Rice RS2-9, which is bound by transcription factor OSH1, blocks enhancer–promoter interactions in plants

Huawei Liu1,2,†, Li Jiang1,2,†, Zhifeng Wen1,2,†, Yingjun Yang1,4,†, Stacy D. Singer5, Dennis Bennett1, Wenyong Xu6, Zhen Su6, Zhihong Yu2, Jonathan Cohn7, Hyunsook Chae7, Qiudeng Que7, Yue Liu8, Chang Liu9, and Zongrang Liu1,†

1USDA-ARS, Appalachian Fruit Research Station, Kearneysville, West Virginia 25430, USA, 2College of Food Science and Technology, Nanjing Agricultural University, Nanjing 210095, China, 3College of Horticulture, Fujian Agriculture and Forestry University, Fuzhou 350002, China, 4Forestry College, Henan University of Science and Technology, Luoyang 471023, China, 5Agriculture and Agri-Food Canada, Lethbridge Research and Development Centre, Lethbridge, Alberta T1J 4B1, Canada, 6State Key Laboratory of Plant Physiology and Biochemistry, College of Biological Sciences, China Agricultural University, Beijing 100193, China, 7Syngenta Crop Protection, LLC, Research Triangle Park, North Carolina 27709, USA, 8College of Horticulture, Qingdao Agricultural University, Qingdao 266109, China, and 9Department of Epigenetics, University of Hohenheim, Stuttgart 70599, Germany

Received 20 August 2021; accepted 2 November 2021; published online 12 November 2021.

*For correspondence (e-mail: zongrang.liu@usda.gov).
†These authors contributed equally to this work.

SUMMARY

Insulators characterized in Drosophila and mammals have been shown to play a key role in the restriction of promiscuous enhancer–promoter interactions, as well as reshaping the topological landscape of chromosomes. Yet the role of insulators in plants remains poorly understood, in large part because of a lack of well-characterized insulators and binding factor(s). In this study, we isolated a 1.2-kb RS2-9 insulator from the Oryza sativa (rice) genome that can, when interposed between an enhancer and promoter, efficiently block the activation function of both constitutive and floral organ-specific enhancers in transgenic Arabidopsis and Nicotiana tabacum (tobacco). In the rice genome, the genes flanking RS2-9 exhibit an absence of mutual transcriptional interactions, as well as a lack of histone modification spread. We further determined that O. sativa Homeobox 1 (OSH1) bound two regions of RS2-9, as well as over 50 000 additional sites in the rice genome, the majority of which resided in intergenic regions. Mutation of one of the two OSH1-binding sites in RS2-9 impaired insulation activity by up to 60%, whereas the mutation of both binding sites virtually abolished insulator function. We also demonstrated that OSH1 binding sites were associated with 72% of the boundaries of topologically associated domains (TADs) identified in the rice genome, which is comparable to the 77% of TAD boundaries bound by the insulator CCCTC-binding factor (CTCF) in mammals. Taken together, our findings indicate that OSH1-RS2-9 acts as a true insulator in plants, and highlight a potential role for OSH1 in gene insulation and topological organization in plant genomes.

Keywords: enhancer–promoter interaction, insulator and binding factor, topologically associated domains (TADs), high-throughput screening, transgenic plants

INTRODUCTION

It is well known that cis-regulatory elements (CREs) in promoter regions control gene expression activity and tissue specificity through their association with transcription factors or protein complexes. Many CREs are also physically separated from promoters, and function as independent regulatory units or enhancers that are situated upstream, downstream or within coding regions (Busch et al., 1999; Ito et al., 2003; Raatz et al., 2011), and even distally (Castelletti et al., 2014; Stam et al., 2002; Zheng et al., 2015). As distal enhancers act in a gene-specific manner rather than promiscuously, mechanisms must exist to insulate or prevent unrestricted interactions.

Plant enhancers in a transgenic state, on the other hand, often become highly interactive, leading to the ectopic activation of neighboring promoters. For example, this has
been demonstrated for several flower-specific enhancers, whereby their activation activities often mirror their native tissue-specific activity (Liu et al., 2008; Wen et al., 2014). Similarly, tissue-specific promoters in a transgenic context have also been found to ectopically interact with, and activate, adjacent promoters (Wen et al., 2014). Hence, both enhancers and promoters share a similar interactive and activation potential. However, not all enhancers are capable of superseding adjacent promoters, and many of these are instead themselves affected by adjacent CREs (Gudynas-Savitch et al., 2010; Liu et al., 2008; Singer et al., 2010a). One prominent example is the flower-specific Arabidopsis AGAMOUS enhancer (AGe), which can be ectopically activated by an adjacent enhancer in vegetative tissues when in a transgenic context (Hily et al., 2009; Singer et al., 2010a). Yet this type of cross activation rarely occurs within plant genomes, which again suggests the existence of an insulation apparatus that prevents promiscuous interactions between CREs, both locally and at a distance. Despite evidence for such a system in plants, very little is yet known about such regulatory processes.

In *Drosophila melanogaster* and in mammals, enhancer–promoter interactions are known to be inhibited by enhancer-blocking (EB) insulators, which have been explicitly demonstrated using transgenic assays in which an EB insulator is interposed between an enhancer and a promoter (Bell et al., 1999; Chung et al., 1993; Kellum and Schedl, 1992). EB insulators are believed to act through bound protein complexes, which act as a physical barrier to block the passage of an activating signal from an enhancer to a promoter (Wallace and Felsenfeld, 2007). Insulators are defined by several criteria, including the fact that they act in a position-dependent manner, impede enhancer-modulated transcription only when an insulator is interposed between the enhancer and the promoter, and do not silence either enhancer or promoter activity (Cai and Levine, 1995; Geyer and Corces, 1992; Kellum and Schedl, 1992; Scott and Geyer, 1995; Scott et al., 1999). In addition to their EB activity, insulators and their bound proteins also act as physical chromatin boundary elements to restrict the spreading of epigenetic marks or a repressive heterochromatin state (Cuddapah et al., 2009; Sigrist and Pirrotta, 1997).

To date, at least five different types of EB insulator have been characterized in *Drosophila*, including *scs, scs’, gypsy, fab-7/8 and SF1*, each of which is bound by distinct insulator-bound proteins, including Boundary Element-Associated Factor of 32 kD (BEAF-32), Suppressor of Hairy-wing (Su(Hw)), GAGA Factor (GAF), Zest White 5 (Zw5), CP190 and the *Drosophila* homolog (dCTCF) of vertebrate CCCTC-binding factor (CTCF) (Ramirez et al., 2018; Valenzuela and Kamakaka, 2006; Van Bortle et al., 2014; Wang et al., 2018). These insulator-bound proteins have been shown to collectively occupy at least 14,000 sites in the *Drosophila* genome (Negre et al., 2010). Although several DNA fragments have also been identified as EB insulators in mammals (Chung et al., 1993; Hark et al., 2000), every one of them requires the binding of the CTCF protein in order to elicit their function (Bell and Felsenfeld, 2000; Bell et al., 1999), which indicates that CTCF-bound insulators are the only insulators that have evolved in mammals. Therefore, it is apparent that insulator-mediated regulation in *Drosophila* is far more diverse than that found in mammals.

