Genomics 109 (2017) 336–340

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

PlanTEnrichment: A tool for enrichment analysis of transposable elements in plants

Gökhan Karakülah [a,*], Aslı Suner [b]

[a] İzmir International Biomedicine and Genome Institute (iBG-İzmir), Dokuz Eylül University, 35340 İnciraltı, İzmir, Turkey
[b] Department of Biostatistics and Medical Informatics, Faculty of Medicine, Ege University, 35040 İzmir, Turkey

ARTICLE INFO

Article history:
Received 31 March 2017
Received in revised form 29 May 2017
Accepted 31 May 2017
Available online 1 June 2017

Keywords:
Transposable elements
Gene regulation
Plant genomes
Database
Bioinformatics

ABSTRACT

Transposable elements (TEs) are mobile DNA sequences that play roles in gene regulation, and have a potential to influence the expression of nearby genes by functioning as *cis*-regulatory sequences. However, bioinformatics tools facilitating analysis of the associations between TEs and nearby genes in plants are still lacking. We therefore reanalyzed the comprehensive annotation data of gene models and TEs of 11 plant species available in Ensembl Plants database, and built an up-to-date, unique tool called PlanTEnrichment, enabling enrichment analysis of TEs located within the upstream regions of a given gene list. PlanTEnrichment takes, for example, a group of differentially expressed genes under a particular biological condition as input and returns the list of TEs associated with those genes, along with their calculated enrichment scores and statistical significances. PlanTEnrichment is freely available at http://tools.ibg.deu.edu.tr/plantenrichment/ and is likely to substantially enhance our understanding of the role of TEs in diverse biological processes.

1. Introduction

Transposable elements (TEs) are mobile DNA sequences that have the ability to insert themselves into new sites of genome and are found abundantly in most eukaryotes, particularly in plants [1]. They dominate major proportions of several plant genomes. For example, >80% and >60% of maize and sorghum genomes comprise TEs, respectively [2]. TEs are usually divided into two major classes based on transposition intermediate: (i) class I elements (also called retrotransposons) act via an RNA intermediate using 'copy-and-paste' mechanism, whereas (ii) class II elements (also called DNA transposons) transpose in a 'cut-and-paste' fashion [3]. Retrotransposons are known to be the main contributors to the expansion of the size of the genome in eukaryotes [4,5], and are typically observed at higher levels than DNA transposons in plants [2].

In addition to the importance of genome evolution [6], TEs are associated with the regulation of gene expression (for review see [7,8]) by altering chromatin structure [9], by providing novel promoters and by functioning as *cis*-regulatory sequences [10]. The activity of TEs, in some cases, is inhibited by epigenetic silencing mechanisms, such as DNA methylation and histone modifications [9]. However, these epigenetic modifications in TEs can also influence the expression of both nearby [11] and host genes [12]. On the other hand, TE insertion at the upstream regions of genes may lead to a gain in novel binding sites (promoters) for transcription factors [13,14]. In a comprehensive study of human regulatory sequences, Jordan et al. have reported ~25% of the promoter regions of human genes contain TE-derived sequences, which indicates the significance of TEs in the regulation of transcription [13]. Additionally, the movement of TEs into new intergenic regions has potential to disrupt existing *cis*-regulatory sequences or to form novel ones in the genome. Anthony Studer and his colleagues have observed that a retrotransposon, *Hopscotch*, can act as an enhancer of *teosinte branched1* (*tb1*) gene in maize, by inserting itself into the regulatory region of the *tb1* locus [15].

Several publications have appeared in the last two decades documenting interesting examples of differential transcription and mobilization of TEs in response to stress in mammalian [16,17] and several plant species [18–23]. In a recent report by Makarevitch et al. [19], it has been revealed that in maize, as many as 20 TE families exhibit stress-responsive transcription and have the potential to influence the expression of nearby genes. Furthermore, one of the well-studied DNA transposons, *mPing*, has been shown to be triggered in response to stress conditions and rewired transcriptional regulatory network of rice [23]. In tobacco plant, the activation of the *Tnt1* retrotransposons by biotic and abiotic stress has been discovered, and their interactions with the host genome have been elucidated [24].

