Theoretical and Applied Genetics (2020) 133:1291–1301
https://doi.org/10.1007/s00122-020-03548-6

ORIGINAL ARTICLE

High-throughput phenotypic screening of random genomic fragments in transgenic rice identified novel drought tolerance genes

Toshiyuki Komori[1] · Yuejin Sun[2] · Masakazu Kashihara[1] · Natsuko Uekawa[1] · Norio Kato[1] · Satoru Usami[1] · Noriko Ishikawa[1] · Yukoh Hiei[1] · Kei Kobayashi[1] · Rise Kum[1] · Esteban Bortiri[2] · Kimberly White[2] · Paul Oeller[2] · Naoki Takemori[1] · Nicholas J. Bate[2,3] · Toshihiko Komari[1]

Received: 11 December 2019 / Accepted: 14 January 2020 / Published online: 24 January 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Key message **Novel drought tolerance genes were identified by screening thousands of random genomic fragments from grass species in transgenic rice.**

Abstract Identification of agronomically important genes is a critical step for crop breeding through biotechnology. Multiple approaches have been employed to identify new gene targets, including comprehensive screening platforms for gene discovery such as the over-expression of libraries of cDNA clones. In this study, random genomic fragments from plants were introduced into rice and screened for drought tolerance in a high-throughput manner with the aim of finding novel genetic elements not exclusively limited to coding sequences. To illustrate the power of this approach, genomic libraries were constructed from four grass species, and screening a total of 50,825 transgenic rice lines for drought tolerance resulted in the identification of 12 reproducibly efficacious fragments. Of the twelve, two were from the mitochondrial genome of signal grass and ten were from the nuclear genome of buffalo grass. Subsequent sequencing and analyses revealed that the ten fragments from buffalo grass carried a similar genetic element with no significant homology to any previously characterized gene. The deduced protein sequence was rich in acidic amino acid residues in the C-terminal half, and two of the glutamic acid residues in the C-terminal half were shown to play an important role in drought tolerance. The results demonstrate that an open-ended screening approach using random genomic fragments could discover trait genes distinct from gene discovery based on known pathways or biased toward coding sequence over-expression.

Introduction

Genetically modified (GM) crops are now commercially planted in 26 countries, and acreage of GM crops reached 191 million hectares globally in 2018 and continues to expand (ISAAA 2018). The ability of agriculture to continue underpinning the rapid increase in the world population in this future world is inconceivable without the aid of biotechnology at some level. Agriculture will increasingly be expected to deliver more food with less land in a sustainable way and with an increasingly unstable climate. Most of the commercialized genetically engineered traits to date in major crops have been limited to relatively straightforward traits that provide insect resistance and herbicide tolerance. GM solutions for more complex multi-genic traits, such as those that impact grain yield or drought tolerance, have not had a significant impact or provided a clear solution to date (Nuccio et al. 2018), suggesting that novel approaches, beyond those successful with more simple traits, are required to provide durable and useful solutions.

With the advent in functional genomic technologies, the genomes of a number of crop and model plants have been completely sequenced, and their transcripts, proteins, metabolites and phenotypes have been comprehensively

Communicated by Lizhong Xiong.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s00122-020-03548-6) contains supplementary material, which is available to authorized users.

✉ Toshiyuki Komori
toshiyuki.komori@jt.com

[1] Plant Innovation Center, Japan Tobacco Inc., 700 Higashibara, Iwata, Shizuoka 438-0802, Japan

[2] Syngenta Crop Protection LLC, 9 Davis Drive, Research Triangle Park, NC 27709, USA

[3] Present Address: Pairwise Plants, 110 TW Alexander Drive, Research Triangle Park, NC 27709, USA

studied (Arabidopsis Genome Initiative 2000; International Rice Genome Sequencing Project 2005; International Wheat Genome Sequencing Consortium 2018; Schnable et al. 2009). The physiological and biochemical characterization of plants has reached an unprecedented level. With a growing knowledge base in plant biology, it has been anticipated for a long time that detailed dissection of biological processes, such as metabolic pathways, responses to environments and tissue/organ development, would lead to the opportunities to devise improved or novel means to generate superior commercial traits in transgenic plants. However, this hypothesis has not yet sufficiently been demonstrated despite the fact that numerous reports of transformation of plants with modified genes involved in the biological processes have been published. In many cases, targeted phenotypes were observed only in in vitro experiments, not reproduced in greenhouses or field. Frequently, the ectopic expression of trait genes is associated with deleterious morphological characteristics or unexpected yield reduction. In reality, agronomic trait gene discoveries often reveal more unsolved questions than answers. One potential solution has been revealed from the ongoing annotation of higher plant genomes and the identification of non-genic regulatory genetic elements, higher-order gene regulation, and noncoding RNAs that regulate gene function. The focus on coding sequences, therefore, has missed a large portion of the genome palette that could potentially be useful for crop improvement, particularly for complex traits such as drought tolerance. In addition, recent genetic characterization of the domestication process in the major crops has suggested that a significant portion of the genome has been left behind with the selection of a small set of "domestication genes" (Doebley et al. 2006). Another potentially rich source of genetic strategies for resistance to abiotic and biotic stresses may exist in wild or less domesticated plant species (Moyers et al. 2018).

Remarkable progress in the studies of site-specific endonucleases active in vivo is now about to generate new technologies to "edit" endogenous genes to develop novel crop traits. Recent tools, such as the CRISPR/Cas9 system, have made the editing process increasingly straightforward (Lee et al. 2019; Yin et al. 2017; Zhang et al. 2019). However, gene modification by transformation and by gene editing is facing the same hurdle: the lack of knowledge of target genes.

A possible approach to discover target genes for agronomic traits, such as drought tolerance, is high-throughput screening to effectively allow the plant to determine what will confer a positive effect. Many model plant species and important crops, including once recalcitrant cereals, may now be transformed with foreign genes at efficiencies that make a high-throughput GM platform possible, by particle bombardment (Liu et al. 2014) and/or a soil bacterium *Agrobacterium tumefaciens* (Hiei et al. 1994; Ishida et al. 1996, 2015) in a short time without much labor. Libraries of genes may be screened for positive effects using a reproducible phenotypic assay. Genes may enter this process without any preselection or after selection by low stringent criteria, e.g., being transcription factors, having specific protein domains, or showing certain expression profiles. Hypothesis-free screening discovery will provide an effective starting point for molecular studies of agronomic traits and plant breeding and a complement to hypothesis-based discovery with known genes and pathways.