The CTCF protein exhibits genome-wide binding activity (Chen et al., 2008; Jothi et al., 2008; Kim et al., 2007; Maurano et al., 2015) and preferentially binds to a 40-bp conserved consensus sequence motif (Klenova et al., 1993; Nakashashi et al., 2013; Rhee and Pugh, 2011). In-depth analyses have revealed that CTCF binding facilitates intrachromosomal contacts between CTCF-bound sites (Yusufzai et al., 2004), and such contact leads to the extrusion of a loop domain that physically separates enhancers and promoters (insulation) or brings them into proximity (facilitation) (Hou et al., 2008), or alternatively blocks the spreading of histone modifications. This loop-based mechanism unifies the insulation and chromatin domain barrier functions of CTCF binding.

Recent analyses have revealed that CTCF binding also mediates the demarcation of chromosomes into distinct topologically associated domains (TADs) with a median size of 880 kb, which often also form sub-loop domains and cover more than 90% of the mouse genome (Dixon et al., 2012; Phillips-Cremins and Corces, 2013; Rao et al., 2014). Both TADs and sub-TADs are frequently anchored at a pair of convergent CTCF binding sites associated with cohesin and other protein factors (Dixon et al., 2012; Phillips-Cremins and Corces, 2013). The deletion, or change in orientation, of CTCF binding sites in TAD boundaries results in the disruption of TADs and/or the formation of a new TAD at an alternative binding site that is located either nearby or at a distance (Flavahan et al., 2016; Narendra et al., 2015; Nora et al., 2012), which collectively highlights the crucial role of CTCF binding sites in defining the boundaries of TADs. A large body of evidence indicates that TAD compartmentation in mammalian genomes is critical for the maintenance of proper enhancer–promoter interactions (Flavahan et al., 2016; Kang et al., 2015; Lupiáñez et al., 2015; Nora et al., 2012), euchromatin and heterochromatin states in specific regions (Narendra et al., 2015; Xu et al., 2014) and transcriptional landscapes (Akdemir et al., 2020), as well as developmental and pathogenesis processes (Flavahan et al., 2016; Kang et al., 2015; Lupiáñez et al., 2015; Nora et al., 2012). In contrast, not all insulators/insulator-bound proteins characterized in *Drosophila* are involved in the demarcation of TADs. Indeed, only the *scs*-binding BEAF-32, together with Chromator and CP190, is significantly co-enriched in the boundaries of TADs (Wang et al., 2018).
This indicates that most insulator-bound proteins in *Drosophila* are exclusively involved in enhancer blocking, but not in the formation of TADs.

Although efforts to identify EB insulator systems in plants have been underway for decades, our current understanding of this field remains limited and has been somewhat complicated by the fact that the putative insulators that have been assessed to date have had diverse origins (e.g. mammalian, bacterial, plant), inconsistent action, and a lack of conserved sequences (van der Geest and Hall, 2003). However, this assay vector is based on a cut-and-paste cloning strategy and is therefore not analogous insulators (*UASrpg* from *Arabidopsis* and *BEAD1* from *Homo sapiens*) are able to block enhancer-promoter interactions in plants. However, somewhat surprisingly, their insulation function was found to require the sequence motifs recognized and bound by their native protein factors, even though no homologous proteins exist in plants (Tran and Johnson, 2020). Furthermore, although several groups have reported prominent TAD/chromatin loop conformation in plant genomes (Dong et al., 2017; Liu et al., 2017; Wang et al., 2015), it remains unknown whether this depends on insulator protein(s), largely because of a lack of well-characterized insulators in plants. In an attempt to further our understanding of EB insulation and chromatin organization in plants, we developed novel approaches and carried out extensive analyses in transgenic plants, leading to the isolation and characterization of an enhancer-promoter blocking insulator from *Oryza sativa* (rice). In addition, we distinguish its requisite binding protein and demonstrate its association with TAD boundaries in the rice genome.

**RESULTS**

**High-throughput screening of EB insulators from the rice genome**

In a previous study, we exploited the strong activation potential of the tandemly duplicated cauliflower mosaic virus (CaMV)35S enhancer (*d35Se*), which exhibits greater activity than a single enhancer (Kay et al., 1987), on the flower-specific promoter derived from the AGAMOUS second intron (*AGIP*) (Liu et al., 2008) as a means of identifying EB insulators that function in plants (Hily et al., 2009; Singer et al., 2010a). However, this assay vector is based on a cut-and-paste cloning strategy and is therefore not practical for use in the high-throughput screening of insulators in transgenic plants. To circumvent this problem, we introduced the integrase recognition sites of the Gateway cloning system to convert the conventional assay vector, *JM69*, into a high-throughput vehicle, *pSYG3A* (Figure 1a). Given that most *Drosophila* and mammalian insulators are located between enhancers and promoters, or in intergenic regions between neighboring genes (Kim et al., 2007; Valenzuela and Kamakaka, 2006), we assumed that insulators in plant genomes are likely to share a similar localization. Hence, we targeted intergenic regions in the rice genome and focused on candidate fragments that were flanked by genes displaying distinct expression patterns or strengths, based on an analysis of publicly available transcriptome data (Jain et al., 2007; Ma et al., 2005; Sato et al., 2013). We isolated 59 intergenic regions between adjacent genes that displayed distinct tissue-specific or differential levels of expression, with sizes ranging from 1 to 2 kb (Table S1 and S2). We then cloned these regions into the high-throughput *pSYG3A* vector (Figure 1a) and introduced the resulting vectors into *Arabidopsis*. Any fragment that was able to attenuate or block the *d35Se*-mediated activation of flower-specific GUS expression driven by the *AGIP* promoter (*AGIP::GUS*) in vegetative tissues was considered to provide an insulation function.

As expected, no GUS staining was noted in wild-type seedlings, and only very low levels of staining were observed in T1 transgenic seedlings bearing the *AGIP::GUS* cassette (Figure 1b,c). Conversely, transgenic seedlings bearing *d35Se* placed immediately upstream of the *AGIP::GUS* fusion (*d35Se//AGIP::GUS*) exhibited GUS staining in up to 85% of seedlings (Figure 1b,c), confirming our previous finding that the CaMV35S enhancer (*35Se*) is capable of ubiquitously activating floral-specific *AGIP* activity in vegetative tissues (Hily et al., 2009; Singer et al., 2010a). As expected, two control constructs, *JM85* (harboring a 4-kb phage fragment) and *CW148* (harboring a 2-kb genomic fragment cloned from an intergenic region in the Arabidopsis genome) (Figure 1a; Table S1), showed no insulation function (Figure 1a; Hily et al., 2009) and exhibited high levels of GUS staining in over 83% of seedlings (Figure 1b,c), which is similar to the rate seen in *d35Se//AGIP::GUS* seedlings. However, two of the 59 rice fragments tested (*RS2-9* and *RS2-22*) displayed a substantial reduction in the frequency of GUS-positive seedlings, with only 11 and 13% of seedlings exhibiting GUS staining, respectively (Figure 1b,c). To rule out false-positive outcomes, we re-cloned the same *RS2-9* and *RS2-22* fragments directly from the transgenic T1 seedlings, confirmed them by sequencing and re-inserted them into the *pSYG3A* vector before re-introduction into Arabidopsis. Histochemical staining yielded a similar reduction in GUS staining frequencies in the re-cloned *RS2-9* and *RS2-22* T1 seedlings (data not shown).