Given the large number of TEs located near gene loci in some plant species [25] and their potential to influence gene expression by providing novel regulatory sequences, TEs are crucial for rewiring regulatory networks that underlie diverse biological events, such as development and response to stress. So far, only Plant Transposable Element-related

* Corresponding author.
E-mail address: gokhan.karakulah@deu.edu.tr (G. Karakülah).

http://dx.doi.org/10.1016/j.ygeno.2017.05.008
0888-7543/

miRNA Database (PlanTE-MIR DB) has been established to facilitate the investigation of overlapping TE-microRNA gene pairs in 10 plants species [26]. To date, however, there is no available platform developed for the study of associations between TEs and nearby protein coding genes in plants. Here we report a regularly updated and simple yet functional analysis tool called PlanTEnrichment, which permits enrichment analysis of TEs located within the upstream regions of a given gene list. By an approach similar to that used in Gene Ontology (GO) term enrichment analysis [27], PlanTEnrichment helps to identify whether members of a group of genes tend to be significantly linked with certain TEs under a specific condition in plants.

2. Materials and methods for database construction

2.1. Collection and processing of annotation data

We downloaded gene model annotation data of 11 plant species (*Arabidopsis thaliana*, *Brachypodium distachyon*, *Glycine* max (soybean), *Hordeum vulgare* (barley), *Medicago truncatula*, *Oryza sativa* (rice), *Populus trichocarpa* (poplar), *Sorghum bicolor*, *Solanum lycopersicum* (tomato), *Triticum aestivum* (wheat), and *Zea mays* (maize)) from the ftp server of Ensembl Plants database (ftp://ftp.ensemblgenomes.org, release 34) [28] using an in-house shell script. We also made use of TE annotation of 11 species, which was provided in "repeat feature" and "repeat consensus" MySQL (https://www.mysql.com/) database tables of Ensembl Plants. These tables include: (i) genomic localization, (ii) class, (iii) type, and (iv) consensus repeat sequence (if available) information of each TE in the respective genome. All download tasks were performed with the wget command line utility of Linux. Using the genomic coordinates of genes and TEs, we detected and classified all TEs located within 1, 2.5 and 5 kb upstream of genes in their respective genomes. Thus, we found all possible TE related genes in each plant species included in the study, and stored their genomic features in our local database. We performed identification of co-localized gene-TE pairs by taking advantage of Structured Query Language (SQL) on a local MySQL database (v5.5.52). The workflow of the study is shown in Fig. 1.

2.2. Calculation of enrichment scores and their statistical significances

The unique functionality of the PlanTEnrichment tool is in permitting the identification of significantly enriched TEs associated with a given gene list. We calculate enrichment score (ES) of each TE observed within the upstream regions of a group of genes as follows. Let X denote a specific TE that is located within the upstream region of any gene in a given gene list and let Y denote all TEs that are located within the upstream region of all genes in the list. Then ES_X is described as follows:

$$ES_X = (a/b)/(c/d)$$

- where a is the total number of X observed within the upstream regions of the list of genes,
- b is the total number of X observed within the upstream regions of all genes in the genome,
- c is the total number Y observed within the upstream regions of the list of genes, and
- d is the total number of Y observed within the upstream regions of all genes in the genome.

The statistical significance of each ES is calculated with the Fisher's exact test. The *p*-value of ES_x can be computed by the following equation:

$$\text{The } p\text{-value of } ES_x = ((a+b)!(c+d)!(a+c)!(b+d)!)/a!b!c!d!N!$$

where N is the total of a, b, c and d. TEs with *p*-value < 0.05 and two or more fold ES can be considered as appropriate cut-off for determining

