



# The impact of transposable elements on the structure, evolution and function of the rice genome

Roland Akakpo, Marie-Christine Carpentier, Yue Ie Hsing, Olivier Panaud

## ► To cite this version:

Roland Akakpo, Marie-Christine Carpentier, Yue Ie Hsing, Olivier Panaud. The impact of transposable elements on the structure, evolution and function of the rice genome. *New Phytologist*, 2020, 10.1111/nph.16356 . hal-02434127

**HAL Id: hal-02434127**

**<https://univ-perp.hal.science/hal-02434127>**

Submitted on 10 Jan 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



## Tansley insight

# The impact of transposable elements on the structure, evolution and function of the rice genome

Author for correspondence:  
Olivier Panaud  
Tel: +33 646460372  
Email: panaud@univ-perp.fr

Received: 2 September 2019  
Accepted: 5 November 2019

Roland Akakpo<sup>1</sup>, Marie-Christine Carpentier<sup>1</sup>, Yue Ie Hsing<sup>2</sup> and  
Olivier Panaud<sup>1,3</sup>

<sup>1</sup>Laboratoire Génome et Développement des Plantes, UMR 5096 CNRS/UPVD, Université de Perpignan, Via Domitia, 52 Avenue Paul Alduy, 66860 Perpignan Cedex, France; <sup>2</sup>Institute of Plant and Microbial Biology, Academia Sinica, 128, Section 2, Yien-chu-yuan Road, Nankang 115, Taipei, Taiwan; <sup>3</sup>Institut Universitaire de France, 1 Rue Descartes, 75231 Paris Cedex 05, France

## Contents

Summary	1	VI. A case study: GWAS of LTR-retrotransposon insertions associated with grain width in rice	5
I. Introduction	1	VII. Conclusion	5
II. Rice transposable elements	2	Acknowledgements	5
III. TE activity in <i>Oryza</i> lineage	2	References	5
IV. TE activity in cultivated rice	3		
V. TE-GWAS – a new strategy to unravel the functional impact of TEs at species level	4		

New Phytologist (2020)  
doi: 10.1111/nph.16356

**Key words:** functional genomics, model crop species, rice, transposable elements (TEs).

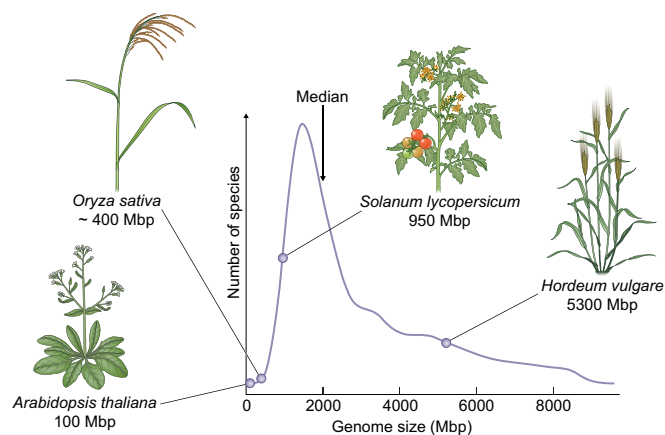
## Summary

Transposable elements (TEs) are ubiquitous in plants and are the primary genomic component of the majority of taxa. Knowledge of their impact on the structure, function and evolution of plant genomes is therefore a priority in the field of genomics. Rice, as one of the most prevalent crops for food security worldwide, has been subjected to intense research efforts over recent decades. Consequently, a considerable amount of genomic resources has been generated and made freely available to the scientific community. These can be exploited both to improve our understanding of some basic aspects of genome biology of this species and to develop new concepts for crop improvement. In this review, we describe the current knowledge on how TEs have shaped rice chromosomes and propose a new strategy based on a genome-wide association study (GWAS) to address the important question of their functional impact on this crop.