A large-scale transformation and screening approach has previously been employed by two research groups. One group reported a high-throughput gene discovery platform "TraitMill™," in which plant genes involved in growth and developments were identified by bioinformatics tools, cDNAs were cloned into expression vectors and transferred to rice mediated by *A. tumefaciens*, and transgenic plants were evaluated by an automated robot system (Reuzeau et al. 2006). Another group transformed Arabidopsis with about 10,000 independent full-length cDNAs from Arabidopsis under the control of a constitutive promoter and monitored the transformants for various categories of phenotype changes (Ichikawa et al. 2006). The method was termed the "FOX hunting" system for full-length cDNA over-expressing gene hunting and further employed for functional analysis of rice genes in transgenic rice (Nakamura et al. 2007) and transgenic Arabidopsis (Kondou et al. 2009). Both groups produced a range of transgenic events with various phenotypic alterations.

We hereby report an approach of screening of a large number of genes employing high-throughput rice transformation. Unlike previous approaches, with over-expressed cDNAs using strong, constitutive promoters, we transformed rice with genomic fragments from four stress-tolerant grass species. Because genes in genomic fragments are likely to retain their own promoters and other genetic information lost from cDNAs, there is a good chance that the genes retain their native expression profiles and functions in transgenic rice. The transgenic rice lines were screened for drought tolerance as a model case, and subsequent analyses demonstrated that unique drought tolerance genes could be discovered by the genomic clone approach that will lead to new avenues of research and the development of novel strategies for crop improvement.

Materials and methods

Construction of genomic libraries

Total DNA or nuclear DNA was extracted from the leaves of donor plants. The DNA was partially digested with *Taq*I

and fractionated by sucrose density gradient ultracentrifugation. DNA fragments between 15 and 20 kb or between 20 and 30 kb in size were collected and ligated into the *Bst*BI site of pLC20GWH or pLC31GWH (Fig. 1a). The ligated DNA samples were packaged with MaxPlax Lambda Packaging Extracts (Epicentre) to construct genomic libraries. DNA manipulations in this study were performed according to standard procedures (Kaiser and Murrey 1985; Sambrook and Russell 2001).

Rice transformation

Recombinant plasmids were cloned into *Escherichia coli* GeneHogs® (Invitrogen) and then mobilized into *A. tumefaciens* LBA4404 (Hoekema et al. 1983) by triparental mating using *E. coli* HB101 (pRK2073) (Lemos and Crosa 1992) as a helper strain. Immature embryos of a rice cultivar Yukihikari were inoculated with the resultant *Agrobacterium*, and transgenic rice plants were created, as previously reported (Hiei and Komari 2008).

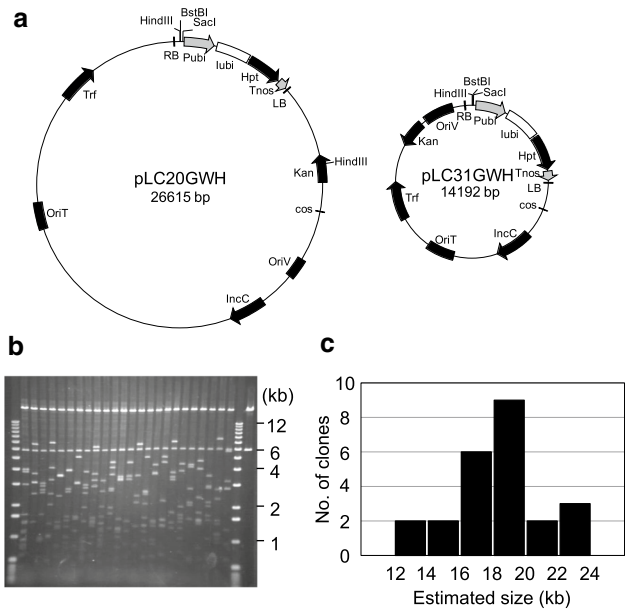

Fig. 1 Construction of genomic libraries. **a** Binary cosmid vectors used for library construction. RB, right border; Pubi, maize *Ubi1* promoter; Iubi, maize *Ubi1* intron; Hpt, hygromycin resistance gene; Tnos, 3′ signal of nopaline synthase gene; LB, left border; Kan, kanamycin resistance gene; cos, cos site of phage lambda; OriV, origin of vegetative replication of IncP plasmid; IncC, gene for IncC protein; OriT, origin of transfer of IncP plasmid; Trf, transacting replication function of IncP plasmid. **b** Restriction analysis with *Hin*dIII and *Sac*I of 24 randomly selected clones from the buffalo grass library. The lane at the right end indicates digests of pLC20GWH. **c** Insert size distribution of the 24 random clones from the buffalo grass library

Drought tolerance assay for screening

T_1 plants were selected for uniformity and transplanted to plastic pots filled with nursery soil. In each pot, six T_1 plants derived from different events were planted together with one Yukihikari plant (susceptible check) and one Suweon 287 plant (tolerant check). Water supply was stopped on the 17th day after transplanting and withheld for 3–5 days, depending on the level of moisture in each pot, which was monitored by the appearance of the check plants. Watering was resumed on a pot-by-pot basis when the whole leaf blade of the youngest fully expanded leaf of the Suweon 287 seedling discolored, following breaching of all leaves of Yukihikari. The level of recovery was individually scored according to the criteria described in Fig. 2. Scores were analyzed by the *U* test (Mann and Whitney 1947), using the T_1 plants assayed in the first cycle of the same screening batch as a control population.

Drought tolerance assay for validation and subclone test

About 100 T_0 rice plants per construct were grown in a greenhouse, and 48 similar-sized plants were transplanted to 12 pots. Each pot contained four T_0 plants to be tested, two control plants (regenerated Yukihikari or vector control T_0), one Yukihikari plant (susceptible check) and one Suweon 287 plant (tolerant check). In T_1 assay, four T_1 plants of the same event, three Yukihikari plants (two for control and one for susceptible check) and one Suweon 287 plant (tolerant check) were tested. Otherwise, the same methods as screening were used.