To determine whether the same insulation function was evident in mature plants, we transferred over 80 transgenic seedlings bearing the *RS2-9*, *RS2-22*, *CW148*, *AGIP::GUS* and *d35Se//AGIP::GUS* cassettes to soil, respectively. X-Gluc staining was analyzed in the leaves and floral buds of 40-day-old plants in each case. As expected, no GUS staining was observed in any tissue of

© 2021 Her Majesty the Queen in Right of Canada. The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd. Reproduced with the permission of the Minister of Agriculture. This article has been contributed to by US Government employees and their work is in the public domain in the USA.
Figure 1. High-throughput screening and characterization of enhancer-blocking (EB) insulators in transformed Arabidopsis thaliana. (a) Assay vector and cloning strategy. Isolated insulator candidates from the Oryza sativa (rice) genome were inserted into pSYG3A via λ integrase-mediated site-specific recombination between attL (L1 and L2) and attR (R1 and R2) sites to replace the RIC-1 fragment. P1 and P2 denote the attR1 and attR2 sites created as a result of recombination between the attL1 and attR1 and the attL2 and attR2 sites, respectively. Arrows indicate the direction of transcription. Abbreviations: LB and RB, T-DNA left and right borders; d35Se, tandem duplication of the 250-bp cauliflower mosaic virus (CaMV) 35S enhancer (35Se [Kay et al., 1987]; CmR, Escherichia coli chloramphenicol resistance gene; ccdB, a gene encoding CcdB that acts as a DNA gyrase toxin in bacteria; AGIP, Arabidopsis AGAMOUS enhancer-derived promoter (Liu et al., 2008); GUS, β-glucuronidase gene; RIC,1, CmR and ccdB gene fragments flanked by attR1/attR2 sites; T, transcriptional terminator. (b) Frequency of GUS-positive T1 seedlings. Data are averaged from three biological replicates, ±SD. More than 80 kanamycin-resistant T1 seedlings were selected and stained in each transformation experiment. (c) X-Gluc-stained phenotype of T1 Arabidopsis seedlings bearing the indicated constructs. (d)–(i) X-Gluc staining of leaf and floral tissues from wild-type (Wt) (d), 35Se/AGIP::GUS (e), AGIP::GUS (f), CW148 (g), RS2-22 (h) and RS2-9 (i) plants. (j) Frequency of GUS-positive leaf and floral tissues in mature transgenic plants (n > 80). (k) Real-time quantitative RT-PCR analysis of GUS transcripts relative to the UBQ10 mRNA level. Data are averaged from three biological replicates, ±SD. Mature leaves were pooled from five lines for each replicate.

untransformed wild-type plants (Figure 1d), and GUS staining was not apparent in the leaves of AGIP::GUS plants but was evident in the flowers (Figure 1f). Conversely, the d35Se/AGIP::GUS, CW148 and d35Se/AGIP::GUS plants exhibited strong GUS staining in both leaves and flowers (Figure 1e,g). The presence of RS2-9 or RS2-22 between the d35Se and AGIP::GUS cassette reduced the incidence of GUS staining in the leaves, but not in the flowers (Figure 1h,i), resulting in a staining pattern resembling that observed in AGIP::GUS plants (Figure 1f). This indicates that the presence of either the RS2-9 or RS2-22 fragment in the transgenic cassette did not repress the transcriptional capacity or flower-specific activity of the AGIP promoter. Although a similar percentage of plants bearing each vector exhibited GUS staining in flowers, only 15 and 17% of the RS2-9 and RS2-22 lines exhibited weak but detectable GUS staining in leaves, respectively, which was higher than the 5% of AGIP::GUS plants, but much lower than the proportion of d35Se/AGIP::GUS and CW148 lines that exhibited GUS staining in vegetative tissues (Figure 1j).

To validate the insulation function of the two fragments at the transcriptional level, GUS expression was assessed in the leaves of selected transgenic lines using quantitative real-time reverse transcription polymerase chain reaction (RT-PCR). As expected, significant reductions in normalized GUS transcript levels were evident in the leaves of RS2-9 and RS2-22 lines compared with d35Se/AGIP::GUS and CW148 lines (Figure 1k). We also performed PCR amplification using a primer pair located in vector regions flanking the inserted fragments, and specifically detected the 1.2-kb RS2-9, 1.0-kb RS2-22 and 2.0-kb CW148 fragments in all selected lines, respectively (Figure S2). Taken together, our results indicate that the presence of either the RS2-9 or RS2-22 fragment between the d35Se and AGIP::GUS cassette within a transgene led to the insulation of the AGIP promoter from activation by d35Se in Arabidopsis leaves.

RS2-9, but not RS2-22, impeded the activation function of a flower-specific enhancer in Nicotiana tabacum (tobacco)

To test whether the RS2-9 and RS2-22 fragments could also block the activation function of a plant-derived,
flower-specific enhancer in tobacco, we introduced the fragments into a phenotype-based assay vector \((pAGI1RACT)\) that was developed previously for the analysis of enhancer-promoter interactions and insulator function in tobacco (Wen et al., 2014) (Figure 2a). The \(pAGI1RACT\), based assay exploits the robust activation function of a petal-, stamen- and carpel-specific tobacco AGAMOUS1 enhancer \((NtAG1e)\) on a pollen- and stigma-specific promoter \((Pps)\) in flowers (Wen et al., 2014). In the \(pBASEACT\) vector, \(Pps\)-driven DT-A (a cytoxic gene) was exclusively expressed in pollen and stigmatic tissues, resulting in a phenotype that will hereafter be dubbed a ‘sterile flower phenotype’ (Figure 2c-f; Wen et al., 2014). However, the insertion of the \(NtAG1e\) 2.5 kb downstream of the \(Pps\) promoter::DT-A cassette in the \(pAGI1RACT\) vector led to the activation of DT-A expression in petals, stamens and carpels, resulting in the conversion of the sterile flower phenotype into an ‘ablated flower phenotype’ in approximately 50% of the transgenic lines (Figure 2b). The resulting ablated flower phenotype resembled that seen in \(rNtAGIP1:\text{DT-A}\) plants (Figure 2g, h; Wen et al., 2014). Thus, any fragment that possesses insulator function should, when interposed between \(NtAG1e\) and \(Pps::\text{DT-A}\) in \(pAGI1RACT\), attenuate the ability of \(NtAG1e\) to activate the \(Pps\) promoter in petals, stamens and carpels, thus reducing the incidence of the ablated flower phenotype and concomitantly increasing the proportion of sterile flower lines.

\(RS2-9\) and \(RS2-22\), along with the control \(CW148\) and \(\lambda\) spacer (shortened to 2 kb from 4 kb in \(JM85\); Figure S1) fragments, were inserted between the \(NtAG1e\) and \(Pps::\text{DT-A}\) fragments in \(pAGI1RACT\) (Figure 2a), and the resulting vectors were introduced into tobacco plants. Approximately half of the transgenic plants bearing the \(pAGI1RACT\) cassette alone produced sterile flowers, and the other half produced an ablated flower phenotype (Figure 2b), which is consistent with previous findings (Wen et al., 2014). The insertion of the \(RS2-9\) fragment reduced the incidence of the ablated phenotype to 7% and increased the proportion of sterile flower phenotype lines to 93%. However, the insertion of either the \(RS2-22\) or control \(CW148\) and \(\lambda\) spacer fragments neither reduced the proportion of lines with the ablated flower phenotype nor increased the incidence of sterile flower lines, and thus apparently failed to block enhancer-promoter interactions (Figure 2b). We also performed additional transformation experiments with these vectors and confirmed that \(RS2-9\) but not \(CW148\) or \(RS2-22\), significantly reduced the frequency of ablated flower lines to below 8% (Figure 2j). PCR amplification led to the detection of intact 1.2-kb \(RS2-9\) and 2.0-kb \(CW148\) fragments in the 14 transgenic lines analyzed (Figure S3), respectively, confirming the role of \(RS2-9\) in the insulation of the flower-specific activation function of \(NtAG1e\) in tobacco.