## I. Introduction

Transposable elements (TEs) are ubiquitous components of eukaryotic genomes. In flowering plants, their contribution to genome size variation has been documented in many studies, and it is now clearly established that transposition is the main factor responsible for such variation, besides polyploidy (Bennetzen & Wang, 2014). Fig. 1 shows the distribution of genome size for 6000

plant species (<https://cvalues.science.kew.org/>). Considering that the gene space in diploid genomes occupies *c.* 100–200 Mbp, the large extent of variation in this distribution, the peak observed at *c.* 700 Mbp, and the median at *c.* 2 Gbp suggest that the majority of plant genomes are composed mostly of TEs or TE-related sequences. This has been confirmed by the many plant genome sequencing projects completed so far (see Chen *et al.*, 2018 for review). For instance, while < 10% of the small genome of the



**Fig. 1** Distribution of genome size for 6000 plant species. Data obtained from the Plant C-value database (<https://cvalues.science.kew.org/>).

model species *Arabidopsis* is composed of TEs (AGI, 2000), TEs make up > 90% of the large genome of hexaploid wheat (Wicker *et al.*, 2018). The dynamics of the process through which TEs actually shape plant chromosomes are not so well understood. A comparative study among several plant genomes showed that most TE-related sequences arose from recent insertions (within the last 2 Myr, El Baidouri & Panaud, 2013) suggesting that older ones were eliminated from the genomes and therefore that there should exist a force counterbalancing TE-driven genomic expansions, which we formulated earlier in an ‘increase/decrease’ model (Vitte & Panaud, 2005) and which was recently addressed more specifically through mathematical models (Dai *et al.*, 2018). The dynamics of such processes (i.e. that of genomic amplification vs TE elimination) are yet to be fully understood.

In addition, the ubiquitous nature of TEs in plant genomes raises the overarching question of their biological impact and how transposition at large could contribute to plant biodiversity. TEs have long been considered mutagenic agents that impede gene function upon insertion into coding sequences, as exemplified by their early discovery as the causal agent of pigmentation loss in maize kernel by McClintock (1953). This negative view has since been challenged by several reports showing that transposition may in fact benefit an organism by regulating gene expression either by providing alternative promoters or novel cis-acting regulatory sites, or by acting as epigenetic mediators (Mirouze & Vitte, 2014; Hirsch & Springer, 2017). Moreover, two reports in mammals implicate TE amplification in the emergence of important biological novelty (i.e. placental pregnancy (Lynch *et al.*, 2011) and innate immunity (Chuong *et al.*, 2016)) indicating that transposition can be an important driver of eukaryote evolution through the rewiring of gene networks resulting in novel traits. In crops, there are few examples for which TEs are implicated in the variation of agronomic traits, such as fruit shape in tomato (Xiao *et al.*, 2008), and red pigmentation in apple (Zhang *et al.*, 2019). However, the extent to which TE-driven genetic variation could be exploited agronomically remains to be established for all crops.

Rice is the staple food for billions of people among the world’s poorest populations. This cereal has therefore been subjected to intense research efforts over recent decades in order to secure food

production world-wide. Consequently, a considerable amount of resources and knowledge have been generated for this species for all aspects of the biology of the plant. In particular, a high quality physical-map based genomic sequence of the rice variety ‘Nipponbare’ was the first crop genome to be released (IRGSP, 2005). As of today, high quality genome assemblies have been made available for five additional rice varieties (Song *et al.*, 2018). The release of the genomic sequence for 3000 rice genomes was an important milestone for rice research, giving access to the diversity of a crop at an unprecedented scale and providing a new approach for accelerating gene discovery through association studies (Rice genome project, 2014). Finally, the release of 10 genome assemblies for several wild relatives of rice allows the study of the evolution of a plant genome over 15 Myr (Stein *et al.*, 2018). In addition to providing rice biologists with new, efficient tools with which to develop future varieties, these resources have allowed us to decipher some basic aspects of TE-driven genome dynamics in rice, from long-term evolutionary aspects to their contribution to the genomic diversity of the crop following its domestication.

In this article, we first provide a synthetic view of these advances and then propose a new strategy to investigate the biological impact of TEs in rice, taking advantage of the genomic resources available for this model crop species.

## II. Rice transposable elements

As in any other plant species, the rice genome harbours most known types of TEs as defined by Wicker *et al.* (2007) – for example, LTR-retrotransposons (Hirochika *et al.*, 1996; Chaparro *et al.*, 2007), transposons (Panaud *et al.*, 2002), miniature inverted-repeat transposable elements (MITEs) (Jiang *et al.*, 2004), long interspersed elements (Komatsu *et al.*, 2003), short interspersed elements (Tsuchimoto *et al.*, 2008) and terminal-repeat retrotransposon in miniatures (Gao *et al.*, 2016). Altogether, TE-related sequences make up *c.* 40% of the rice genome (Jiang & Panaud, 2013). Interestingly, this contrasts with the scarcity of known active TEs in rice, where ‘activity’ is defined as their mobility under clearly defined and repeatable experimental procedures. These are three retrotransposons – *Tos17* (Hirochika *et al.*, 1996), *Lullaby* (Picault *et al.*, 2009) and *Karma* (Komatsu *et al.*, 2003) – that can be activated during *callus vitro* culture. Additionally, the transposon *nDart* can be activated when rice plants are treated with methylation inhibitor (Eun *et al.*, 2012), and *mPing* can be activated using laser irradiation (S. Li *et al.*, 2017). To date, there is no other TE family among the hundreds that the rice genome harbours for which such direct evidence of activity has been established. However, *Oryza* genomes harbour traces of strong transpositional activity for many families (Copetti *et al.*, 2015). This paradox can only be solved when the conditions of the activation of transposition *in natura* are fully elucidated.