Analyses of transcripts in transgenic rice

RNA was extracted from leaves and roots with RNeasy Plant Mini Kit (QIAGEN) and subjected to DNase treatment with Ambion TURBO DNA-free™ Kit (Invitrogen). RT-PCR was performed after the first-strand synthesis with QuantiTect Reverse Transcription Kit (QIAGEN) using six pairs of primers (Online Resource Table S1). RACE to determine the transcription unit was carried out with GeneRacer™ Kit (Invitrogen) according to the supplier's instruction manual using gene-specific primers in Online Resource Table S2.

Results

Construction of genomic libraries

Genomic libraries were constructed from the DNA of signal grass (*Urochloa decumbens* cv. Basilisk), proso millet (*Panicum miliaceum* L.), buffalo grass (*Bouteloua*

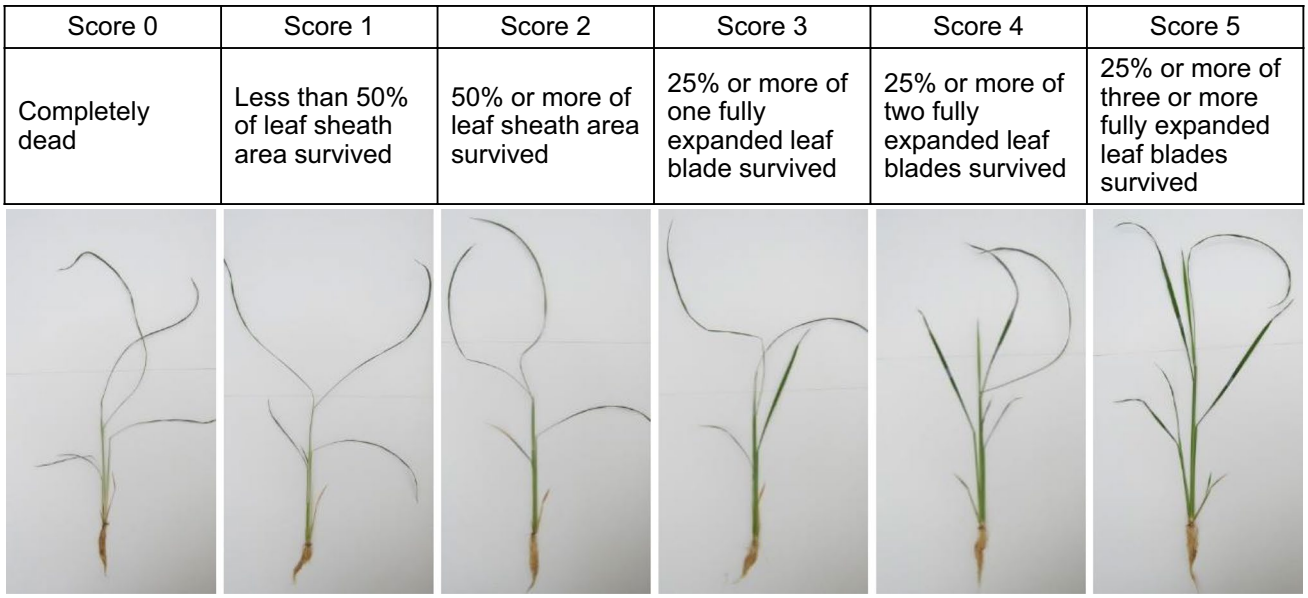

Fig. 2 Criteria of drought tolerance scores. Plants were individually scored after drought treatment followed by resuming of watering

dactyloides) and moor grass (*Sesleria heufleriana*), as summarized in Table 1. Binary cosmid vectors pLC20GWH and pLC31GWH were used to clone 15–20 and 20–30 kb fragments efficiently by the cosmid cloning system, respectively. Plasmids were isolated from randomly selected clones of each library and subjected to restriction analysis with *Hin*dIII and *Sac*I, as exemplified in Fig. 1b. The results indicated that distributions of insert size were mostly as expected; in the case of the buffalo grass library (Fig. 1c), the average inset size was estimated to be 18.3 kb.

Production of transgenic rice

The binary vectors carrying genomic fragments were individually transferred to *Agrobacterium*, and the resultant recombinant *Agrobacterium* strains were used to transform rice. With each of the *Agrobacterium* strains, one immature embryo of rice was infected, and up to two T_0 transgenic plants were grown in a greenhouse. In total, 55,696 *Agrobacterium* strains were used to infect rice, and T_1 seeds were harvested from 88,769 T_0 rice plants (Table 1).

Screening of drought-tolerant events

Transgenic rice events with 55 or more T_1 seeds were screened for drought tolerance. Out of 50,825 events assayed in the first screening cycle, in which six T_1 plants per event were used, 3782 showed drought tolerance scores (Fig. 2) higher than the corresponding control population at the 5% level (Table 2). The selected events, except for one event for which sufficient T_1 plants were not prepared due to a low germination rate, were assayed in the second cycle, in which 12 T_1 plants were tested, and 453 events were higher in the drought tolerance score at the 5% level. The 453 events were further assayed in the third cycle, in which 24 T_1 plants were tested, and 141 events passed the 5% threshold and designated as Hit Events. While the 141 Hit Events were mostly from different fragments, two from proso millet and two from buffalo grass were from the same respective fragments.

Table 1 Construction of genomic libraries and rice transformation

Donor	DNA type	Target size (kb)	Vector	No. of clones used to transform rice	No. of plants harvested	
					Total	≥55 seeds
Signal grass	Total DNA	15–20	pLC20GWH	11,599	19,814	10,407
Proso millet	Total DNA	15–20	pLC20GWH	12,049	19,562	11,107
Buffalo grass	Nuclear DNA	15–20	pLC20GWH	15,049	24,266	17,187
Moor grass	Nuclear DNA	20–30	pLC31GWH	16,999	25,127	15,699
Total				55,696	88,769	54,400

Table 2 Summary of drought tolerance screening

Donor	1st cycle			2nd cycle			3rd cycle		
	No. of events assayed	No. of events passed	Pass rate (%)	No. of events assayed	No. of events passed	Pass rate (%)	No. of events assayed	No. of events passed	Pass rate (%)
Signal grass	10,276	787	7.7	787	95	12.1	95	18	18.9
Proso millet	10,915	772	7.1	771	95	12.3	95	32	33.7
Buffalo grass	14,834	1141	7.7	1141	136	11.9	136	59	43.4
Moor grass	14,800	1082	7.3	1082	127	11.7	127	32	25.2
Total	50,825	3782	7.4	3781	453	12.0	453	141	31.1

Thus, 139 fragments corresponded to the Hit Events and designated as Hit Fragments (HF001–HF139). The selection history of HF050 and HF051, which are from buffalo grass and addressed in detail in this study, is shown in Online Resource Table S3 as an example.