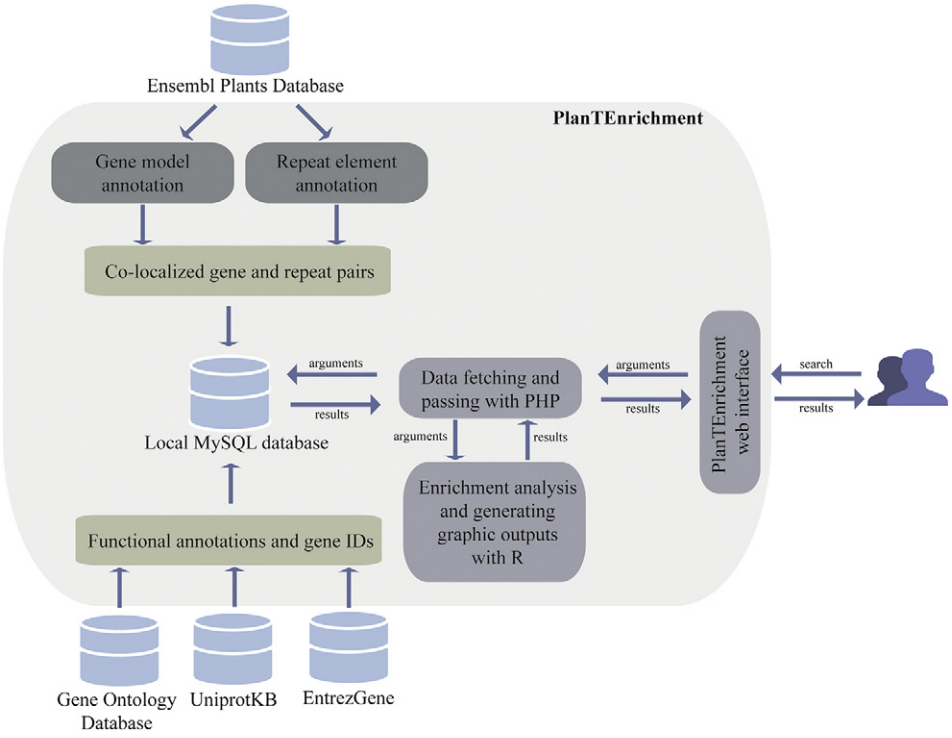

Fig. 1. Workflow diagram of PlanTEnrichment. Genomic locations of both genes and repeat elements were extracted from Ensembl Plants Database, and co-localized gene and repeat element pairs were identified. Functional gene annotations and gene identifiers of other external data sources, including Gene Ontology, UniprotKB and EntrezGene were also stored in our local MySQL database. A simple yet functional web interface was developed with PHP, HTML5 and CSS3 for retrieving the results of database queries and allowing enrichment analysis of repeat elements based on given gene lists by the users with R scripts; (http://php.net; http://www.mysql.com/; http://www.r-project.org./).

statistically significant TEs. However, a more stringent cut-off can also be applied by the users.

2.3. Development of database and user interfaces

PlanTEnrichment comprises comprehensive annotation data of TE related genes and their functions that are collected from several sources, such as GO [29] and UniProtKB [30]. We hosted all genomic features and gene annotations of TE related genes on a Linux server (CentOS Linux release 7.3(1611) and Apache v2.4.6). The user interfaces of PlanTEnrichment were built using Hyper Text Markup Language (HTML) 5 with responsive Pure CSS v0.6.2 modules (https://purecss.io/). Database queries were developed with SQL database programming language, and the communication between the user interfaces of PlanTEnrichment and the MySQL database was implemented in PHP v5.4.16 (http://www.php.net/). The user interfaces of PlanTEnrichment were also enhanced with JQuery (https://jquery.com/) AJAX (asynchronous HTTP) methods to provide its users with a simple and easy-to-use searching experience. ES and its statistical significance are calculated via execution of our in-house R v3.3.2 [31] scripts from PHP. Additionally, bar graphs that are generated by PlanTEnrichment are drawn in R statistical calculation environment using the ggplot2 package.