## III. TE activity in *Oryza* lineage

The first comparative genomic analyses in the genus, focusing on several key loci such as *ADH* (Ammiraju *et al.*, 2007), *Hd1* (Sanyal *et al.*, 2010) and *Monoculm 1* (Lu *et al.*, 2009) revealed a good

**Box 1** The rationale for a transposable element (TE)-genome-wide association study (GWAS).

Since the late neolithic and development of agriculture some 10,000 years ago, farmers have moved many crops outside their centers of origin to new and often extreme environments. Only crops that could adapt to these adverse conditions were retained in production.

**In this regard, germplasm collections and in particular traditional varieties represent a reservoir (mostly untapped) of genetic factors involved in adaptation.**

TEs are known to be activated upon stress. Our model posits that transposition was triggered during the early phase of crop dissemination due to the stresses of domestication and adaptation to new environments.

Some TEs (e.g. the LTR-retrotransposons) harbour promoters that can be stress-inducible.

On the other hand, other TEs, such as MITEs, have been found to frequently contain transcription factor binding sites, which in some cases can wire new genes into stress-regulatory networks.

**Thus, while grown in new and adverse conditions, crops accumulated new copies of TEs, thereby spreading new stress-inducible promoters throughout their genomes.**

In some instances, the genomic amplification of new regulatory sequences **may have lead to the re-programming of gene networks, resulting in new phenotypes.**

We anticipate that some TE insertions are causative of adaptive traits and may not be tagged by SNPs, due to recent movement, and, thus, can **only be detected using the TIPs themselves as markers.**