**Genes flanking RS2-9 in the rice genome exhibit distinct tissue specificity and histone modifications**

Insulators in *Drosophila* and mammals restrict enhancer-promoter-promoter interactions between neighboring genes to maintain the spatiotemporal expression specificity
of each gene (Narendra et al., 2015; Xu et al., 2014). Previously, it was demonstrated that two of three flower-specific promoters tested in plants were also capable of ectopically activating a target promoter in floral tissues, whereby the stronger promoter often overrode and imposed its activity on the weaker promoter (Wen et al., 2014), which is indicative of an interactive tendency of adjacent enhancers and promoters in plants. To determine whether the genes flanking RS2-9 in the rice genome are engaged in an interaction, we analyzed their tissue expression specificities and transcript abundances by searching the RiceXPro database (http://ricexpro.dna.affrc.go.jp). Two genes, Os04g0101700 and Os04g0101800, immediately flank (<500 bp) the RS2-9 region and are arranged in a tail-to-tail arrangement on chromosome 4 (Figure 3a), which rules out the possibility that RS2-9 acts as a potential promoter for either of them. The two flanking genes exhibit distinct tissue specificity, with Os04g0101700 being primarily expressed in the late stages of anther development and with Os04g010800 being predominantly expressed in roots (Figure 3b,c). In addition, the normalized signal intensity (>200), which correlates with transcript abundance, of Os04g010800 in many stages of stem, inflorescence, anther, pistil, lemma/palea, ovary and embryo development was at least 10 times higher than the highest signal intensity (approx. 200) of Os04g0101700 in the late stages of anther development (Figure 3b,c). These results indicate that in a native context, the two genes are not simultaneously regulated by mutual regulatory elements, suggesting a lack of interaction between them.

Insulators in other organisms are also known to block the spreading of histone marks into neighboring genes, thus preventing ectopic repression or activation (Cuddapah et al., 2009; Narendra et al., 2015; Xu et al., 2014). As such, we also assessed the histone modification state of the two RS2-9 flanking genes to determine whether active and repressive histone marks are strictly restricted to specific sites or whether they are spread across the RS2-9 insulator. Of a total of 15 histone modifications analyzed (Table S4), 10 were differentially partitioned across RS2-9 (Figure 3d). Active histone modifications, including the acylation of H3K9 (H3K9ac), H4K12 (H4K12ac), H3K23 (H3K23ac), H4K16 (H4K16ac) and H3K4 (H3K4ac), as well as the di- and tri-methylation of H3K36 (H3K36me2 and H3K36me3, respectively), predominantly accumulated at the highly expressed Os04g0101800 locus. Conversely, repressive histone modifications, including the mono-, di- and tri-methylation of H3K9, were preferentially enriched at the less active Os04g0101700 locus. In particular, H3K9me2, a key repressive mark linked to DNA methylation that is prevalent in heterochromatin and in silenced genes in plant genomes (Liu et al., 2010), was preferentially accumulated in the promoter region of Os04g0101700 but was rarely observed in Os04g0101800. Thus, active and repressive histone modifications are differentially partitioned in the genes flanking RS2-9.

**Oryza sativa** homebox 1 (OSH1) binds to RS2-9

The function of insulators in *Drosophila* and mammals is dependent upon the binding of protein factors (Ali et al., 2016; Bell et al., 1999; Gaszner et al., 1999; Szabo et al., 2019; Zhao et al., 1995). To search for such binding factors in rice, we retrieved and analyzed publicly available chromatin immunoprecipitation sequencing (ChIP-seq) data sets (generated with antibodies against five rice protein factors) collected in the Plant Chromatin State Database (PCSD) database (http://systemsbiology.cau.edu.cn/chromstates; Table S4) (Liu et al., 2018). We found that OSH1, which is a key transcription factor involved in the control of indeterminate growth of the shoot apical meristem in rice (Tsuda et al., 2011), bound to the RS2-9 region with two strong binding peaks (Figure 4a), each of which was significantly enriched at least ninefold (Fisher’s test, $P = 6.934 \cdot 10^{-10}$; Table S3). A smaller, single binding peak was also observed within an intronic region of Os10g0548400, but no evident binding activity was detected in the remaining 11-kb genomic region assessed (Figure 4a). Interestingly, the patterns of DNase hypersensitivity and cleavage of micrococcal nuclease (MNase), which specifically releases mononucleosomes from RS2-9 chromatin, appeared to remain invariant across the RS2-9 chromatin region (Figure 4a).

The two OSH1-binding peaks are clustered with distinct binding motifs

To define the two OSH1 binding sites precisely, we remapped the ChIP-seq reads to the RS2-9 sequence and showed that the two binding peaks occupied an approximately 100-bp region between 400 and 500 bp in peak I (PI) and 900 and 1000 bp in peak II (PII), respectively (Figure 4b). We then examined the sequences for OSH1-binding motifs in the two binding peaks. Previous studies identified and functionally validated a TGAC motif as a primary binding site for both OSH1 and its Zea mays (maize) ortholog, KN1 (Bolduc and Hake, 2008; Tsuda et al., 2011). Recent bioinformatic analysis of the OSH1 ChIP-seq data also delineated at least 10 further consensus binding motifs, hereafter referred to as non-TGAC motifs, with motifs I and III preferentially located in the binding peaks and the remainder residing in peripheral binding regions (Tsuda et al., 2014). Hence, unlike the CTCF binding site, which harbors a 40-bp conserved consensus sequence motif (Klenova et al., 1993; Nakahashi et al., 2013; Rhee and Pugh, 2011), the OSH1 binding motif is less well defined, and appears to be diverse and to have a tendency to be clustered in nature (Bolduc and Hake, 2008; Tsuda et al., 2011, 2014). We scanned for the TGAC motif, as well as motifs I and III, in the RS2-9 region and identified a TGAC motif as a primary binding site for both OSH1 and its Zea mays (maize) ortholog, KN1 (Bolduc and Hake, 2008; Tsuda et al., 2011, 2014). We scanned for the TGAC motif, as well as motifs I and III, in the RS2-9
sequence and identified a total of 21 motifs, including 12 motifIs, six TGAC motifs and three motifs sharing a signature sequence (GAT/ATC) with motif III (hereafter referred to as putative motif III) (Figure 4b). Although all motifIs were randomly distributed along RS2-9, the three putative motif IIIIs were exclusively clustered in PI and five of the six TGAC motifs were clustered in PII (Figure 4b,c). Notably, the TGAC and putative motif III/non-TGAC motifs were not co-located but were preferentially clustered in specific binding peaks.
Both PI and PII binding sites are required for RS2-9 activity, but PI is more critical

To understand whether these motifs are required for insulation function, we performed mutagenesis analysis. A previous study demonstrated that the abolishment of OSH1 regulatory function requires the simultaneous mutation of multiple, clustered TGAC motifs (Tsuda et al., 2011), which suggests that multiple TGAC motifs are required for the stable binding and regulatory function of OSH1. This mechanism appears to be distinct from other transcription factors, which typically bind to a single, specific motif rather than multiple clustered motifs (Busch et al., 1999). Accordingly, we created mutations in three putative motif IIIs in PI (\(M1\)), five TGAC motifs in PII (\(M2\)) and a combination of mutations (\(M1M2\)) in both peak regions of RS2-9, respectively (Figure 4c; Figure S4). We then assessed their insulation capacities in transgenic Arabidopsis plants, as described above (see Figure 1c). Both \(M1\) and \(M2\) mutations significantly increased the frequency of GUS staining in T1 seedlings, compared with unaltered RS2-9 plants. However, this increase varied substantially between the two mutated sites, with \(M1\) seedlings exhibiting a GUS staining frequency of 60% and \(M2\) seedlings displaying a GUS staining frequency of 18%, both of which were significantly higher than the 10% frequency observed in the RS2-9 seedling population (Figure 4d). These results indicate that the \(M1\) mutations were threefold more effective at impairing insulation function than the \(M2\) mutations. Combining the \(M1\) and \(M2\) mutations (\(M1M2\)) further increased the frequency of GUS-positive seedlings up to 82%, which approached that observed in control CW148 seedlings (Figure 4d). Evidently, although the two binding sites differentially contribute to the insulation function of RS2-9, both are required for RS2-9 activity.