Validation of selected fragments

Hit Fragments were re-introduced into rice to test the reproducibility of the effects, except for six fragments that were shown to share the same or highly homologous segments with previously validated fragments or to be derived from bacteria. Out of the 133 Hit Fragments tested, two (HF001 and HF003) from signal grass and ten (HF050, HF051, HF052, HF058, HF063, HF067, HF068, HF073, HF081 and HF099) from buffalo grass led to clear drought-tolerant responses compared to the control plants in the T_0 drought tolerance assay (Table 3). For the 12 fragments, T_1 seeds were harvested from two or three selected events from the T_0 assay, and drought tolerance was tested in the T_1 generation. As a result, at least one event of each fragment was more drought tolerant than control Yukihikari plants at the 5% level, as exemplified in Online Resource Table S4. Thus, the 12 fragments were shown to confer drought tolerance on rice in a reproducible manner.

Sequence analyses of validated fragments

Entire sequences of four of the 12 validated Hit Fragments, two of signal grass fragments (HF001 and HF003) and two (HF050 and HF051) from buffalo grass, were determined. HF001 and HF003 proved to partially overlap each other (Fig. 3), and BLAST search (Altschul et al. 1990) showed that they had high levels of homology with mitochondrion sequences of higher plants. The results strongly suggested that a DNA sequence present in the 9970-bp shared segment from a mitochondrial genome of signal grass was introduced into the nuclear genome of rice to give drought tolerance. End sequencing and *Hin*dIII/*Sac*I analysis of another Hit fragment HF016 from signal grass suggested that HF016 also contained the 9970-bp segment (Fig. 3) and supported the hypothesis above. On the other hand, HF050 and HF051 shared similar sequences that contained putative *Ty3-Gypsy* (Xiong and Eickbush 1990) retrotransposons (Fig. 4a). Unlike the validated fragments from signal grass, HF050 and HF051 were likely from the nuclear genome because (1) the buffalo grass library was constructed from a nuclear DNA preparation (Table 1), (2) HF050 and HF051 sequences did not match chloroplast or mitochondrial genomes reported to date, (3) *Ty3-Gypsy* has not been found in any organelle genomes and (4) HF050 and HF051 did not contain exactly the same segment. More detailed dissection of HF001 and HF003 and the biology underlying the phenomenon of a mitochondrial fragment that confers drought tolerance will be described elsewhere (in preparation for submission). Toward further characterization of the causative genetic element underlying HF050 and HF051, additional experiments were conducted to systematically subdivide the large fragment and characterize the underlying sequence.

Subcloning tests for HF050 and HF051

Five subclones (Sub050-1 and Sub050-2 from HF050 and Sub051-1, Sub051-2 and Sub051-3 from HF051, Fig. 4a) were created and subjected to T_0 drought tolerance assay in rice. Sub050-2, Sub051-1 and Sub051-2 gave drought tolerance to rice, whereas Sub050-1 and Sub051-3 were not effective (Online Resource Table S5). The T_0 results were confirmed by the T_1 drought tolerance assay for some events of each of the five subclones (data not shown). In order to narrow down the candidate region of the drought tolerance element in HF051, four more HF051 subclones (Sub051-4, Sub051-5, Sub051-6 and Sub051-7) were created and assayed for the capability to make rice drought tolerant in T_0 and T_1 generations. Sub051-7 was negative, whereas the other three were positive (Online Resource Table S5). Taken together, it was suggested that both HF050 and HF051 carried the same kind of drought tolerance elements in the shared region in the downstream of the *Ty3-Gypsy* retrotransposons (Fig. 4a). Seven more validated fragments (HF052,

Table 3 T_0 drought tolerance assay for 12 fragments

Line	No. of plants							*P* value	Result
	Score (0: susceptible–5: tolerant)						Total		
	0	1	2	3	4	5			
HF001	23	11	8	3	2	1	48	0.00166	Tolerant
Regenerated Yukihikari	20	4	0	0	0	0	24	–	–
HF003	20	15	8	2	3	0	48	3.14×10^{-5}	Tolerant
Regenerated Yukihikari	23	0	1	0	0	0	24	–	–
HF050	17	6	12	2	5	6	48	3.60×10^{-6}	Tolerant
Regenerated Yukihikari	23	0	1	0	0	0	24	–	–
HF051	17	3	4	6	2	16	48	3.88×10^{-6}	Tolerant
Regenerated Yukihikari	22	2	0	0	0	0	24	–	–
HF052	25	4	10	5	1	3	48	1.90×10^{-4}	Tolerant
Regenerated Yukihikari	23	1	0	0	0	0	24	–	–
HF058	23	8	8	9	0	0	48	2.07×10^{-4}	Tolerant
Regenerated Yukihikari	22	2	0	0	0	0	24	–	–
HF063	20	12	12	3	1	0	48	0.00133	Tolerant
Regenerated Yukihikari	19	4	1	0	0	0	24	–	–
HF067	29	6	9	2	2	0	48	0.00994	Tolerant
Vector control T_0	21	3	0	0	0	0	24	–	–
HF068	31	12	4	1	0	0	48	0.00395	Tolerant
Vector control T_0	23	1	0	0	0	0	24	–	–
HF073	39	5	4	0	0	0	48	0.0246	Tolerant
Vector control T_0	24	0	0	0	0	0	24	–	–
HF081	27	9	5	2	1	4	48	0.00249	Tolerant
Vector control T_0	22	1	1	0	0	0	24	–	–
HF099	25	10	11	1	1	0	48	6.28×10^{-5}	Tolerant
Vector control T_0	24	0	0	0	0	0	24	–	–

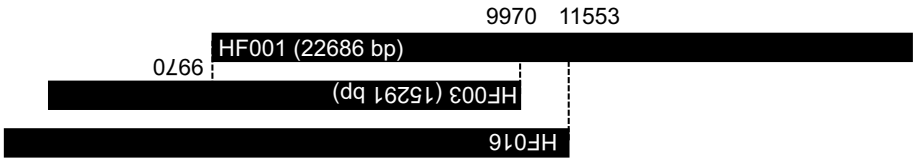

Fig. 3 Physical relationship of three Hit Fragments from signal grass, HF001, HF003 and HF016

HF058, HF063, HF067, HF068, HF073 and HF081) were fully sequenced, and each of them was shown to carry a segment containing a Sub051-5-like sequence adjacent to a *Ty3-Gypsy* retrotransposon (data not shown).