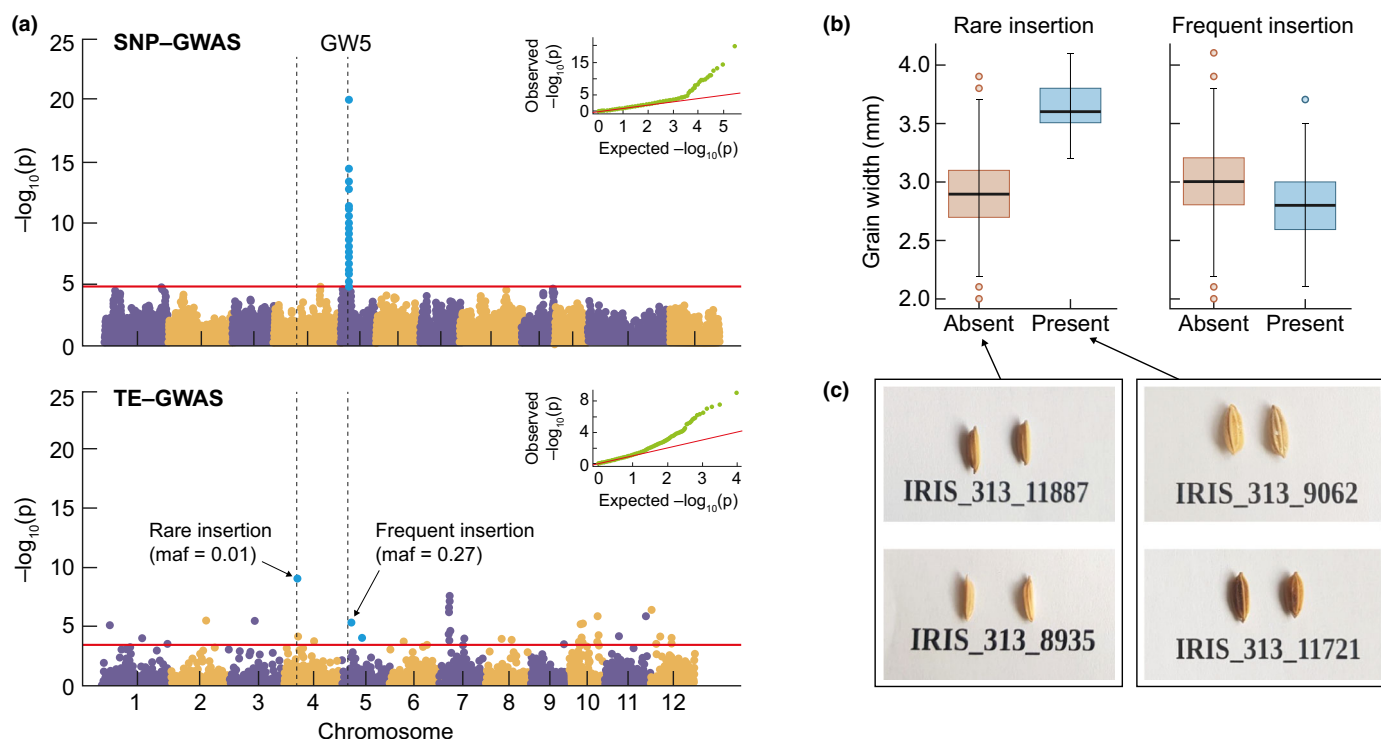
conservation of synteny among *Oryza* species, but with very limited correspondence in intergenic regions. A close examination of these loci confirmed that these regions were mostly comprised of TEs. These early reports therefore suggested that the rate of TE-driven genomic turnover was high, with effects observable at a moderate evolutionary timescale (< 3 Myr). Such turnover could only arise from the combined effect of successive waves of transposition that would quickly be eliminated from the genome through deletion or recombination, as posited by the increase/decrease model mentioned in the Introduction section. This was further illustrated by several genomic studies: Piegu *et al.* (2006) and Ammiraju *et al.* (2007) indeed showed that large and rapid genome size increases in *Oryza australiensis* and *Oryza granulata*, respectively, could be accounted for by the activity of LTR-retrotransposons. Interestingly, such dramatic genomic invasions occurred following speciation in both lineages, which provides indirect evidence that TEs may contribute to a large extent to lineage-specific genomic differentiation. This interpretation was further supported by the comparative analysis of eight A-genome *Oryza* species showing differential transpositional activity of several families among

lineages within the last million years (Zhang & Gao, 2017). More recently, the sequencing and assembly of the genome of 10 relatives of rice (Stein *et al.*, 2018) provided a better view of the evolutionary fate of TE-related sequences: a comparative survey of orthologous insertions of LTR-retrotransposons among the species of the A-genome type (i.e. the most closely related to cultivated rice) allowed estimation of the elimination rate of TE-related sequences through deletions. This estimation translated into a half-life of 1.2 Myr, which is much faster than what had been established earlier for *Drosophila* (14 Myr) and mammals (> 800 Myr). Together, these studies provide an explanation for earlier observations by providing an estimation of the parameters of the increase/decrease model. One consequence of this is that TE-driven genomic turnover in the *Oryza* lineage occurs at such high rate that it may lead to genomic diversification within species. When it comes to rice, this may be of primary importance if one considers that TE-associated structural variation may have a biological impact that could be exploited if at least part of it could be agronomically favorable, as we will show in the fifth section of this article.

#### IV. TE activity in cultivated rice

As mentioned above, only a few TE families were shown to be transpositionally active in cultivated rice. However, earlier studies clearly showed that TEs could contribute to the genetic diversity of the crop; for example, the mPing MITE was found to be polymorphic between 'Indica' and 'Japonica' varieties (Jiang *et al.*, 2003). The release of genomic sequences of 3000 rice varieties (Rice Genome Project, 2014) provided a unique opportunity to tentatively estimate at what level TEs actually contribute to the genomic diversity of the crop in the recent past. The identification of TE insertion polymorphisms (TIPs) from large datasets requires the development of suitable software that is fast and has a low false discovery rate (FDR). Several methods have been recently developed for this purpose. RELOCATE2 (Chen *et al.*, 2017) is based on parsing mapping files of paired-read Illumina sequences against a reference genome. It was used recently to track the transpositional activity of the mPing MITE in the 3000 rice genomes (Chen *et al.*, 2019). Carpentier *et al.* (2019) developed a new software package, TRACKPOSON, that was used to identify retrotransposon TIPs among the 3000 rice genomes (Rice genome project, 2014). This method is a two-step procedure consisting of first mapping the Illumina reads onto a consensus TE sequence and then mapping the un-mapped paired reads onto the reference rice genome. This method was found to be very fast, reliable and particularly well suited to large datasets such as the 3000 rice genomes. The authors showed that the three retrotransposon families that are known to be active *in vitro* (i.e. *Tos17*, *Lullaby* and *Karma*) are polymorphic in the cultivated gene pool, thus confirming that these TEs may in fact be active *in planta* and *in agro*. Surprisingly, the 32 TE families that were investigated in that study are also polymorphic. This may be a result of lineage sorting of polymorphisms that existed prior to domestication in the wild populations of rice progenitors, or of the fact that these families have been transpositionally active since domestication. In fact, the latter possibility is more parsimonious since most of the TIPs