3. Database content and utility

3.1. The proportions of TE related genes in plant genomes

PlanTEnrichment was developed with the help of available public gene and TE annotation data from the Ensembl Plants database. Based on the co-localization analysis of plant TEs and genes, we observed that the proportions of genes that were found to be associated with TEs substantially varied across plant species (Table 1). Rice was found to be the species with the most abundant TE insertions at the upstream regions of its genes. We found 48.28%, 74.42% and 89.35% of rice genes include at least one repeat element within the 1, 2.5 and 5 kb upstream regions of their transcription start sites, respectively. However, *Arabidopsis thaliana*, *Brachypodium distachyon*, poplar and wheat were almost identical in terms of the proportions of the genes showing co-localizations with plant repeat elements. Overall, we have observed that a large number of genes in the genomes of 11 plant species include at least one TE within 5 kb upstream regions, which possibly indicates the significance of TEs in the regulation of a wide range of biological processes.

3.2. Enrichment analysis by gene IDs

PlanTEnrichment allows searching up to 1000 genes of plant species of interest at once by taking as input either multiple genomic coordinates in Chromosome:Start:End:Strand format or any of the following valid gene identifiers: (i) Ensembl gene IDs, (ii) Ensembl transcript IDs, (iii) EntrezGene IDs, or (iv) UniprotKB IDs. The main page of PlanTEnrichment tool offers a user-friendly web interface to its users so that one can perform his/her search efficiently and as simply as possible (Fig. 2a). First, the user chooses any of 11 plant species, and then defines the length of upstream sequence and gene identifier type from the pull down menus, respectively. At the final step, the user pastes his/her gene list or multiple genomic regions into the text area underneath the pull down menus and clicks the search button. When the search operation is performed, depending on the number of genes in the search list and the server workload, the retrieval of search results may take up to a few minutes.

In the result page of PlanTEnrichment, all TE names associated with the given gene list appear along with their class and superfamily information as well as corresponding ESs and *p*-values (Fig. 2b). Additionally, the tool provides an option to download results in text and graphic formats. A "Download repeat positions" link appearing in the result page makes available the download of genomic intervals of TEs and their associated genes. This interval information can be utilized, for instance, to quantify the abundance of the TEs of interest in RNA sequencing samples via incorporating with other genome arithmetic tools, such as BEDTools [32] and BEDOPS [33]. Providing interval information in text format is particularly important when the user compares expression levels of TEs from two or more groups. We also provided gene descriptions, UnipotKB IDs and descriptions, and associated GO terms of the given gene list via "the Download repeat positions" link in the result page. The values used in the calculation of ESs and *p*-values are downloadable and provided as a separate link via the "Download enrichment results" link. Enrichment results (up to 10 TEs) can also be downloaded in portable network graphics using the "See results as bar graph" link in the result page (Fig. 2c). Brief information regarding the tool and its usage, data sets and genome assemblies used in the study is provided in the tutorial page of PlanTEnrichment.

3.3. Enrichment analysis of heat stress responsive genes in maize

As an example, we tested PlanTEnrichment with a previously published gene set to examine whether our tool generates consistent results with the current literature. By using the genomic coordinates of differentially regulated TE associated genes under heat stress in maize [19] (Supplementary Table 1, sample list #3 of PlanTEnrichment) as input, we performed an enrichment analysis to detect potential heat stress related TEs. Our enrichment analysis found 48 TEs of which 4 (RLX_naiba_AC195481_139, RLX_etug_AC187099_1770, RLG_gyma_AC189750_2238 and RLC_bipide_AC205969_9058) were statistically significant ($p < 0.001$) and had >2 fold enrichment. These results agree with the findings from Makarevitch et al. [19], which have shown that naiba, etug, gyma families are associated with heat stress response. PlanTEnrichment also suggested one of the bipide family members to be a novel heat responsive TE. However, this finding needs to be investigated further and validated.

Table 1
The percentages of plant genes that have at least one TE within 1, 2.5 and 5 kb regions of their transcription start sites in the Ensembl Plants database (release 34) gene model and repeat element annotation.

