE data repository for dioecious plants. DPTEdb provides interactive access to the database content via a series of user interfaces and batch/bulk download opportunities. Several analysis tools were also impeded in the DPTEdb for convenient and efficient analysis of TEs, such as homology comparison using BLAST and classification of TEs using GetORF and HMMER. Furthermore, DPTEdb offers information on commonly studied dioecious plants, as well as news and

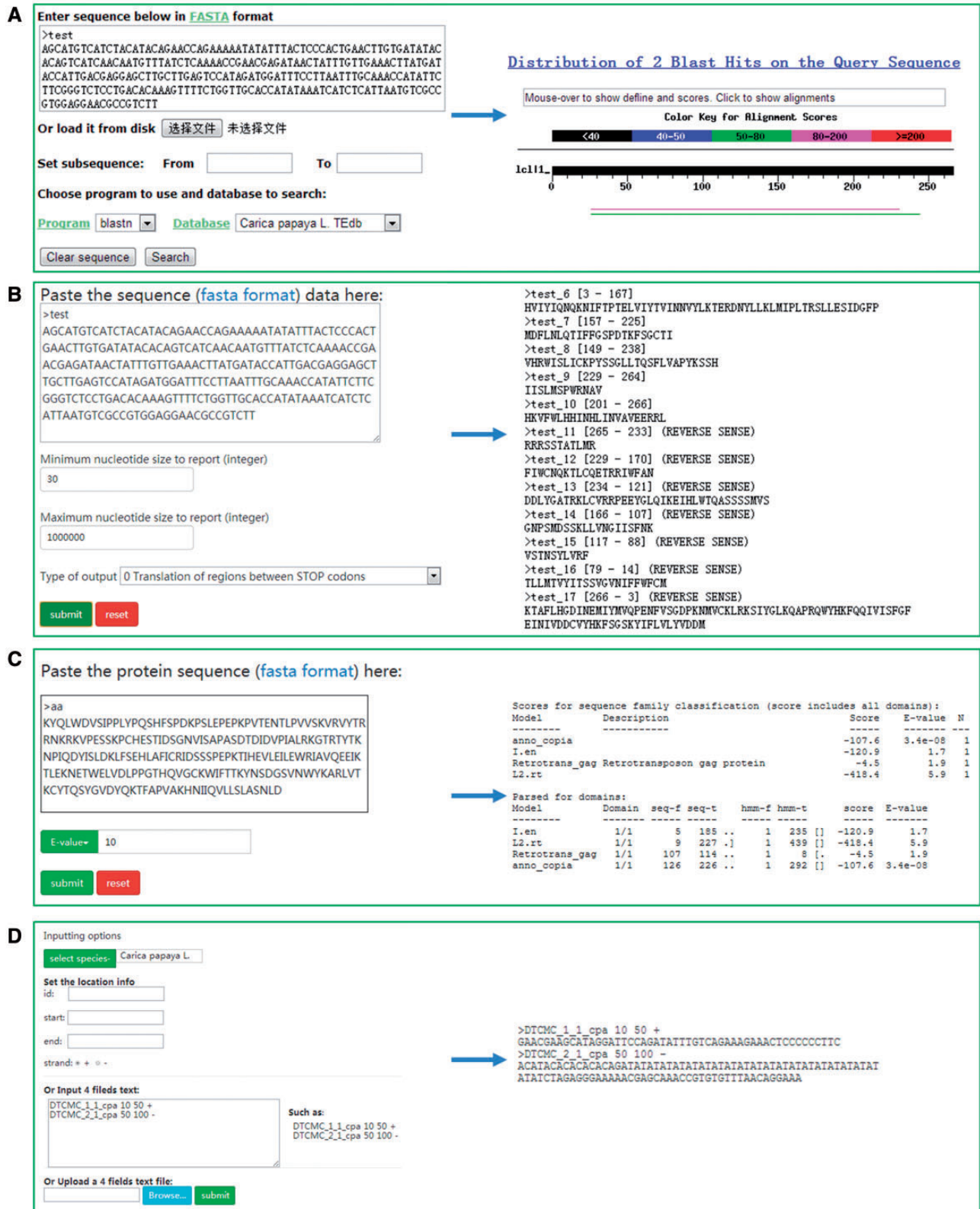


Figure 2. Different analysis tools provided in DPTedB. (A) BLAST interface of DPTedB. A sample of BLASTn results is shown. (B) GetORF interface of a test DNA sequence and snapshots of output results. (C) Interface of HMMER and a sample of protein sequence analyzed by HMMER. (D) Snapshots of the interface of the Cut sequence tool.