**Fig. 2** Transposable element (TE)-genome-wide association study (GWAS) for grain width. (a) GWAS results (Manhattan and quantile–quantile plot) detected using SNP-GWAS and TE-GWAS for grain width. The solid red line represents FDR = 0.05. The dots above the red line are candidates for insertions/SNPs. The green dots represent the most significant associations. The top panel in (a) shows the SNP-GWAS using SNPs with minor allele frequencies (MAF) > 0.1. We identified a candidate SNP that is probably located downstream of the GW5 gene, a major gene for grain shape in rice (Weng *et al.*, 2008). The bottom left panel in (a) shows our TE-GWAS result using TIPs with minor frequencies (MF) > 0.01. Although it is not as saturated as the SNP map, the TIP dataset enabled the detection of significant associations (green dots) at two loci. One is in the GW5 region and was detected by an 'old' insertion of the LTR retrotransposon *rn215-225*, found in 306 accessions. The other one on chromosome 4, a 'recent' insertion of the same family, is found in only 12 accessions. (b) Box plot of grain width for the old and recent candidate insertions. The boxes represent the grain width variation between the first and third quartiles. The two whiskers represent the variation between the 'minimum' ( $Q1 - 1.5(\text{interquartile range (IQR)})$ ) and the 'maximum' ( $Q3 + 1.5(\text{IQR})$ ). Dots represent outliers. We observe that the varieties that hold the recent insertion exhibit a large grain (mean grain width c. 3.4 mm) compared to those without the insertion (mean grain width c. 2.8 mm), suggesting that this insertion has a strong effect on grain width. (c) Large (IRIS\_313\_9062; IRIS\_313\_11721) and narrow (IRIS\_313\_8935; IRIS\_313\_11887) rice grains.

identified are shared by very few rice varieties, most being private. More recently, Fuentes *et al.* (2019) conducted an exhaustive search of structural variations (SVs) among the 3000 rice genomes. For this, the authors benchmarked available software programs, combining several of them into a single pipeline that yielded over 63 million SVs of various origins, 8.7% of which were found to be TE-related. One conclusion of these studies is that transposition did indeed contribute to the diversity of cultivated rice. The next task is to investigate the putative biological impact of these polymorphisms.

## V. TE-GWAS – a new strategy to unravel the functional impact of TEs at species level

Until recently, gene discovery in crops relied on the use of map-based cloning ( Tanksley *et al.*, 1995). However, the advent of multi-genome sequencing projects has led to the development of new strategies based on genome-wide association surveys (Atwell *et al.*, 2010). In rice, the availability of such resources has fueled many such projects (Si *et al.*, 2016; X. Li *et al.*, 2017; Sales *et al.*, 2017),

demonstrating the efficiency of association studies in the crop. Given the availability of TIP data from the 3000 rice genomes for several TE families (Carpentier *et al.*, 2019), we propose a new strategy based on a genome-wide association survey in an attempt to unravel TE-associated genetic factors involved in agronomic traits in rice. This strategy is built on the rationale described in Box 1.

An important aspect of the TE-GWAS procedure is the risk of false discovery because of low allele frequency. The efficiency of GWAS is such that it is usually able to detect associations between traits and alleles with a minimum allele frequency (MAF) of at least 5% (Willard, 2013). However, the majority of TIPs in rice are found at a much lower frequency (Carpentier *et al.*, 2019). Consequently, we had to adapt our method to this peculiarity of TE insertions. First, we applied a correction for genetic structure (Thornsberry *et al.*, 2001). Next, we implemented four validation criteria in order to limit the risk of false positive detection: application of a stringent threshold of an FDR < 1% for significance, followed by the selection of the top 10% most significant candidate insertions; selection of insertions in genic regions (excluding TIPs in TE-rich, pericentromeric regions, for example); selection of insertions with

significant effects on the trait, contrasting the phenotypes of the varieties without the insertion against those with the insertion; selection of recent insertions, discarding those in the same regions as the ones detected with SNP-GWAS.