Species	Number of Genes	1 kb (%)	2.5 kb (%)	5 kb (%)
Arabidopsis thaliana	32,833	11.98	26.81	43.05
Brachypodium distachyon	26,552	13.71	33.22	51.98
Glycine max	54,174	16.97	39.78	62.64
Hordeum vulgare	26,067	23.31	45.78	65.69
Medicago truncatula	54,073	38.31	67.77	84.96
Oryza sativa	91,080	48.28	74.42	89.35
Populus trichocarpa	41,377	13.73	35.12	55.36
Solanum lycopersicum	38,735	23.40	48.83	69.39
Sorghum bicolor	34,567	38.23	67.65	85.01
Triticum aestivum	114,428	11.98	30.59	52.13
Zea mays	44,474	22.13	38.47	48.48

4. Conclusions and future works

Herein, we developed a regularly updated unique analysis tool called PlanTEnrichment to facilitate investigation of the role of TEs under diverse biological conditions. By an approach similar to that used in GO term enrichment analysis, PlanTEnrichment takes, for example, a group of genes sharing a common feature (e.g. differentially expressed genes between normal and stress conditions) and returns a list of candidate TEs that have a potential to serve as novel regulatory sequences for nearby genes. To test the association of significantly enriched TEs with a certain condition, quantification of TE expression in samples should be considered as the next step. Therefore, in our future research we intend to concentrate on incorporating expression profiles of TEs, which will be collected from public RNA sequencing databases. We currently include 11 plant species in our database. However, PlanTEnrichment has a