OSH1 binding occurs genome wide, but its binding sites are predominantly associated with the boundaries of TADs

We analyzed the ChIP-seq data (Tsuda et al., 2014) further and identified more than 50 000 OSH1 binding sites in the rice genome.
rice genome (Table S3). Over 80% of these binding sites resided within intergenic and gene-flanking regions, whereas less than 20% were present in either exonic or intronic regions (Figure 4e). This distribution pattern is similar to that identified for CTCF in human cells (Kim et al., 2007) and BEAF-32 in Drosophila (Emberley et al., 2008), respectively. As the binding of CTCF and BEAF-32 are associated with the demarcation of TAD boundaries (Dixon et al., 2012; Phillips-Cremins and Corces, 2013; Rao et al., 2014; Wang et al., 2018), we performed association analysis between the boundaries (approx. 5 kb) of TADs identified from the analysis of Hi-C data derived from rice seedlings (Figure S5a,b; Liu et al., 2017), and the OSH1 binding sites (Table S3) identified from the analysis of ChIP-seq data from panicle tissues (Tsuda et al., 2014). In a recent study, prominent differences in local chromatin structure were not observed across different plant tissues, even when the tissues exhibited distinct transcriptional profiles (Dong et al., 2020; Zhou et al., 2019). Therefore, in our opinion, the TAD pattern from rice seedlings is likely to be comparable overall with that from panicles, making it reasonable to explore a potential relationship between TAD annotation and OSH1 targets using these two distinct tissues. We found that OSH1 was predominantly enriched in TAD boundaries compared with randomly assigned regions (control) within the rice genome, with more enrichment observed within, rather than outside of, TADs (Figure 5). Of the 49 144 consensus OSH1-binding sites shared by both Rep1 and Rep2 in the ChIP-seq data (Table S3), 3614 overlapped with 2412 of the total 3370 TAD boundaries reported (Figure S5; Liu et al., 2017). To determine whether the overlapping regions between TAD borders and OSH1 binding sites occurred through random chance, we calculated the number of expected overlapping sites after randomly permuting the location of TADs in individual chromosomes, whereby TAD borders were shuffled. This process led to the presence of approximately 1700 overlapping sites, which is considerably lower than our observation (Figure S5). Indeed, OSH1 bound to 72% of TAD borders (Figure S5), which is comparable to the 77 and 74% of CTCF- and BEAF-32-bound TAD boundaries found in mammals (Dixon et al., 2012) and Drosophila (Wang et al., 2018), respectively.

DISCUSSION

The function of an EB insulator is defined by its ability to block the action of an enhancer on an adjacent promoter in a transgenic assay when the insulator is interposed between the enhancer and the promoter. Such a system has been used extensively for the isolation and characterization of a multitude of insulators in Drosophila and mammals, to date (Bell et al., 1999; Chung et al., 1993; Kellum and Schedl, 1992). We adopted the same strategy in the current study, and demonstrated that the RS2-9 element, which resides in an intergenic region between divergent genes in the rice genome (Figure 3a), effectively attenuated the activation function of the strong, constitutive, virus-derived CaMV35Se in Arabidopsis (Figure 1b–k), as well as the flower-specific, plant-derived NtAG1e in tobacco (Figure 2a–j). In neither case did the presence of the RS2-9 fragment affect the activity of the promoters upon which the two enhancers act (Figures 1b–k and 2b–j), ruling out the possibility that RS2-9 acts directly as a transcriptional repressor.

Given that the enhancer blocking function and binding motifs of CTCF are conserved between organisms as distant as vertebrates and insects (Heger et al., 2012; Heger and Wiehe, 2014), it is likely that the function of a plant-derived insulator would also be conserved between dicot and monocot species. Thus, the enhancer blocking function of RS2-9 demonstrated in a transgenic context in Arabidopsis and tobacco is expected to reflect its endogenous function in rice. In the rice genome, the genes immediately flanking RS2-9 exhibit distinct tissue specificity and expression levels (Figure 3b,c), which corresponds well with the differential expression patterns frequently observed in genes flanking insulators in Drosophila and mammals (Phillips and Corces, 2009). Similarly, the absence of spreading of active or repressive histone marks across RS2-9 (Figure 3d) is also in line with reports that insulators hinder the migration of histone modifications in Drosophila and mammals (Bowman et al., 2014; Narendra et al., 2015; Xu et al., 2014). As such, it appears that RS2-9 shares similar functional characteristics with its Drosophila and mammalian counterparts.

Interestingly, although the RS2-22 element was able to effectively block the activation function of the CaMV35Se in Arabidopsis (Figure 1b–k), it failed to obstruct activation by the plant-derived, flower-specific NtAG1e in tobacco (Figure 2b–j), which raises the possibility that the function of RS2-22 may be either plant or enhancer specific. In Drosophila, certain insulators display enhancer-specific action (Ohhtsuki and Levine, 1998), a phenomenon that is particularly pronounced for the gypsy insulator, the function of which depends primarily upon insulator anatomy and enhancer strength, and often becomes much less effective in the presence of strong enhancers (Scott and Geyer, 1999). Conceivably, the loss of the insulation function of RS2-22 in tobacco could be ascribed to a lack of specificity for, or the strong action of, the NtAG1 enhancer.

The activities of insulators characterized to date are known to depend upon the binding of specific protein factors (Bell et al., 1999; Ramírez et al., 2018; Van Bortle et al., 2014; Wang et al., 2018). In line with this, we also demonstrated that OSH1 strongly and significantly bound RS2-9 (Figure 4a,b), with at least ninefold read enrichment (Fisher’s test, $P = 6.934E-101$; Table S3). OSH1 has been previously identified as a class-1 KNOTTED 1-LIKE HOMEBOX
(KNOX) gene encoding a homeodomain transcription factor that primarily controls shoot apical meristem (SAM) maintenance in rice. This function is evidenced by the fact that osh1 loss-of-function mutants display termination of growth immediately after germination because of the lack of a shoot apical meristem (Tsuda et al., 2011). Similarly, loss-of-function mutants of its orthologous genes, such as KNOTTED 1 (KN1) in maize and SHOOT MERISTEMLESS (STM) in Arabidopsis, also lack a shoot apical meristem (Long et al., 1996; Vollbrecht et al., 2000), indicating that the function of OSH1 and its orthologs are highly conserved in both dicot and monocot plants. Interestingly, OSH1 is also known to be involved in the activation and repression of gene expression in rice (Song et al., 2018; Tsuda et al., 2011), and reporter gene expression is not influenced by the mutation of an OSH1 binding motif in the associated promoter in transgenic rice (Tsuda et al., 2011). Furthermore, the maize ortholog, KN1, previously demonstrated a lack of direct transcriptional activation or repression function in a transgenic assay (Bolduc and Hake, 2009). Taken together, these findings contradict the typical transcription factor-mediated gene regulation that is normally observed in plants (Busch et al., 1999; Ito et al., 2003; Raatz et al., 2011).