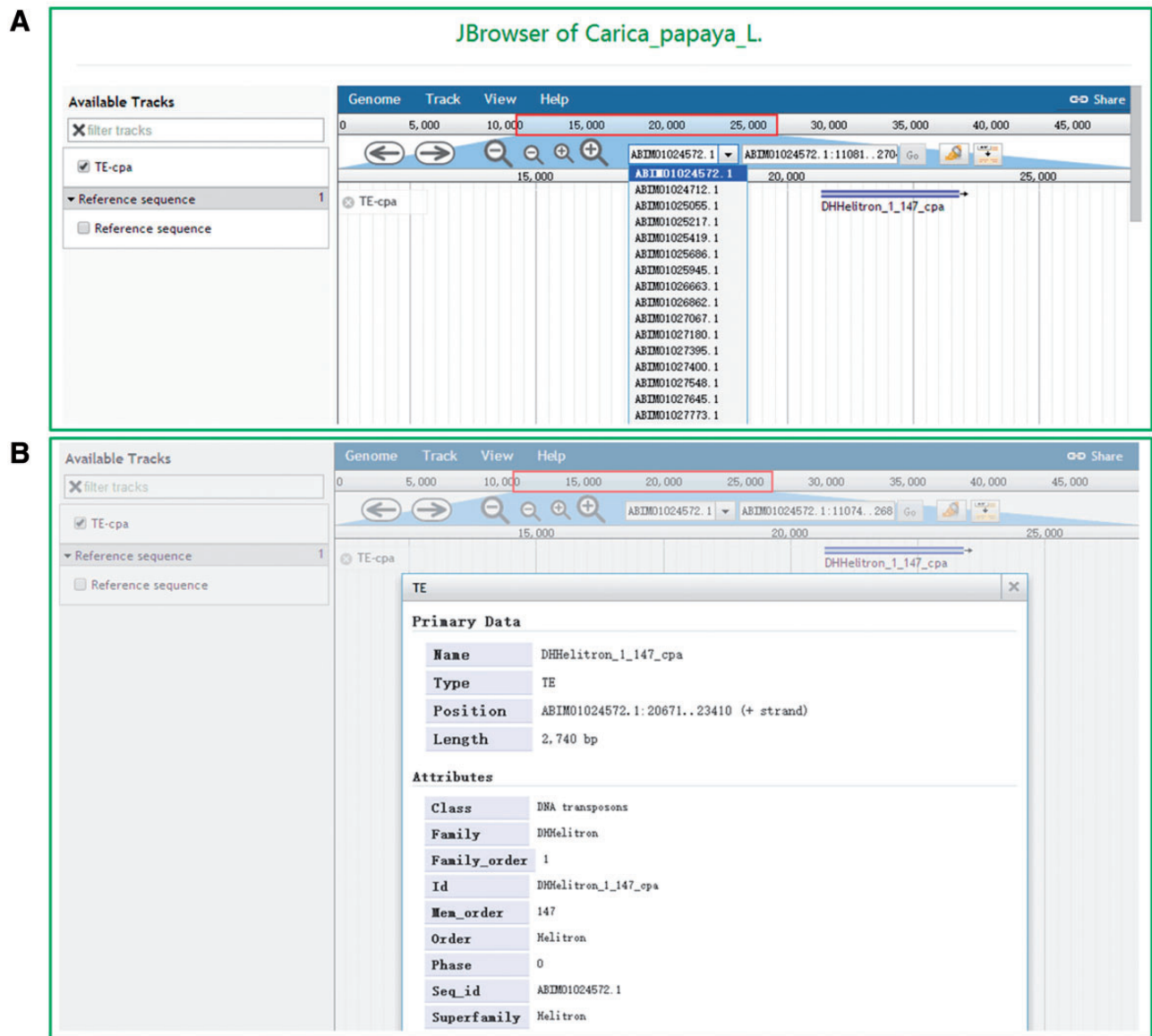


Figure 3. Snapshots of the Jbrowse tool of DPTEdb. (A) Genome sequence view of a region in scaffold ABIM01024572.1 from *C. papaya*. (B) Detailed information of the TE DHHelitron_1_147_cpa in the graphic interface.

publications related to dioecious plants. This can help users understand dioecious plants conveniently.

Submissions of new TE data to DPTEdb for dioecious plants are invited and highly encouraged. Among the eight included dioecious plants, a large number of TEs were identified and classified in the five species with draft genomes. In contrast, only a small number of TEs were detected in *A. officinalis*, *H. scandens* and *S. latifolia*, in which the assemblies only represent a minority of their entire genome. With increasing genome sequences of dioecious plants, we will improve and continuously update the DPTEdb for the TEs of the existing species, other dioecious plant species, and closely related hermaphroditic taxa.

Conclusion

We have generated a comprehensive TE database for dioecious plants called DPTEdb. The current DPTEdb includes 31 340 TEs from eight dioecious plants, along with classification information. DPTEdb allows users to search and browse TEs within the eight dioecious plants and permits batch download and analysis of TE sequences using the interfaces and analysis tools. Therefore, this database can contribute to the study of TEs in dioecious plants, and establishes a foundation for further comparative and evolutionary studies of sex chromosomes in plants.

Accessibility

The DPTEdb database is freely available at <http://genedeno.voweb.ticp.net:81/DPTEdb/index.php>.

Funding

This work was supported by grants from the National Natural Science foundation of China (31300202, 31470334) and a grant from Key Technologies R&D Program of Henan province (132102110195). Funding for open access charge: National Natural Science Foundation of China.

Conflict of interest. None declared.