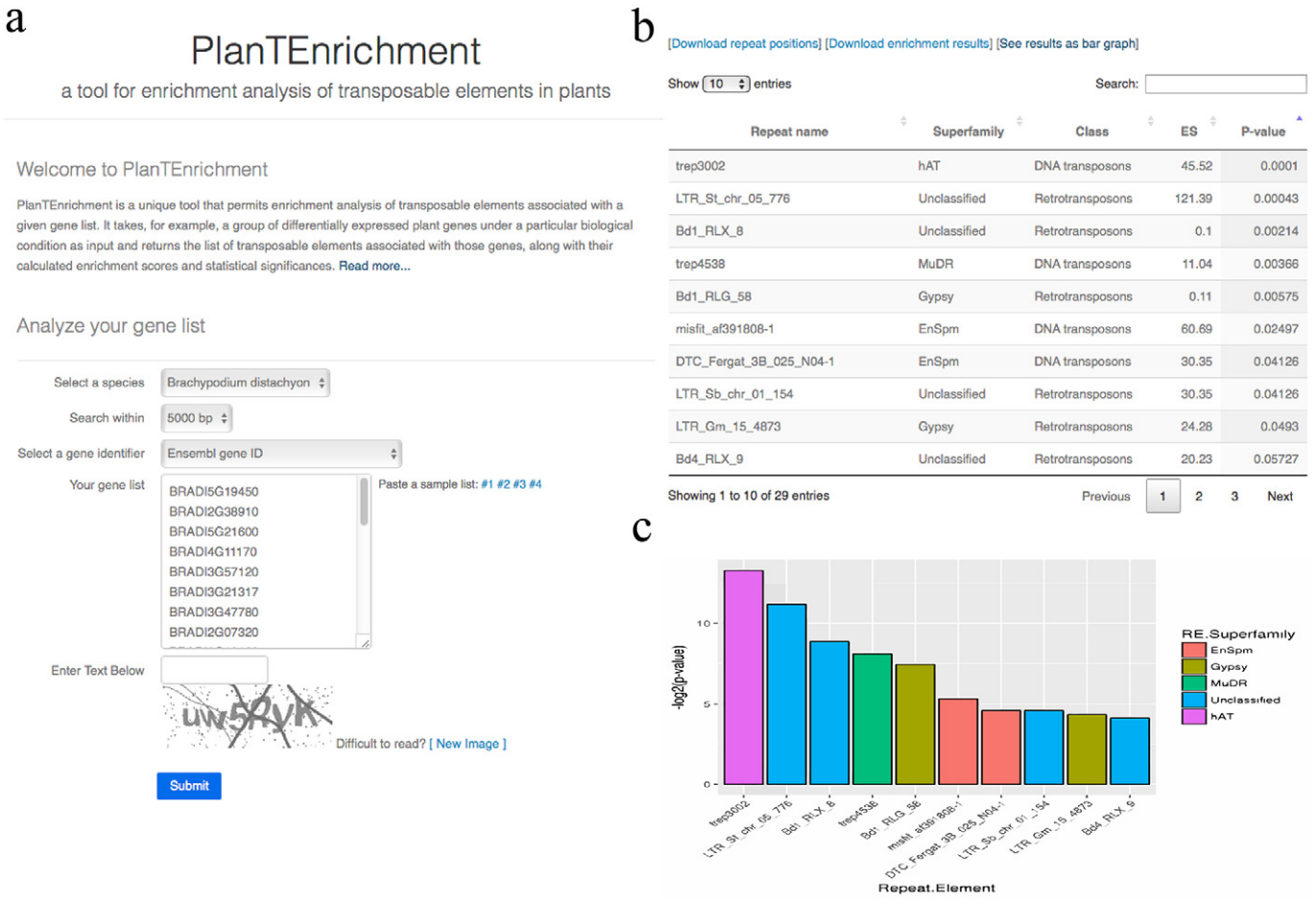

Fig. 2. Screenshots of PlanTEnrichment search and results pages. (a) Users can search up to 1000 genes at a time using the form fields in the home page of PlanTEnrichment. (b) Search results are provided in both searchable table format and tab delimited text format for further downstream analysis steps. (c) Statistical significances of TEs are also illustrated as bar graphs and can be downloaded as a portable network graphics (PNG) file.

modular and expandable architecture. Therefore it is open to grow, and we plan to increase the number of available species for the enrichment analysis. Our tool is freely available at http://tools.ibg.deu.edu.tr/plantenrichment/ and is likely to substantially enhance our understanding of the role of TEs in diverse biological processes.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.ygeno.2017.05.008.

Acknowledgements

The authors would like to thank their colleagues Prof. Dr. Bilge Karaçalı from İzmir Institute of Technology, İzmir, Turkey for his valuable insights and recommendations, and Mr. Alirıza Arıbaş and Specialist Ömür Dalan from Dokuz Eylül University for their technical assistance. The authors also would like to extend their thanks to İzmir Biomedicine and Genome Institute (iBG-İzmir), Dokuz Eylül University, İzmir, Turkey for its support in implementing the project.

References

[1] C. Biemont, C. Vieira, Genetics: junk DNA as an evolutionary force, Nature 443 (2006) 521–524.
[2] L.J. Kelly, I.J. Leitch, Exploring giant plant genomes with next-generation sequencing technology, Chromosom. Res. 19 (2011) 939–953.
[3] T. Wicker, F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A.H. Schulman, A unified classification system for eukaryotic transposable elements, Nat. Rev. Genet. 8 (2007) 973–982.
[4] A. Kumar, J.L. Bennetzen, Plant retrotransposons, Annu. Rev. Genet. 33 (1999) 479–532.
[5] C. Sun, D.B. Shepard, R.A. Chong, J. Lopez Arriaza, K. Hall, T.A. Castoe, C. Feschotte, D.D. Pollock, R.L. Mueller, LTR retrotransposons contribute to genomic gigantism in plethodontid salamanders, Genome Biol. Evol. 4 (2012) 168–183.