The binding of OSH1 to two sites approximately 400-bp apart in RS2-9 is unique (Figure 4a,b), especially given that both binding sites are required for full enhancer blocking activity (Figure 4c). Each of the two binding sites is clustered with distinct classes of binding motifs, with PI encompassing three copies of the putative motif III, and with PII harboring five copies of the TGAC motif (Figure 4b). A similar binding motif cluster is also found in the binding sites of Drosophila BEAF-32 and su(Hw) insulator proteins, with the number of binding motifs in each cluster appearing to positively correlate with the strength of the insulators (Geyer and Corces, 1992; Zhao et al., 1995). Whether the number of binding motifs in PI and PII sites is correlated with the strength of the RS2-9 insulator remains to be investigated. Although the PI binding site encompassing three copies of the putative motif III was found to be more critical than the PII site harboring five copies of the TGAC motif, both sites were indispensable for full insulator function (Figure 4d). One might argue that the mutated motifs could impede the binding of protein factors other than OSH1 orthologs in Arabidopsis. However, this appears highly unlikely because both the putative motif III and TGAC binding motifs are consistently present in the thousands of OSH1-binding sites in the rice genome (Tsuda et al., 2014; Table S3), and the TGAC motif has been experimentally validated in vivo and in vitro as a binding site of OSH1 (Tsuda et al., 2011) and its maize homolog KN1 (Bolduc and Hake, 2009). Hence, mutation of the OSH1 binding motifs at PI and PII sites are likely to abolish the binding of the counterpart of OSH1 in Arabidopsis. It is feasible that STM may have played this role because of the fact that OSH1, maize KN1 and Arabidopsis STM are so highly conserved, both physically and functionally (Bolduc and Hake, 2009; Long et al., 1996; Tsuda et al., 2011).

It remains unclear why RS2-9 requires two OSH1 binding sites rather than one, as has been demonstrated for mammalian and Drosophila insulators (Bell et al., 1999; Geyer and Corces, 1992; Roseman et al., 1993; Zhao et al., 1995). However, a subset of insulators (approx. 1700) bound by BEAF-32 also contain two binding sites separated by an approximately 200-bp spacer (referred to as a dual-core binding site), and each binding site similarly harbors multiple BEAF-32 binding motifs (Emberly et al., 2008), which is very similar to the binding pattern of OSH1 in RS2-9 (Figure S4b,c). Mutagenesis and molecular characterization similarly revealed that both binding sites are required for EB function of the dual-core insulators. Dual-core insulators are associated with a distinct set of genes specifically involved in chromosome organization/segregation, protecting them from repressive chromatin modifications (Emberly et al., 2008). Thus, RS2-9 appears to represent a similar dual-core insulator with a strong insulation function in plants.

One of the features that insulator-binding proteins exhibit is genome-wide binding activity. BEAF-32 binds up to 10 000 sites in the Drosophila genome (Emberly et al., 2008) and CTCF also displays a genome-wide occupancy tendency in mammalian genomes, with 13 000 sites identified in IMR90 human fibroblasts (Kim et al., 2007), 26 814 sites in resting human CD4+ T cells (Jothi et al., 2008) and 39 609 sites in mouse embryonic stem cells (Chen et al., 2008). In both cases, the insulator proteins preferentially bind within intergenic and gene-flanking regions (Emberly et al., 2008; Kim et al., 2007). Similarly, OSH1 also binds up to 50 000 sites in the rice genome (Table S3), with a similar preference for intergenic and gene-flanking regions (Figure 4e). Functionally, CTCF and BEAF-32 are essential for embryonic development, even though they share no evolutionary relationship (Heath et al., 2008; Emberly et al., 2008). Interestingly, the osh1 loss-of-function mutant exhibits impaired meristem activity that is arrested immediately after germination (Tsuda et al., 2011), which indicates that OSH1 shares not only a similar genome-binding activity but also an essential function with CTCF and BEAF-32.

The predominant association of OSH1 binding sites with the boundaries of TADs identified in the rice genome suggests that OSH1 may also share a similar role with CTCF and BEAF-32 in the regulation of TADs. In mammals, CTCF, together with the cohesin complex, exploits a loop extrusion mechanism to define the boundaries where TADs anchor (Dixon et al., 2012; Fudenberg et al., 2016; Nora et al., 2012; Sexton et al., 2012; Vian et al., 2018). In Drosophila, such a function is carried out mainly by the
insulator complexes BEAF-32/CP190 or BEAF-32/Chromator, but not by dCTCF, GAF, mod (mdg4), su(Hw) or ZW5 (Rowley et al., 2017; Wang et al., 2018). Plant genomes appear to undergo similar topological organization, but the size and density of TADs vary from plant to plant (Dong et al., 2017; Liu et al., 2017; Ouyang et al., 2020; Wang et al., 2015). In rice, over 1700 typical TADs are identified with a median size of 45 kb, which collectively covers approximately 25% of the genome (Liu et al., 2017). OSH1 binding sites are associated with 72% of the boundaries of rice TADs (Figure S5), which is akin to the 74 and 77% association rates of BEAF-32 and CTCF binding sites, respectively (Dixon et al., 2012; Wang et al., 2018). Hence, OSH1 binds to not only the RS2-9 insulator but also the boundaries of TADs, implying its involvement in the demarcation of chromatin loops in rice.

Interestingly, the \textit{in silico} predicted consensus sequence motifs potentially recognized by the TCP gene family are also abundantly enriched in the boundaries of TADs in rice (Liu et al., 2017). Furthermore, BORDER1, 2 and 3 protein factors, which are involved in transcriptional pausing in Arabidopsis, are also required for insulating interactions between neighboring genes located tandemly on the same strand (Yu et al., 2019), but whether they are associated with the boundaries of TADs remains unknown. Further genetic, molecular and genomic analyses should shed light on the role of OSH1, TCP and BORDER proteins in genetic insulation, chromatin loop dynamics and chromosome architecture in plants. Nevertheless, the present work identifies RS2-9 as a true EB insulator, and also characterizes its binding factor OSH1. These findings are of practical and fundamental significance for remediating ectopic enhancer–and promoter–promoter interactions that are prevalent in a transgenic context, and also for understanding potential topological chromatin loop-based transcriptional regulation in plants.

\textbf{EXPERIMENTAL PROCEDURES}

\textbf{Plasmid construction, isolation of genomic fragments from the rice genome, plant transformation and histochemical GUS assays}