## VI. A case study: GWAS of LTR-retrotransposon insertions associated with grain width in rice

We implemented TE-GWAS in rice using TIPs from the *RN\_215* LTR-retrotransposon family among 1132 accessions of *Oryza sativa* 'Indica'. The association study was performed using grain width as the phenotype (<http://snp-seek.irri.org/>). The results are shown in Fig. 2. Interestingly, a major quantitative trait locus (QTL; *GW5*) that was previously characterized through map-based cloning (Weng *et al.*, 2008) can be identified through classical GWAS (i.e. using SNP data as genotype; Fig. 2a). The use of TE-GWAS allowed us to identify a significant negative correlation between TE insertion and grain width (Fig. 2a,b) on chromosome 5 at the *GW5* locus. This insertion is found at high frequency (0.27) in the rice population and is probably in linkage disequilibrium with the genetic factor that causes the trait variation. In this regard, the TIP on chromosome 5 can be used as a genetic marker, like the other SNPs from the same region that show a significant association with the trait. More interestingly, another insertion found on chromosome 4 is highly significant (Fig. 2a) and is positively correlated with grain width (Fig. 2b). There is no association between any SNP from the same region and the trait, suggesting that the insertion is recent, which is supported by the low frequency of the insertion (0.01). Although this result requires confirmation through wet-lab experiments (e.g. an expression study of nearby genes throughout panicle development and CRISPR-Cas9 removal of the *RN\_205* insertion from large seeded accessions), it nevertheless suggests that recent TE insertions may have led, to some extent, to the diversification of agronomic traits in rice.

## VII. Conclusion

The availability of genomic resources for large germplasm collections opens a new perspective that can help us to understand some basic aspects of plant genome dynamics, as well as for the development of new strategies for gene discovery and crop improvement. Transposable elements, although ubiquitous in eukaryotes, have long been overlooked as important factors in genetic variation that could be beneficial for crops. The development of new strategies of genome-wide association surveys may help unravel some of the hidden heritability for major agronomic traits. In this context, rice, with unprecedented resources, will pave the way to the development of such new strategies.

## Acknowledgements

This work was supported by Institut Universitaire de France, Agropolis Fondation (grant no. CFP 2015-02) and Academia Sinica thematic project AS-TP-107-L02. This study was conducted within the framework of the Laboratoires d'Excellences (LABEX) TULIP (ANR-10-LABX-41).