[6] N.V. Fedoroff, Presidential address. Transposable elements, epigenetics, and genome evolution, Science 338 (2012) 758–767.
[7] R.A. Elbarbary, B.A. Lucas, L.E. Maquat, Retrotransposons as regulators of gene expression, Science 351 (2016), aac7247. .
[8] C.D. Hirsch, N.M. Springer, Transposable element influences on gene expression in plants, Biochim. Biophys. Acta 1860 (2017) 157–165.
[9] R.K. Slotkin, R. Martienssen, Transposable elements and the epigenetic regulation of the genome, Nat. Rev. Genet. 8 (2007) 272–285.
[10] C. Feschotte, Transposable elements and the evolution of regulatory networks, Nat. Rev. Genet. 9 (2008) 397–405.
[11] M.R. Estecio, J. Gallegos, M. Dekmezian, Y. Lu, S. Liang, J.P. Issa, SINE retrotransposons cause epigenetic reprogramming of adjacent gene promoters, Mol. Cancer Res. 10 (2012) 1332–1342.
[12] T.N. Le, Y. Miyazaki, S. Takuno, H. Saze, Epigenetic regulation of intragenic transposable elements impacts gene transcription in *Arabidopsis thaliana*, Nucleic Acids Res. 43 (2015) 3911–3921.
[13] I.K. Jordan, I.B. Rogozin, G.V. Glazko, E.V. Koonin, Origin of a substantial fraction of human regulatory sequences from transposable elements, Trends Genet. 19 (2003) 68–72.
[14] G.J. Faulkner, Y. Kimura, C.O. Daub, S. Wani, C. Plessy, K.M. Irvine, K. Schroder, N. Cloonan, A.L. Steptoe, T. Lassmann, K. Waki, N. Hornig, T. Arakawa, H. Takahashi, J. Kawai, A.R. Forrest, H. Suzuki, Y. Hayashizaki, D.A. Hume, V. Orlando, S.M. Grimmond, P. Carninci, The regulated retrotransposon transcriptome of mammalian cells, Nat. Genet. 41 (2009) 563–571.
[15] A. Studer, Q. Zhao, J. Ross-Ibarra, J. Doebley, Identification of a functional transposon insertion in the maize domestication gene tb1, Nat. Genet. 43 (2011) 1160–1163.
[16] P.D. Mariner, R.D. Walters, C.A. Espinoza, L.F. Drullinger, S.D. Wagner, J.F. Kugel, J.A. Goodrich, Human Alu RNA is a modular transacting repressor of mRNA transcription during heat shock, Mol. Cell 29 (2008) 499–509.