A high-throughput insulator assay vector, \textit{pSYG3A}, was constructed by converting a conventional insulator assay vector, \textit{JM69} (Hily et al., 2009), into a Gateway cloning-compatible vector. Specifically, a \textit{RcI} 1 fragment containing the chloramphenicol resistance gene (\textit{Cmr}) and the \textit{ccdB} gene flanked by attR and attB sites (Invitrogen, now ThermoFisher Scientific, http://www.thermofisher.com/us/en/home.html) was directly cloned into an end-filled Ncol site in \textit{pAXU3121} (Goderis et al., 2002), where the same \textit{RcI} 1 fragment was released by \textit{I-SceI} digestion and cloned into the same site in \textit{JM69} to create \textit{pSYG3A} (Figure 1a). \textit{pSYG3A} served as a recipient vector for the high-throughput cloning of genomic fragments through attL-attR (LR) site-specific recombination reaction with the donor plasmid, \textit{pCR8/GW/TOPO} (Invitrogen, now ThermoFisher Scientific). Potential EB insulator candidates were identified through the analysis of genome-wide transcriptome profiles. Intergenic regions that are flanked by genes with distinct tissue specificity or transcriptional activity were identified as potential insulator candidates, isolated by PCR amplification and cloned into the \textit{pCR8/GW/TOPO} vector according to the manufacturer's instructions (Invitrogen, now ThermoFisher Scientific). A total of 59 genomic fragments ranging from 1 to 2 kb were isolated and successfully inserted into \textit{pSYG3A} (Figure 1a; Table S1). The design of sequence mutation of OSH1 binding motifs (M1, M2 and M1/2M2) is detailed in Figure S4, and entire insulator sequences bearing specific mutations were commercially synthesized by Synbio Technologies (https://www.synbio-tech.com). The synthesized fragments were cloned into \textit{pSYG3A} before introduction into Arabidopsis. The construction of plasmids \textit{JM85}, d35S\textit{AEG1::GUS/}\textit{JM69}, \textit{pBASE\textit{ACT, nTAgI1::DT-A}} and \textit{pAGI1\textit{R\textit{ACT}}} has been described previously (Hily et al., 2008; Liu and Liu, 2008; Wen et al., 2014). Cloning of the \textit{RS2-9}, \textit{RS2-22}, \textit{CW148} and a 2-kb bacterial\textit{λ} sequence from \textit{JM85} (hereafter referred to \textit{CW/\textit{P}}) fragments into \textit{pAGI1\textit{R\textit{ACT}}} was carried out as described previously (Wen et al., 2014; Figure 2a), with the exception of the insertion of the \textit{RcI} 1 fragment between the \textit{NtAgI1} and \textit{Pps::DT-A} cassette in \textit{pAGI1\textit{R\textit{ACT}}} to allow Gateway-mediated site-specific recombination between the donor and recipient plasmids, as described above.

The resulting binary vectors were introduced into \textit{Agrobacterium tumefaciens} strain GV3101 via electroproporation and were then transformed into Arabidopsis Col-0 and tobacco \textit{SR1} accessions, as described previously (Hily et al., 2009; Wen et al., 2014; Yang et al., 2010). Transformed \textit{T\textsubscript{0}} Arabidopsis seeds were screened on MS medium containing 50 mg L\textsuperscript{-1} kanamycin, and selected \textit{T\textsubscript{0}} seedlings were transferred to soil for downstream analysis. Transformed \textit{T\textsubscript{0}} tobacco plants were transferred to a glasshouse for the evaluation of flower phenotypes in mature plants. Histochemical GUS assays were performed as described previously (Hily et al., 2009).

\textbf{Verification of transgene insertion by PCR amplification}

Genomic DNA was isolated from leaf tissue of selected transgenic lines using the dodecyltrimethylammonium bromide (D-TAB) method, as described previously (Hily et al., 2009). Primers \textit{SGY4AscI1U1/AttB1} (5'-ATCGCTCGCGACCTGCAGGCATAA-3') and \textit{SGY4AscID1/AttB2R} (5'-GTTAGCGAGCTTTAATATTAGGGA-3') were used for the amplification of insulator sequences in both transgenic \textit{Arabidopsis} and tobacco plants. In parallel, the tobacco \textit{NtACT\textit{B6}} and \textit{Arabidopsis EF1\textit{A}} genes were amplified as internal controls, as described previously (Hily et al., 2009; Wen et al., 2014). Approximately 100 ng of genomic DNA was used as the template for amplification, and PCR reactions were performed using GoTag DNA Polymerase according to the manufacturer's instructions (Promega, https://www.promega.com). Cycling conditions were as follows: 94°C for 5 min, 30 cycles of 94°C for 30 s, 55-65°C for 1-3 min and 72°C for 3 min, followed by a single extension cycle at 72°C for 7 min. Amplified DNA samples were resolved on 0.8% agarose gels.

\textbf{Real-time quantitative RT–PCR analysis}

Real-time quantitative RT–PCR was carried out as described previously (Hily et al., 2009; Singer et al., 2010a). Briefly, total RNA was extracted from young leaf tissues of five independent transgenic lines bearing each construct, respectively, using the RNaseasy Plant Mini kit (Qiagen, https://www.qiagen.com) and treated with DNAase (Invitrogen, now ThermoFisher Scientific, http://www.thermofisher.com/us/en/home.html) to remove genomic DNA. Approximately 400 ng of treated RNA was reverse transcribed into...
cDNA in each case using the Superscript VILO cDNA synthesis kit (Invitrogen, now ThermoFisher Scientific). Following dilution by 50-fold, an aliquot of 1.5 μl cDNA was used for qRT-PCR analysis with the ABI Prism 7900HT detection system and the SYBR Green Master Mix (Invitrogen, now ThermoFisher Scientific). All reactions were performed in triplicate. Primer pairs for the amplification of GUS and UBQ10 have been reported previously (Hily et al., 2009). The PCR cycling conditions were as follows: 95°C for 10 min, followed by 40 cycles of 95°C for 15 sec and 60°C for 1 min. Relative gene expression was determined from standard curves produced using a series of cDNA dilutions. The resulting GUS mRNA levels were normalized to the level of UBQ10 mRNA and statistically analyzed using the Mann–Whitney U-test (P ≤ 0.01).

Tissue-specific expression analysis

Examination of the development- and tissue-specific expression patterns of the genes flanking RS2-9 was performed within the RiceXPro database (http://ricexpro.dna.affrc.go.jp), where 12 data sets encompassing 572 microarray experiments from six tissue types at various developmental stages were deposited and analyzed (Sato et al., 2013), including leaf, root and stem tissues collected from rice plants at vegetative, reproductive and ripening stages at 12:00 h and 0:00 h, and inflorescence, anther, pistil, lemma and palea, ovary, embryo and endosperm tissues collected at various stages of their development (detailed stages are described in Figure 3b,c). The web interface enabled the extraction and snapshot of graphed expression. A similar analysis was also carried out in the eFP Rice database (http://bar.utoronto.ca/eFP/ cgi-bin/eFPWeb.cgi), where two sets of independent microarray data sets were previously deposited (Jain et al., 2007; Ma et al., 2005).

Computational analyses

To analyze histone modifications in the RS2-9 fragment, as well as the potential binding of protein factors, a total of 41 data sets (Table S4) were downloaded from the Sequence Read Archive (http://www.ncbi.nlm.nih.gov/sra) (Kodama et al., 2012), respectively. The details of such computational analyses have been described previously (Liu et al., 2018). Briefly, quality control of the downloaded data was performed using FASTQC 0.10.1 (http://www.bioinformatics.babraham.ac.uk/projects/fastqc) and low-quality reads were filtered out using the FASTX TOOLKIT 0.0.13 (http://hannonlab.cshl.edu/fastx_toolkit). Adaptor sequences were trimmed with CUTADAPT 1.4.1 (http://cutadaptr.readthedocs.io/en/stable). Bowtie2 2.1.0 was used to align sequencing reads with the rice reference genome TIRG6.1 or IRGSP-1.0, with default parameters. MACS 1.4.1 was used to call enriched regions and calculate read abundance (represented as reads per million mapped reads, RPM, at a resolution of 10 nt), which was then graphically represented using the UCSC Genome Browser in Plant Chromatin State Database (Liu et al., 2018). Regions with significant enrichment (compared with the expectation from Poisson distribution with local lambda) of the OSH1-bound regions were identified from 3,564,691 intersecting peaks using the MACS 1.4.1 with set parameters ‘-g 3.8e8’ (genome size was set as 3.5e8), P < 0.05 and ‘nomodel’ (see results in Table S3). All other parameters were set as the default. The genome distribution of OSH1 binding sites was analyzed using CEGAS 1.0.2 and all information regarding the genomic locations, sizes and fold changes is presented in Table S3. Further verification of mapping patterns and peaks of ChIP-seq data (DRX000457) in the 1.2-kb RS2-9 region was conducted using CLC IO 20.