## ORCID

Olivier Panaud  <https://orcid.org/0000-0002-9292-503X>

## References

- AGI. 2000. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.
- Ammiraju JSS, Zuccolo A, Yu Y, Song X, Piegu B, Chevalier F, Walling JG, Ma J, Talag J, Brar DS *et al.* 2007. Evolutionary dynamics of an ancient retrotransposon family provides insights into evolution of genome size in the genus *Oryza*. *The Plant Journal* 52: 342–351.
- Atwell S, Huang YS, Vilhjálmsson BJ, Willems G, Horton M, Li Y, Meng D, Platt A, Tarone AM, Hu TT *et al.* 2010. Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. *Nature* 465: 627–631.
- Bennetzen JL, Wang H. 2014. The contributions of transposable elements to the structure, function, and evolution of plant genomes. *Annual Review of Plant Biology* 65: 505–530.
- Carpentier M-C, Manfroi E, Wei F-J, Wu H-P, Lasserre E, Llauro C, Debladis E, Akapko R, Hsing Y-I, Panaud O. 2019. Retrotranspositional landscape of Asian rice revealed by 3000 genomes. *Nature Communications* 10: 24.
- Chaparro C, Guyot R, Zuccolo A, Piégu B, Panaud O. 2007. RetriOryza: a database of the rice LTR-retrotransposons. *Nucleic Acids Research* 35: D66–D70.
- Chen F, Dong W, Zhang J, Guo X, Chen J, Wang Z, Lin Z, Tang H, Zhang L. 2018. The sequenced angiosperm genomes and genome databases. *Frontiers in Plant Science* 9: 418.
- Chen J, Lu L, Benjamin J, Diaz S, Hancock CN, Stajich JE, Wessler SR. 2019. Tracking the origin of two genetic components associated with transposable element bursts in domesticated rice. *Nature Communications* 10: 641.
- Chen J, Wrightsman TR, Wessler SR, Stajich JE. 2017. RelocaTE2: a high resolution transposable element insertion site mapping tool for population resequencing. *PeerJ* 5: e2942.
- Chuong EB, Elde NC, Feschotte C. 2016. Regulatory evolution of innate immunity through co-option of endogenous retroviruses. *Science* 351: 1083–1087.
- Copetti D, Zhang J, El Baidouri M, Gao D, Wang J, Barghini E, Cossu RM, Angelova A, Maldonado LCE, Roffler S *et al.* 2015. RiTE database: a resource database for genus-wide rice genomics and evolutionary biology. *BMC Genomics* 16: 538.
- Dai X, Wang H, Zhou H, Wang L, Dvořák J, Bennetzen JL, Müller H-G. 2018. Birth and death of LTR-retrotransposons in *Aegilops tauschii*. *Genetics* 210: 1039–1051.
- El Baidouri M, Panaud O. 2013. Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution. *Genome Biology and Evolution* 5: 954–965.
- Eun C-H, Takagi K, Park K-I, Maekawa M, Iida S, Tsugane K. 2012. Activation and epigenetic regulation of DNA transposon nDart1 in rice. *Plant & Cell Physiology* 53: 857–868.
- Fuentes RR, Chebotarov D, Duitama J, Smith S, De la Hoz JF, Mohiyuddin M, Wing RA, McNally KL, Tatarinova T, Grigoriev A *et al.* 2019. Structural variants in 3000 rice genomes. *Genome Research* 29: 870–880.
- Gao D, Li Y, Kim KD, Abernathy B, Jackson SA. 2016. Landscape and evolutionary dynamics of terminal repeat retrotransposons in miniature in plant genomes. *Genome Biology* 17: 7.
- Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, Kanda M. 1996. Retrotransposons of rice involved in mutations induced by tissue culture. *Proceedings of the National Academy of Sciences, USA* 93: 7783–7788.
- Hirsch CD, Springer NM. 2017. Transposable element influences on gene expression in plants. *Biochimica et Biophysica Acta (BBA) – Gene Regulatory Mechanisms* 1860: 157–165.
- International Rice Genome Sequencing Project. 2005. The map-based sequence of the rice genome. *Nature* 436: 793–800.
- Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, McCouch SR, Wessler SR. 2003. An active DNA transposon family in rice. *Nature* 421: 163–167.
- Jiang N, Feschotte C, Zhang X, Wessler SR. 2004. Using rice to understand the origin and amplification of miniature inverted repeat transposable elements (MITEs). *Current Opinion in Plant Biology* 7: 115–119.