[17] T.A. Allen, S. Von Kaenel, J.A. Goodrich, J.F. Kugel, The SINE-encoded mouse B2 RNA represses mRNA transcription in response to heat shock, Nat. Struct. Mol. Biol. 11 (2004) 816–821.
[18] S.R. Wessler, Turned on by stress. Plant retrotransposons, Curr. Biol. 6 (1996) 959–961.
[19] I. Makarevitch, A.J. Waters, P.T. West, M. Stitzer, C.N. Hirsch, J. Ross-Ibarra, N.M. Springer, Transposable elements contribute to activation of maize genes in response to abiotic stress, PLoS Genet. 11 (2015), e1004915. .
[20] T. Beguiristain, M.A. Grandbastien, P. Puigdomenech, J.M. Casacuberta, Three Tnt1 subfamilies show different stress-associated patterns of expression in tobacco. Consequences for retrotransposon control and evolution in plants, Plant Physiol. 127 (2001) 212–221.
[21] H. Ito, T. Yoshida, S. Tsukahara, A. Kawabe, Evolution of the ONSEN retrotransposon family activated upon heat stress in Brassicaceae, Gene 518 (2013) 256–261.
[22] E. Bucher, J. Reinders, M. Mirouze, Epigenetic control of transposon transcription and mobility in *Arabidopsis*, Curr. Opin. Plant Biol. 15 (2012) 503–510.
[23] K. Naito, F. Zhang, T. Tsukiyama, H. Saito, C.N. Hancock, A.O. Richardson, Y. Okumoto, T. Tanisaka, S.R. Wessler, Unexpected consequences of a sudden and massive transposon amplification on rice gene expression, Nature 461 (2009) 1130–1134.
[24] M.A. Grandbastien, C. Audeon, E. Bonnivard, J.M. Casacuberta, B. Chalhoub, A.P. Costa, Q.H. Le, D. Melayah, M. Petit, C. Poncet, S.M. Tam, M.A. Van Sluys, C. Mhiri, Stress activation and genomic impact of Tnt1 retrotransposons in Solanaceae, Cytogenet. Genome Res. 110 (2005) 229–241.
[25] P.S. Schnable, D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, P. Minx, A.D. Reily, L. Courtney, S.S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S.M. Rock, E. Belter, F. Du, K. Kim, R.M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S.M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M.J. Levy, L. McMahan, P. Van Buren, M.W. Vaughn, K. Ying, C.T. Yeh, S.J. Emrich, Y. Jia, A. Kalyanaraman, A.P. Hsia, W.B. Barbazuk, R.S. Baucom, T.P. Brutnell, N.C. Carpita, C. Chaparro, J.M. Chia, J.M. Deragon, J.C. Estill, Y. Fu, J.A. Jeddeloh, Y. Han, H. Lee, P. Li, D.R. Lisch, S. Liu, Z. Liu, D.H. Nagel, M.C. McCann, P. SanMiguel, A.M. Myers, D. Nettleton, J. Nguyen, B.W. Penning, L. Ponnala, K.L. Schneider, D.C. Schwartz, A. Sharma, C. Soderlund, N.M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T.K. Wolfgruber, L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J.L. Bennetzen, R.K. Dawe, J. Jiang, N. Jiang, G.G. Presting, S.R. Wessler, S. Aluru, R.A. Martienssen, S.W. Clifton, W.R. McCombie, R.A. Wing, R.K. Wilson, The B73 maize genome: complexity, diversity, and dynamics, Science 326 (2009) 1112–1115.
[26] A.P.R. Lorenzetti, G.Y.A. de Antonio, A.R. Paschoal, D.S. Domingues, PlanTE-MIR DB: a database for transposable element-related microRNAs in plant genomes, Funct. Integr. Genomics 16 (2016) 235–242.
[27] D.W. Huang, B.T. Sherman, Q. Tan, J.R. Collins, W.G. Alvord, J. Roayaei, R. Stephens, M.W. Baseler, H.C. Lane, R.A. Lempicki, The DAVID Gene functional classification tool: a novel biological module-centric algorithm to functionally analyze large gene lists, Genome Biol. 8 (2007) R183.
[28] P.J. Kersey, J.E. Allen, I. Armean, S. Boddu, B.J. Bolt, D. Carvalho-Silva, M. Christensen, P. Davis, L.J. Falin, C. Grabmueller, J. Humphrey, A. Kerhornou, J. Khobova, N.K. Aranganathan, N. Langridge, E. Lowy, M.D. McDowall, U. Maheswari, M. Nuhn, C.K. Ong, B. Overduin, M. Paulini, H. Pedro, E. Perry, G. Spudich, E. Tapanari, B. Walts, G. Williams, M. Tello-Ruiz, J. Stein, S. Wei, D. Ware, D.M. Bolser, K.L. Howe, E. Kulesha, D. Lawson, G. Maslen, D.M. Staines, Ensembl Genomes 2016: more genomes, more complexity, Nucleic Acids Res. 44 (2016) D574–D580.
[29] M. Ashburner, C.A. Ball, J.A. Blake, D. Botstein, H. Butler, J.M. Cherry, A.P. Davis, K. Dolinski, S.S. Dwight, J.T. Eppig, M.A. Harris, D.P. Hill, L. Issel-Tarver, A. Kasarskis, S. Lewis, J.C. Matese, J.E. Richardson, M. Ringwald, G.M. Rubin, G. Sherlock, Gene ontology: tool for the unification of biology. The Gene Ontology Consortium, Nat. Genet. 25 (2000) 25–29.
[30] M. Schneider, L. Lane, E. Boutet, D. Lieberherr, M. Tognolli, L. Bougueleret, A. Bairoch, The UniProtKB/Swiss-Prot knowledgebase and its plant proteome annotation program, J. Proteome 72 (2009) 567–573.
[31] R.C. Team, R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015, http://www.R-project.org 2016.
[32] A.R. Quinlan, BEDTools: The Swiss-Army Tool for Genome Feature Analysis, Current Protocols in Bioinformatics/Editoral Board, Andreas D. Baxevanis … [et al.], 47, 2014 11–34 11–12.
[33] S. Neph, M.S. Kuehn, A.P. Reynolds, E. Haugen, R.E. Thurman, A.K. Johnson, E. Rynes, M.T. Maurano, J. Vierstra, S. Thomas, R. Sandstrom, R. Humbert, J.A. Stamatoyannopoulos, BEDOPS: high-performance genomic feature operations, Bioinformatics 28 (2012) 1919–1920.