Transcriptional analysis for the Sub051-5 segment

RNA was prepared from leaves and roots of hygromycin-resistant T_1 rice plants derived from a drought-tolerant event of Sub051-5, and RT-PCR was performed with six pairs of primers illustrated in Fig. 4b. As a result, transcripts were detected for Pair 2 in both leaves and roots, but not for the other five pairs (data not shown). Then, 5′-RACE and 3′-RACE were carried out for the Pair 2 segment, and the transcription unit and direction of transcription were revealed (Fig. 4b). The results indicated that four ORFs (ORF171, ORF140, ORF120 and ORF161, Fig. 4b) could be encoded by the transcript.

Identification of the drought tolerance gene in HF051

In order to investigate the roles of the four ORFs in drought tolerance, start codons (ATGs) of ORF171, ORF140, ORF120 and ORF161 were disrupted in Sub051-5 and created four mutated versions, Sub051-5 Ma, Mb, Mc and Md, respectively. The four constructs were introduced into rice, and drought tolerance was examined in T_0 (Online Resource Table S6) and T_1 (data not shown). Sub051-5 Ma lost the capability to make rice drought tolerant, whereas

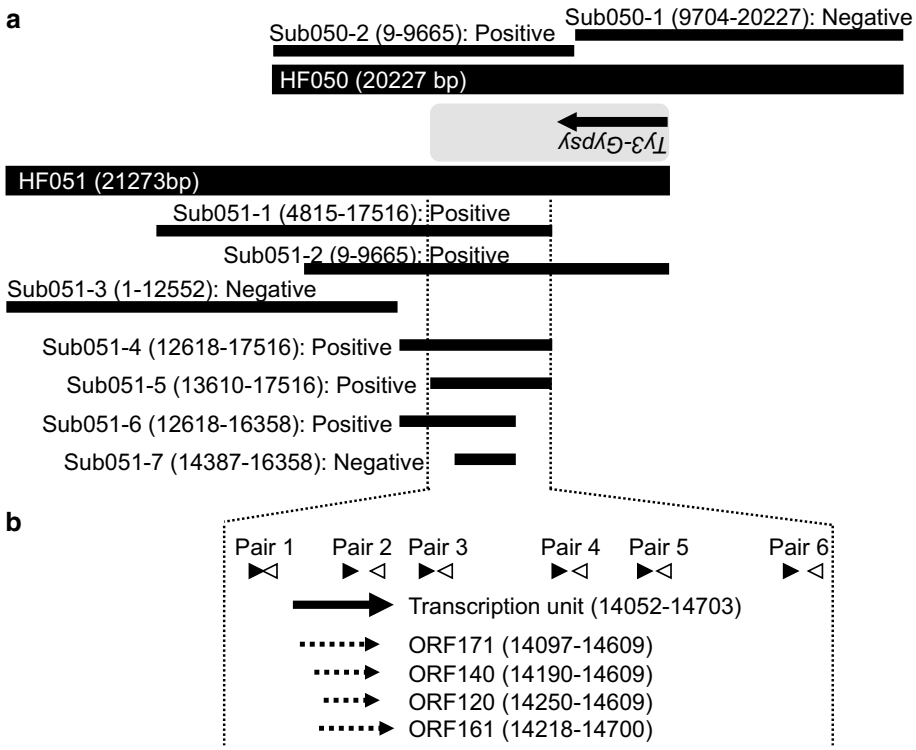

Fig. 4 Summary of analyses to identify a drought tolerance gene in HF051. **a** Subclone tests for HF050 and HF051. The two fragments share a similar segment (indicated by a gray box), which contains a putative gene for *Ty3-Gypsy* type of retrotransposon. **b** Transcriptional analyses of the Sub051-5 segment. RT-PCR was carried out with six pairs of primers (Pair 1 to Pair 6). Because transcripts were detected for only Pair 2, 5′- RACE and 3′-RACE were carried out for the Pair 2 segment. As a result, the transcription unit and direction of transcription were determined. When ORFs of more than 100 amino acids were surveyed within the transcription unit, four possible ORFs (ORF171, ORF140, ORF120 and ORF161) were found

the other three retained it, which suggested that ORF171 would be responsible for drought tolerance of HF051. The causative gene was named *BdDT* (*Bouteloua dactyloides* drought tolerance)-*ORF171*. BLAST search (Altschul et al. 1990) found no homologues of *BdDT-ORF171* in the public database.

Characterization of the BdDT-ORF171 peptide sequence

BdDT-ORF171 encodes a 171-aa protein that was rich in acidic amino acid residues (glutamic acid and aspartic acid) in the C-terminal half (Fig. 5). The other nine reproducibly effective fragments from buffalo grass (HF050, HF052, HF058, HF063, HF067, HF068, HF073, HF081 and HF099) proved to encode proteins similar to the 171-aa protein. In addition, two Hit Fragments from buffalo grass (HF086 and HF095) were ineffective in the validation test but later shown to encode BdDT-ORF171 homologues. The deduced amino acid sequences of the ten positives and the two negatives are shown in Fig. 5. Glutamic acid (E) residues at position 94 and position 108 in BdDT-ORF171 were conserved in all of the positive homologues but replaced with lysine (K) residues in both of the negative homologues, which suggested the possibility that E94K and/or E108K mutations would be critical for drought tolerance. In order to test the hypothesis, Sub051-5 Me (E94K mutant), Mf (E108K mutant) and Mg (E94K E108K double mutant) were created and assayed for drought tolerance. In T_0 assay, Sub051-5 Me and Mf gave drought tolerance to rice, whereas Sub051-5 Mg did not (Online Resource Table S6). T_1 lines of Sub051-5 Me and Mf were drought tolerant (Online Resource Table S7) and confirmed the T_0 results. T_1 assay was not conducted for Sub051-5 Mg because all of the four T_0 plants that survived in the T_0 assay were complete or almost sterile. The results that only the double mutation led to the loss of the capability to give drought tolerance to rice indicated that the double mutation is crucial and that single mutation at the two positions does not disrupt gene function.