- Jiang N, Panaud O. 2013. Transposable element dynamics in rice and its wild relatives. In: Zhang Q, Wing RA, eds. *Plant genetics and genomics: crops and models. Genetics and genomics of rice*. New York, NY, USA: Springer, 55–69.
- Komatsu M, Shimamoto K, Kyojuka J. 2003. Two-step regulation and continuous retrotransposition of the rice LINE-type retrotransposon Karma. *Plant Cell* 15: 1934–1944.
- Li S, Xia Q, Wang F, Yu X, Ma J, Kou H, Lin X, Gao X, Liu B. 2017. Laser irradiation-induced DNA methylation changes are heritable and accompanied with transpositional activation of mPing in rice. *Frontiers in Plant Science* 8: 363.
- Li X, Guo Z, Lv Y, Cen X, Ding X, Wu H, Li X, Huang J, Xiong L. 2017. Genetic control of the root system in rice under normal and drought stress conditions by genome-wide association study. *PLoS Genetics* 13: e1006889.
- Lu F, Ammiraju JSS, Sanyal A, Zhang S, Song R, Chen J, Li G, Sui Y, Song X, Cheng Z *et al.* 2009. Comparative sequence analysis of MONOCULM1-orthologous regions in 14 *Oryza* genomes. *Proceedings of the National Academy of Sciences, USA* 106: 2071–2076.
- Lynch VJ, Leclerc RD, May G, Wagner GP. 2011. Transposon-mediated rewiring of gene regulatory networks contributed to the evolution of pregnancy in mammals. *Nature Genetics* 43: 1154–1159.
- McClintock B. 1953. Induction of instability at selected loci in maize. *Genetics* 38: 579–599.
- Mirouze M, Vitte C. 2014. Transposable elements, a treasure trove to decipher epigenetic variation: insights from Arabidopsis and crop epigenomes. *Journal of Experimental Botany* 65: 2801–2812.
- Panaud O, Vitte C, Hivert J, Muzlak S, Talag J, Brar D, Sarr A. 2002. Characterization of transposable elements in the genome of rice (*Oryza sativa* L.) using Representational Difference Analysis (RDA). *Molecular Genetics and Genomics* 268: 113–121.
- Picault N, Chaparro C, Piegu B, Stenger W, Formey D, Llauro C, Descombin J, Sabot F, Lasserre E, Meynard D *et al.* 2009. Identification of an active LTR retrotransposon in rice. *The Plant Journal* 58: 754–765.
- Piegu B, Guyot R, Picault N, Roulin A, Saniyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA *et al.* 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research* 16: 1262–1269.
- Rice Genome Project. 2014. The 3,000 rice genomes project. *GigaScience* 3: 7.
- Sales E, Viruel J, Domingo C, Marqués L. 2017. Genome wide association analysis of cold tolerance at germination in temperate japonica rice (*Oryza sativa* L.) varieties. *PLoS ONE* 12: e0183416.
- Sanyal A, Ammiraju JSS, Lu F, Yu Y, Rambo T, Currie J, Kollura K, Kim H-R, Chen J, Ma J *et al.* 2010. Orthologous comparisons of the Hd1 region across genera reveal Hd1 gene lability within diploid *Oryza* species and disruptions to microsynteny in Sorghum. *Molecular Biology and Evolution* 27: 2487–2506.
- Si L, Chen J, Huang X, Gong H, Luo J, Hou Q, Zhou T, Lu T, Zhu J, Shangguan Y *et al.* 2016. *OsSPL13* controls grain size in cultivated rice. *Nature Genetics* 48: 447–456.
- Song S, Tian D, Zhang Z, Hu S, Yu J. 2018. Rice genomics: over the past two decades and into the future. *Genomics, Proteomics & Bioinformatics* 16: 397–404.
- Stein JC, Yu Y, Copetti D, Zwickl DJ, Zhang L, Zhang C, Chougule K, Gao D, Iwata A, Goicoechea JL *et al.* 2018. Genomes of 13 domesticated and wild rice relatives highlight genetic conservation, turnover and innovation across the genus *Oryza*. *Nature Genetics* 50: 285–296.
- Tanksley SD, Ganai MW, Martin GB. 1995. Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. *Trends in Genetics* 11: 63–68.
- Thornberry JM, Goodman MM, Doebley J, Kresovich S, Nielsen D, Buckler ES. 2001. Dwarf8 polymorphisms associate with variation in flowering time. *Nature Genetics* 28: 286–289.
- Tsuchimoto S, Hirao Y, Ohtsubo E, Ohtsubo H. 2008. New SINE families from rice, OsSN, with poly(A) at the 3' ends. *Genes & Genetic Systems* 83: 227–236.
- Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Research* 110: 91–107.
- Weng J, Gu S, Wan X, Gao H, Guo T, Su N, Lei C, Zhang X, Cheng Z, Guo X *et al.* 2008. Isolation and initial characterization of GW5, a major QTL associated with rice grain width and weight. *Cell Research* 18: 1199–1209.
- Wicker T, Gundlach H, Spannagl M, Uauy C, Borrill P, Ramírez-González RH, De Oliveira R, Mayer KFX, Paux E, Choulet F *et al.* 2018. Impact of transposable elements on genome structure and evolution in bread wheat. *Genome Biology* 19: 103.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capi P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics* 8: nrg2165.
- Willard HF. 2013. Chapter 1 – The human genome: a window on human genetics, biology, and medicine. In: Ginsburg GS, Willard HF, eds. *Genomic and personalized medicine* (2<sup>nd</sup> edn). London, UK: Elsevier/Academic Press, 4–27.
- Xiao H, Jiang N, Schaffner E, Stockinger EJ, van der Knaap E. 2008. A retrotransposon-mediated gene duplication underlies morphological variation of tomato fruit. *Science* 319: 1527–1530.
- Zhang L, Hu J, Han X, Li J, Gao Y, Richards CM, Zhang C, Tian Y, Liu G, Gul H *et al.* 2019. A high-quality apple genome assembly reveals the association of a retrotransposon and red fruit colour. *Nature Communications* 10: 1494.
- Zhang QJ, Gao LZ. 2017. Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome *Oryza* species. *G3 (Bethesda)* 7: 1875–1885.