References

- Sabot,F., Guyot,R., Wicker,T. *et al.* (2005) Updating of transposable element annotation from large wheat genomic sequences reveals diverse activities and gene associations. *Mol. Genet. Genomics*, 274, 119–130.
- Mehrotra,S. and Goyal,V. (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinformatics*, 12, 164–171.
- Wicker,T., Sabot,F., Hua-Van,A. *et al.* (2007) A unified classification system for eukaryotic transposable elements. *Nat. Rev. Genet.*, 8, 973–982.
- Feschotte,C. and Pritham,E.J. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annu. Rev. Genet.*, 41, 331–368.
- Bennetzen,J. and Wang,H. (2014) The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annu. Rev. Plant Biol.*, 65, 505–530.
- Guyot,R., Darre,T., Crouzillat,D. *et al.* (2015) Transposable element distribution, abundance and impact in genome evolution in the genus *Coffea*. *Plant Animal Genome*, XXIII.
- Hollister,J.D. and Gaut,B.S. (2009) Epigenetic silencing of transposable elements: a trade-off between reduced transposition and deleterious effects on neighboring gene expression. *Genome Res.*, 19, 1419–1428.
- Babushok,D.V., Ostertag,E.M. and Kazazian,H.H. Jr. (2007) Current topics in genome evolution: molecular mechanisms of new gene formation. *Cell. Mol. Life Sci.*, 64, 542–554.
- Ellison,C.E. and Bachtrog,D. (2013) Dosage compensation via transposable element mediated rewiring of a regulatory network. *Science*, 342, 846–850.
- Zhou,Q., Ellison,C.E., Kaiser,V.B. *et al.* (2013) The epigenome of evolving *Drosophila* neo-sex chromosomes: dosage compensation and heterochromatin formation. *PLoS Biol.*, 11, e1001711.
- Faber-Hammond,J.J., Phillips,R.B. and Brown,K.H. (2015) Comparative analysis of the shared sex-determination region (SDR) among salmonid fishes. *Genome Biol. Evol.*, 7, 1972–1987.
- Charlesworth,D. (2013) Plant sex chromosome evolution. *J. Exp. Bot.*, 64, 405–420.
- Charlesworth,D. (2015) Plant contributions to our understanding of sex chromosome evolution. *New. Phytol.*, 208, 52–65.
- Wang,J.P., Na,J.K., Yu,Q.Y. *et al.* (2012) Sequencing papaya X and Y^h chromosomes reveals molecular basis of incipient sex chromosome evolution. *Proc. Natl. Acad. Sci. USA*, 109, 13710–13715.
- VanBuren,R. and Ming,R. (2013) Dynamic transposable element accumulation in the nascent sex chromosomes of papaya. *Mob. Genet. Elements*, 3, e23462.
- VanBuren,R., Zeng,F., Chen,C. *et al.* (2015) Origin and domestication of papaya Y^h chromosome. *Genome Res.*, 25, 524–533.
- Matsunaga,S., Yagisawa,F., Yamamoto,M. *et al.* (2002) LTR retrotransposons in the dioecious plant *Silene latifolia*. *Genome*, 45, 745–751.