Discussion

We have randomly screened a large number of genomic fragments from four grass species by introducing them into rice through *Agrobacterium*-mediated transformation and examining the transgenic plants for phenotypic changes in a high-throughput platform screening approach. A high level of stringency and confirmation through retransformation ensured that only fragments with a robust and reproducible drought phenotype were carried forward, while a certain number of efficacious fragments might have been lost as false-negatives, especially in the screening step. Two kinds of previously uncharacterized drought tolerance elements were found from screening over 50,000 genomic fragments, demonstrating the capability of this trait gene hunting approach to identify novel genetic elements. As drought tolerance was evaluated at a relatively early stage

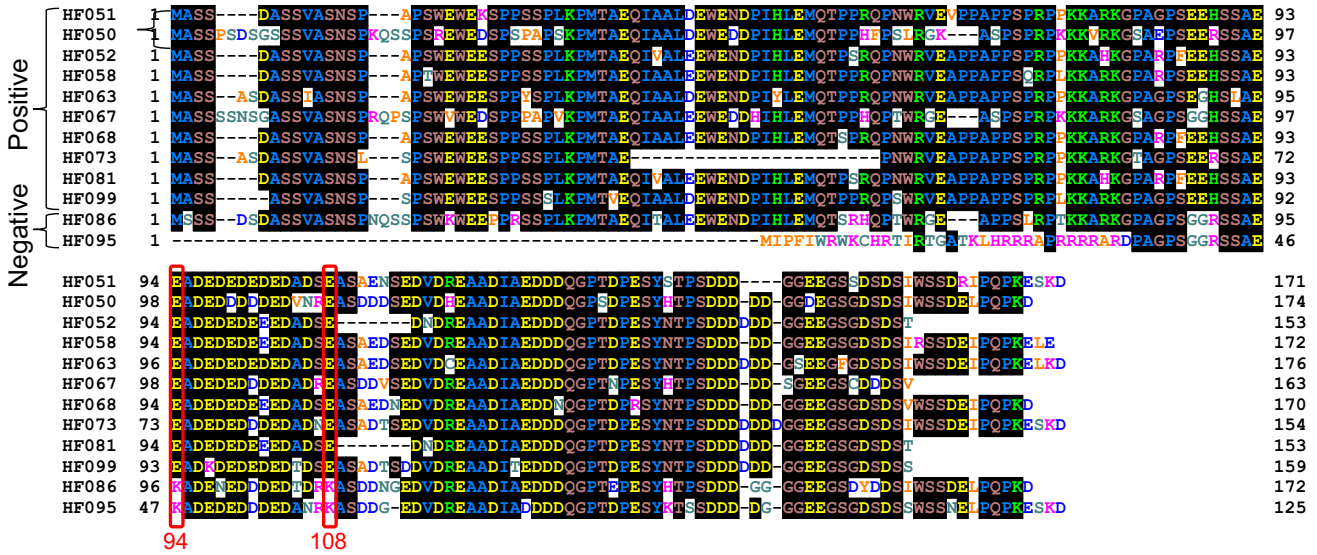

Fig. 5 Alignment of deduced amino acid sequences of BdDT-ORF171 homologues encoded by ten positive fragments and two negative fragments. Glutamic acid (E) residues at position 94 and position 108 were conserved in all of the positive homologues in HF050, HF051, HF052, HF058, HF063, HF067, HF068, HF073 HF081 and HF099 but replaced with lysine (K) residues in both of the negative homologues in HF086 and HF095

(4 weeks after sowing) in this study, it will be important to test drought tolerance of selected transgenic rice at a later stage for practical use in breeding. A similar genomic library screening approach could be applied to other complex traits, such as high yielding and other biotic/abiotic stress tolerance traits, to identify novel genetic elements of importance.

Recently, several publications have reported genes that confer drought tolerance to crop plants based on established or known hypotheses related to drought tolerance (Habben et al. 2014; Nuccio et al. 2015). A hypothesis-free approach, such as described here, serves as a complement to knowledge-based approaches and should ultimately broaden the repertoire of genes for crop trait improvement. Neither of the genetic elements identified in this study would have been identified through conventional means. Using genomic DNA from non-crop species also had the advantage to identify efficacious genetic elements that have been lost through domestication and breeding since neither element type (represented by HF001/HF003 or HF050/HF051) is present in crop species. Both signal grass and buffalo grass represent hardy prairie grasses grown for their resilience to abiotic stresses and their ability to support growth and biomass under drought conditions (Bor 1960; Qu et al. 2008). Within hypothesis-free approaches, the cDNA approach and our genomic clone approach are fully complementary. Strong, ectopic expression of cDNAs by the CaMV 35S (Benfey and Chua 1990) or maize ubiquitin (Christensen et al. 1992) promoter may cause phenotypic changes more frequently than the expression of genomic sequences by their native promoters. Thus, the cDNA approaches may find "Hit" genes more frequently. However, it may be challenging to select further genes that do not give undesired characteristics and to properly regulate the gene expression by suitable promoters and other elements. On the other hand, once genomic clones give target traits, the causative expression units, which may have been optimized through the evolution of the original plants, may be employed without much modification, although it may take some time to identify the units from "Hit" fragments. An important factor is that intron(s) and other elements lost from cDNA are retained by genomic clones. Introns could play important roles in gene function by enhancing gene expression, as exemplified by a maize gene for pyruvate, orthophosphate dikinase (PPDK) (Fukayama et al. 2001), and/or generating various transcripts from a gene through alternative splicing, as exemplified by the tobacco mosaic virus (TMV) resistance gene *N* in tobacco (Dinesh-Kumar and Baker 2000).

Another advantage of the genomic clone approach is that the construction of genomic libraries is very simple and straightforward compared to that of cDNA libraries, which needs preparation of normalized full-length cDNA clones. Genes in the genome of a donor species have equal chances to be represented in a genomic library regardless of the expression levels in the tissue used, which may enable identification of genes that are expressed at quite low levels and/or only in specific spatiotemporal manners. In addition, possible effects of gene clusters can be identified. Genes involved in the same function sometimes make a cluster (Lawrence and Roth 1996). The introduction of a genomic

fragment containing such a cluster might accelerate the identification of relevant genes.