ACKNOWLEDGEMENTS

This project was supported by Syngenta Crop Protection LLC (CRADA no. 58-3K95–0-1402) and United States Department of Agriculture (USDA) Agricultural Research Service (ARS) in-house funding.

AUTHOR CONTRIBUTIONS

ZL and QQ conceived and designed the experiments. HL, LJ, ZW, YY, DB, JC and ZL performed the experiments. WX, SS, JC, HC, QQ, ZS, ZY, YL, CL and ZL analyzed the data. ZL wrote the article.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest associated with this work.

DATA AVAILABILITY STATEMENT

Data supporting the findings of this work are available within the article and in the Supporting Information. The RNA-seq and ChIP-seq data used in this work were obtained from public databases, which are cited in the text or listed in Tables S2–S4. The experimental materials and data sets generated and analyzed during the current study are available from the corresponding author, upon request.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Evaluation of the lack of insulation function of randomly cloned DNA fragments (approx. 2 kb) from Arabidopsis (CW146, CW147, CW148) and Glycine max (soybean; CW12044) genomes that were inserted between the duplicated CaMV 35S enhancer (d35Se) and AGIP::GUS.

Figure S2. Verification of the transgene in transformed Arabidopsis lines.

Figure S3. Verification of the transgene in transformed tobacco lines.

Figure S4. Mutagenesis analysis of the function of OSH1 binding motifs.

Figure S5. Preferential association of OSH1-binding sites with TAD boundaries.

Table S1. Nucleotide sequences of 59 rice fragments and control spacer fragments used for the evaluation of the insulation function in Arabidopsis.

Table S2. Retrieved genome sequence of the ChIP-seq-identified OSH1 binding region overlapping the RS2-9 sequence.

Table S3. List of total number of identified OSH1 binding sites with significant read enrichment in the rice genome (IRGSP-1.0) in two ChIP-seq replicates (DRX000321 and DRX000457).

Table S4. List of the downloaded data sets used in this work.

REFERENCES

Akdemir, K.C., Le, V.T., Chandran, S., Li, Y., Verhaak, R.G., Beroukhim, R. et al. (2020) Disruption of chromatin folding domains by somatic genomic rearrangements in human cancer. Nature Genetics, 52, 294–305.
Maurano, M.T., Wang, H., John, S., Shafer, A., Canfield, T., Lee, K. et al. (2015) Role of DNA methylation in modulating transcription factor occupancy. Cell Reports, 12, 1184–1195.

Nakahashi, H., Kwon, K.-R., Resch, W., Vian, L., Dose, M., Stavreva, D. et al. (2013) A genome-wide map of CTCF multivalency redefines the CTCF code. Cell Reports, 3, 1678–1689.

Narendra, V., Rocha, P.P., An, D., Raviram, R., Skok, J.A., Mazzeni, E.O. et al. (2015) Transcription. CTCF establishes discrete functional chroma-
tin domains at the Hox clusters during differentiation. Science, 347, 1017–1021.

Negre, N., Brown, C.D., Shah, P.K., Kheradpour, P., Morrison, C.A., Henikoff, J.G. et al. (2010) A comprehensive map of insulator elements for the droso-
phila genome. PLoS Genetics, 6(1), e1000814.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N. et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature, 485, 381–385.

Ohnuki, S. & Levine, M. (1998) GAGA mediates the enhancer blocking activ-
ity of the eve promoter in the Drosophila embryo. Genes & Develop-
ment, 12, 3236–3330.

Ouyang, W., Xiong, D., Li, G. & Li, X. (2020) Unraveling the 3D genome archi-
tecture in plants: present and future. Molecular Plant, 13, 1678–1693.

Phillips, J.E. & Corces, V.G. (2009) CTCF: master weaver of the genome. Cell, 137, 1194–1211.

Phillips-Cremins, J. & Corces, V. (2013) Chromatin insulators: linking genome organization to cellular function. Molecular Cell, 50, 461–474.

Raatz, B., Eicker, A., Schmitz, G., Fuss, E., Muller, D.R., Rossmann, S. et al. (2011) Specific expression of LATERAL SUPPRESSOR is controlled by an evolutionarily conserved 3' enhancer. Plant Journal, 68, 402–412.

Ramirez, F., Bhardwaj, V., Arrington, L., Lam, K.C., Groning, B.A., Villaveces, J. et al. (2010) High-resolution TADs reveal DNA sequences underlying genome organization in flies. Nature Communications, 9, 189.

Rao, S., Huntley, M., Durand, N., Stamenova, E., Bochkov, I., Robinson, J. et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell, 159, 1685–1688.

Rhee, H.S. & Pugh, B.F. (2011) Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell, 147, 1409–1419.

Rosenman, R.R., Pirrotta, V. & Geyer, P.K. (1993) The su(Hw) protein insu-
lates expression of the Drosophila melanogaster white gene from chro-
mosomal position-effects. The EMBO Journal, 12, 435–442.

Rowley, M.J., Nichols, M.H., Lyu, X., Ando-Kuri, M., Rivera, I., Herzenz, K. et al. (2017) Evolutionarily conserved principles predict 3D chromatin organization. Molecular Cell, 67, 837.

Sato, Y., Takehisa, H., Kamatsuki, K., Minami, H., Namiki, N., Ikawa, H. et al. (2013) RiceXpro version 3.0: expanding the informatics resource for rice transcriptome. Nucleic Acids Research, 41, 1206–1213.

Scott, K.S. & Geyer, P.K. (1995) Effects of the su(Hw) insulator protein on the expression of the divergently transcribed Drosophila yolk protein genes. The EMBO Journal, 14, 6258–6267.

Scott, K.C., Taubman, A.D. & Geyer, P.K. (1999) Enhancer blocking by the Drosophila gypsy insulator depends upon insulator anatomy and enhancer strength. Genetics, 153, 787–798.

Sexton, T., Yaffe, E., Kenigsberg, E., Bantignies, F., Leblanc, B., Hoichman, M. et al. (2012) Three-dimensional folding and functional organization principles of the drosophila genome. Cell, 148, 458–472.

Sigrist, C.J. & Pirrotta, V. (1997) Chromatin insulator elements block the silencing of a target gene by the Drosophila polycistronic repeat response element (PRE) but allow transinteractions between PREs on different chromo-
somes. Genetics, 147, 209–221.

Singer, S.D. & Cox, K.D. (2015) A gypsy-like sequence from Arabidopsis thaliana exhibits enhancer-blocking activity in transgenic plants. Journal of Plant Biochemistry and Biotechnology, 22, 35–42.

Singer, S.D., Cox, K.D. & Liu, Z. (2010a) Both the constitutive Cauliflower Mosaic Virus 35S and tissue-specific AGAMOUS enhancers activate trans-
scription autonomously in Arabidopsis thaliana. Plant Molecular Biology, 74, 293–305.

Singer, S.D., Hilly, J.-M. & Liu, Z. (2010b) A 1-kb Bacteriophage lambda frag-
ment functions as an insulator to effectively block enhancer-promoter interactions in Arabidopsis thaliana. Plant Molecular Biology Reporter, 28, 69–76.