- Kejnovsky,E., Kubat,Z., Macas,J. *et al.* (2006) *Retand*: a novel family of gypsy-like retrotransposon harboring an amplified tandem repeat. *Mol. Genet. Genomics*, 276, 254–263.
- Cermak, T., Kubat, Z., Hobza, R. *et al.* (2008) Survey of repetitive sequence in *Silene latifolia* with respect to their distribution on sex chromosome. *Chromosome Res.*, 16, 961–976.
- Kralova,T., Cegan,R., Kubat,Z. *et al.* (2014) Identification of a novel retrotransposon with sex chromosome-specific distribution in *Silene latifolia*. *Cytogenet. Genome Res.*, 143, 87–95.
- Steflova,P., Tokan,V., Vogel,I. *et al.* (2013) Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY₁Y₂) in the dioecious plant *Rumex acetosa*. *Genome Biol. Evol.*, 5, 769–782.
- Sakamoto,K., Ohmido,N., Fukui,K. *et al.* (2000) Site-specific accumulation of a LINE-like retrotransposon in a sex chromosome of the dioecious plant *Cannabis sativa*. *Plant Mol. Biol.*, 44, 723–732.
- Sakamoto,K., Abe,T., Matsuyama,T. *et al.* (2005) RAPD markers encoding retrotransposable elements are linked to the male sex in *Cannabis sativa* L. *Genome*, 48, 931–936.
- Oyama,R.K., Silber,M.V. and Renner,S.S. (2010) A specific insertion of a solo-LTR characterizes the Y-chromosome of *Bryonia dioica* (Cucurbitaceae). *BMC Res. Notes*, 3, 166.
- Ming,R., Hou,S., Feng,Y. *et al.* (2008) The draft genome of the transgenic tropical fruit tree papaya (*Carica papaya* Linnaeus). *Nature*, 452, 991–997.
- van Bakel,H., Stout,J.M., Cote,A.G. *et al.* (2011) The draft genome and transcriptome of *Cannabis sativa*. *Genome Biol.*, 12, R102.
- Tuskan,G.A., DiFazio,S., Jansson,S. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, 313, 1596–1604.
- He,N., Zhang,C., Qi,X. *et al.* (2013) Draft genome sequence of the mulberry tree *Morus notabilis*. *Nat. Commun.*, 4, 2445.
- Al-Dous,E.K., George,B., Al-Mahmoud,M.E. *et al.* (2011) *De novo* genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nat. Biotechnol.*, 29, 521–527.
- Al-Mssallem,I.S., Hu,S., Zhang,X. *et al.* (2013) Genome sequence of the date palm *Phoenix dactylifera* L. *Nat. Commun.*, 4, 2274.
- Li,S.F., Gao,W.J., Zhao,X.P. *et al.* (2014) Analysis of transposable elements in the genome of *Asparagus officinalis* from high coverage sequence data. *PLoS One*, 9, e97189.
- Macas,J., Kejnovský,E., Neumann,P. *et al.* (2011) Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant *Silene latifolia*. *PLoS One*, 6, e27335.
- Llorens,C., Futami,R., Covelli,L. *et al.* (2011) The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.*, 39, D70–D74.