Validated fragments from signal grass carried the same segment from the mitochondrial genome (Fig. 3). HF001 and HF003 were repeatedly confirmed to confer tolerance in the rice drought assay, but it is likely that the high titer of mitochondrial genome DNA in the preparation increased the frequency of detection for the same HF001/HF003 element. Leaf cells of Arabidopsis were reported to contain roughly 50 copies of the mitochondrial genome (Draper and Hays 2000), and it is likely that the preparation of signal grass contained a similar high concentration of organelle DNA. Despite a potentially high titer, the repeated confirmation of HF001/HF003 fragments in a number of retransformation experiments indicates that this fragment is not an artifact but rather plays some role in stress tolerance. Mitochondrial genes have not often been studied in the context of plant drought responses. A rare example was the involvement of a rice cytoplasmic male sterile (CMS) gene *orfH79* in drought tolerance on the basis of the observation that the introduction of the corresponding fertility restorer gene *Rf5* into a CMS line resulted in improved drought tolerance (Yu et al. 2015). As far as we know, this is the first report that observed improved drought tolerance by introducing a mitochondrial sequence into the nuclear genome of a higher plant. Additional studies are being conducted to understand how this sequence was expressed and gave drought tolerance in rice and expected to generate important information related to drought responses.

Ten validated fragments from buffalo grass were shown to share similar sequences containing the drought tolerance genes, *BdDT-ORF171* or its homologue. Thus, the *BdDT-ORF171* gene family highly likely exists in high copy number in the buffalo grass genome although homologues of *BdDT-ORF171* have never been reported in other species. It should be noted that the *BdDT-ORF171* gene and its homologues were located immediately adjacent to the core sequences of *Ty3-Gypsy* retrotransposons (Fig. 4). Perhaps, an original gene of *BdDT-ORF171* was introduced into an ancestor of buffalo grass together with a *Ty3-Gypsy* retrotransposon in some way, e.g., through horizontal gene transfer (El Baidouri et al. 2014), and then spread in the genome in association with the propagation of the retrotransposon in the process of evolution.

BdDT-ORF171 was shown to encode a 171-aa protein, rich in acidic amino acid residues (glutamic acid and aspartic acid) in the C-terminal half (Fig. 5). When deduced amino acid sequences were compared among *BdDT-ORF171* and its homologues with or without drought tolerance capability, glutamic acid residues at position 94 and position 108 were conserved specifically in drought tolerance homologues (Fig. 5). The importance of the glutamic acid residues at the two positions was confirmed from the observation that E94K E108K double mutant of *BdDT-ORF171* resulted in the loss of the drought tolerance function (Online Resource Table S6). *BdDT-ORF171* is a novel drought tolerance gene, and the underpinning mechanisms remain to be elucidated. However, it is likely that the acidic amino acid region plays an important role in drought tolerance. The acidic stretch of the BdDT-ORF171 protein is reminiscent of activation domains in some transcription factors (Liu et al. 1999). For example, C1, which is a transcription activator of genes involved in anthocyanin pigmentation in maize, has an acidic activation domain at the C-terminus, and one aspartic acid at position 262 in the domain was shown to be essential by mutagenesis analysis (Sainz et al. 1997). BdDT-ORF171 might be involved in the regulation of gene expression as proposed for proteins with structural aspartic acid/glutamic acid-rich repeats (Chou and Wang 2015). Transposable element activity increases following abiotic stress (Miousse et al. 2015), and one speculative mechanism is that *BdDT-ORF171* is similarly amplified by association with *Ty3-Gypsy* retrotransposon mobilization to serve as a means to propagate a drought response.

Frequencies of validated fragments were different among the four DNA donor species used. Although similar scale of screening was conducted for the four species (Table 1), multiple validated fragments were identified from signal grass and buffalo grass, whereas no validated fragment was found from proso millet and moor grass. On the other hand, multiple validated fragments from each of signal grass and buffalo grass were shown to have the same kind of drought tolerance gene. Therefore, drought tolerance genes might be more efficiently discovered by increasing the number of donor species while reducing the number of fragments to be screened per donor species.

The use of a hypothesis-free discovery approach identified two classes of genetic elements that have not previously been characterized. These results clearly indicate that there is much still to learn about molecular mechanisms of the stress response and that limiting our focus to known genes or simply to coding sequences may be missing important genetic strategies employed by non-crop plants that could be crucial for developing drought tolerance solutions for crop improvement. Finally, screening of genomic libraries from four species for one trait identified two kinds of interesting genes. Increasing the number of donor species, scaling up the screening and adapting the assay to additional trait targets will continue to identify novel genes, pathways and crop solutions.

Acknowledgements The authors thank the members of our research institutes for valuable discussions and technical assistance.

Accession numbers pLC20GWH, LC506529; pLC31GWH, LC506530; HF001, LC507188; HF003, LC507189; HF050, LC507217; HF051, LC507218.

Author contribution statement TKomo, YS, NK, NT, NB and TKoma designed the research and experiments. TKomo, YS, MK, NU, SU, NI, YH, KK, RK, EB and NB performed the experiments. TKomo, YS, KW, PO, NT, NB and TKoma regularly discussed the research progress and developed the research strategy. TKomo, YS, NB and TKoma wrote the manuscript.

Compliance with ethical standards

Conflict of interest All authors are or were affiliated with Japan Tobacco Inc. or Syngenta Crop Protection LLC, who were also funders of this study.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Benfey PN, Chua NH (1990) The cauliflower mosaic virus 35S promoter: combinatorial regulation of transcription in plants. Science 250:959–966

Bor NL (1960) The grasses of Burma, Ceylon, India and Pakistan (excluding Bambuseae). Pergamon Press, London

Chou CC, Wang AH (2015) Structural D/E-rich repeats play multiple roles especially in gene regulation through DNA/RNA mimicry. Mol BioSyst 11:2144–2151

Christensen AH, Sharrock RA, Quail PH (1992) Maize polyubiquitin genes: structure, thermal perturbation of expression and transcript splicing, and promoter activity following transfer to protoplasts by electroporation. Plant Mol Biol 18:675–689

Dinesh-Kumar SP, Baker BJ (2000) Alternatively spliced *N* resistance gene transcripts: their possible role in tobacco mosaic virus resistance. Proc Natl Acad Sci USA 97:1908–1913

Doebley JF, Gaut BS, Smith BD (2006) The molecular genetics of crop domestication. Cell 127:1309–1321

Draper CK, Hays JB (2000) Replication of chloroplast, mitochondrial and nuclear DNA during growth of unirradiated and UVB-irradiated *Arabidopsis* leaves. Plant J 23:255–265

El Baidouri M, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O (2014) Widespread and frequent horizontal transfers of transposable elements in plants. Genome Res 24:831–838