34. Jurka,J., Kapitonov,V.V., Pavlicek,A. *et al.* (2005) Repbase Update, a database of eukaryotic repetitive elements. *Cytogenet. Genome Res.*, 110, 462–467.
35. Xu,H.E., Zhang,H.H., Xia,T. *et al.* (2013) BmTEdb: a collective database of transposable elements in the silkworm genome. *Database*, 2013, bat055.
36. Ma,B., Li,T., Xiang,Z. *et al.* (2015) MnTEdb, a collective resource for mulberry transposable elements. *Database*, 2015, bav004.
37. Skinner,M.E., Uzilov,A.V., Stein,L.D. *et al.* (2009) JBrowse: a next-generation genome browser. *Genome Res.*, 19, 1630–1638.
38. Xu,Z. and Wang,H. (2007) LTR_FINDER: an efficient tool for the prediction of full-length LTR retrotransposons. *Nucleic Acids Res.*, 35, W265–W268.
39. Rho,M. and Tang,H. (2009) MGEScan-non-LTR: computational identification and classification of autonomous non-LTR retrotransposons in eukaryotic genomes. *Nucleic Acids Res.*, 37, e143.
40. Xiong,W., He,L., Lai,J. *et al.* (2014) HelitronScanner uncovers a large overlooked cache of Helitron transposons in many plant genomes. *Proc. Natl Acad. Sci. U S A*, 111, 10263–10268.
41. Xiong,W.W., Li,T.H., Chen,K. *et al.* (2009) Local combinational variables: an approach used in DNA-binding helixturn-helix motif prediction with sequence information. *Nucleic Acids Res.*, 37, 5632–5640.
42. Han,Y. and Wessler,S.R. (2010) MITE-Hunter: a program for discovering miniature inverted-repeat transposable elements from genomic sequences. *Nucleic Acids Res.*, 38, e199.
43. Edgar,R.C. and Myers,E.W. (2005) PILER: identification and classification of genomic repeats. *Bioinformatics*, 21, i152–i158.
44. Price,A.L., Jones,N.C. and Pevzner,P.A. (2005) De novo identification of repeat families in large genomes. *Bioinformatics*, 21, i351–i358.
45. Hobza,R., Kubat,Z., Cegan, R. *et al.* (2015) Impact of repetitive DNA on sex chromosome evolution in plants. *Chromosome Res.*, 23, 561–570.
46. Li,S.F., Zhang,G.J., Yuan,J.H. *et al.* (2016) Repetitive sequences and epigenetic modification: inseparable partners play important roles in the evolution of plant sex chromosomes. *Planta*, 243, 1083–1095.
47. Steflava,P., Hobza,R., Vyskot,B. *et al.* (2014) Strong accumulation of chloroplast DNA in the Y chromosomes of *Rumex acetosa* and *Silene latifolia*. *Cytogenet. Genome Res.*, 142, 59–65.