Fukayama H, Tsuchida H, Agarie S, Nomura M, Onodera H, Ono K, Lee BH, Hirose S, Toki S, Ku MS, Makino A, Matsuoka M, Miyao M (2001) Significant accumulation of C_4-specific pyruvate, orthophosphate dikinase in a C_3 plant, rice. Plant Physiol 127:1136–1146

Habben JE, Bao X, Bate NJ, DeBruin JL, Dolan D, Hasegawa D, Helentjaris TG, Lafitte RH, Lovan N, Mo H, Reimann K, Schussler JR (2014) Transgenic alteration of ethylene biosynthesis increases grain yield in maize under field drought-stress conditions. Plant Biotechnol J 12:685–693

Hiei Y, Komari T (2008) *Agrobacterium*-mediated transformation of rice using immature embryos or calli induced from mature seed. Nat Protoc 3:824–834

Hiei Y, Ohta S, Komari T, Kumashiro T (1994) Efficient transformation of rice (*Oryza sativa* L.) mediated by *Agrobacterium* and sequence analysis of the boundaries of the T-DNA. Plant J 6:271–282

Hoekema A, Hirsch PR, Hooykaas PJJ, Schilperoort RA (1983) A binary plant vector strategy based on separation of *vir*- and T-region of the *Agrobacterium tumefaciens* Ti-plasmid. Nature 303:179–180

Ichikawa T, Nakazawa M, Kawashima M et al (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. Plant J 48:974–985

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

International Wheat Genome Sequencing Consortium (2018) Shifting the limits in wheat research and breeding using a fully annotated reference genome. Science 361:eaar7191

ISAAA (2018) Global status of commercialized biotech/GM crops in 2018: biotech crops continue to help meet the challenges of increased population and climate change. ISAAA Brief No. 54. ISAAA, Ithaca, NY

Ishida Y, Saito H, Ohta S, Hiei Y, Komari T, Kumashiro T (1996) High efficiency transformation of maize (*Zea mays* L.) mediated by *Agrobacterium tumefaciens*. Nat Biotechnol 14:745–750

Ishida Y, Tsunashima M, Hiei Y, Komari T (2015) Wheat (*Triticum aestivum* L.) transformation using immature embryos. Methods Mol Biol 1223:189–198

Kaiser K, Murrey N (1985) Construction of representative genomic DNA libraries. In: Glover DM (ed) DNA cloning, vol 1. IRL Press Ltd., Oxford, pp 1–47

Kondou Y, Higuchi M, Takahashi S et al (2009) Systematic approaches to using the FOX hunting system to identify useful rice genes. Plant J 57:883–894

Lawrence JG, Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. Genetics 143:1843–1860

Lee K, Zhu H, Yang B, Wang K (2019) An *Agrobacterium*-mediated CRISPR/Cas9 platform for genome editing in maize. Methods Mol Biol 1917:121–143

Lemos ML, Crosa JH (1992) Highly preferred site of insertion of Tn7 into the chromosome of *Vibrio anguillarum*. Plasmid 27:161–163

Liu L, White MJ, MacRae TH (1999) Transcription factors and their genes in higher plants functional domains, evolution and regulation. Eur J Biochem 262:247–257

Liu G, Campbell BC, Godwin ID (2014) Sorghum genetic transformation by particle bombardment. In: Henry R, Furtado A (eds) Cereal genomics: methods in molecular biology (methods and protocols). Humana Press, Totowa, pp 219–234

Mann HB, Whitney DR (1947) On a test of whether one of two random variables is stochastically larger than the other. Ann Math Stat 18:50–60

Miousse IR, Chalbot MC, Lumen A, Ferguson A, Kavouras IG, Koturbash I (2015) Response of transposable elements to environmental stressors. Mutat Res Rev Mutat Res 765:19–39

Moyers BT, Morrell PL, McKay JK (2018) Genetic costs of domestication and improvement. J Hered 109:103–116

Nakamura H, Hakata M, Amano K et al (2007) A genome-wide gain-of function analysis of rice genes using the FOX-hunting system. Plant Mol Biol 65:357–371

Nuccio ML, Wu J, Mowers R, Zhou HP, Meghji M, Primavesi LF, Paul MJ, Chen X, Gao Y, Haque E, Basu SS, Lagrimini LM (2015) Expression of trehalose-6-phosphate phosphatase in maize ears improves yield in well-watered and drought conditions. Nat Biotechnol 33:862–869

Nuccio ML, Paul M, Bate NJ, Cohn J, Cutler SR (2018) Where are the drought tolerant crops? An assessment of more than two decades of plant biotechnology effort in crop improvement. Plant Sci 273:110–119

Qu R, Luo H, Meier VD (2008) Turfgrass. In: Kole C, Hall TC (eds) Compendium of transgenic crop plants: transgenic plantation crops, ornamentals and trufgrasses. Blackwell Publishing, Oxford, pp 177–218

Reuzeau C, Frankard V, Hatzfeld Y, Sanz A, Van Camp W, Lejeune P, De Wilde C, Lievens K, de Wolf J, Vranken E, Peerbolte R, Broekaert W (2006) Traitmill™: a functional genomics platform for the phenotypic analysis of cereals. Plant Genet Resour 4:20–24

Sainz MB, Goff SA, Chandler VL (1997) Extensive mutagenesis of a transcriptional activation domain identifies single hydrophobic and acidic amino acids important for activation in vivo. Mol Cell Biol 17:115–122

Sambrook J, Russell DW (2001) Molecular cloning: a laboratory manual, 3rd edn. Cold Spring Harbor Laboratory Press, New York

Schnable PS, Ware D, Fulton RS et al (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112–1115

Xiong Y, Eickbush TH (1990) Origin and evolution of retroelements based upon their reverse transcriptase sequences. EMBO J 9:3353–3362

Yin K, Gao C, Qiu JL (2017) Progress and prospects in plant genome editing. Nat Plants 3:17107

Yu C, Wang L, Xu S, Zeng Y, He C, Chen C, Huang W, Zhu Y, Hu J (2015) Mitochondrial ORFH79 is essential for drought and salt tolerance in rice. Plant Cell Physiol 56:2248–2258

Zhang Y, Malzahn AA, Sretenovic S, Qi Y (2019) The emerging and uncultivated potential of CRISPR technology in plant science. Nat Plants 5:778–794

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.