DR YOSHIYUKI TANAKA (Orcid ID : 0000-0002-9298-8671)

Article type : Original Article

Positional differences of intronic transposons in *pAMT* affect the pungency level in chili pepper through altered splicing efficiency

Yoshiyuki Tanaka[2]*, Takaya Asano[1], Yorika Kanemitsu[1], Tanjuro Goto[1], Yuichi Yoshida[1], Kenichiro Yasuba[1], Yuki Misawa[3], Sachie Nakatani[3], Kenji Kobata[3]

[1]Graduate School of Environmental and Life Science, Okayama University, Okayama 700-8530, Japan

[2]Graduate School of Agriculture, Kyoto University, Sakyo-ku, Kyoto 606-8502, Japan

[3] Graduate School of Pharmaceutical Sciences, Josai University; 1–1 Keyakidai, Sakado, Saitama 350-0295, Japan

*Corresponding author: ystanaka@kais.kyoto-u.ac.jp

Running title: Intronic transposon positions affect pungency

Key words: Chili pepper, *Capsicum chinense,* capsaicinoid, capsinoid, *hAT* transposon

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process, which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/tpj.14462

Summary

Capsaicinoids are unique compounds that give chili pepper fruits their pungent taste. Capsaicinoid levels vary widely among pungent cultivars, ranging from low-pungency to extremely pungent. However, the molecular mechanisms underlying its quantitative variation have not been elucidated. Our previous study identified various loss-of-function alleles of the *pAMT* gene, which led to low-pungency. The mutations in these alleles are commonly defined by *Tcc* transposon insertion and its footprint. In this study, we identified two leaky *pamt* alleles ($pamt^{L1}$ and $pamt^{L2}$) with different levels of pAMT activity. Notably, both alleles had a *Tcc* transposon insertion in intron 3, but the locations of the insertions within the intron were different. Genetic analysis revealed that $pamt^{L1}$, $pamt^{L2}$ and a loss-of-function *pamt* allele reduced capsaicinoid levels to about 50%, 10%, and less than 1%, respectively. $pamt^{L1}$ and $pamt^{L2}$ encoded functional pAMT proteins, but they exhibited lower transcript levels compared with the functional-type. RNA-seq analysis showed that intronic transposons disrupted splicing in intron 3, which resulted in simultaneous expression of functional *pAMT* mRNA and non-functional splice variants containing partial sequences of *Tcc.* The non-functional splice variants were more dominant in $pamt^{L2}$ than that in $pamt^{L1}$. This suggested that the difference in position of the intronic transposons could alter splicing efficiency, which led to different pAMT activities and reduced capsaicinoid content to different levels. Our results provide a striking example where intronic transposons caused allelic variations, which contributed to quantitative differences in secondary metabolite contents.

Introduction

Plants produce a diverse range of secondary metabolites for protection against environment stress, herbivory, and pathogens (Wink, 1988, 2003). The content of some metabolites alters a plant's worth for humans, for example for use as food and ornamental crops. This phenomenon has enhanced artificial selection (domestication and breeding) for the development of variation in metabolite contents (Gross and Olsen, 2010). This has included altering flower color, changing taste and flavor, and reducing the levels of undesirable compounds in food. The fruits of chili pepper (*Capsicum*) produce a unique alkaloid, capsaicin, and it and its analogs are collectively called capsaicinoids (Aza-Gonzalez et al., 2011). Capsaicinoids causes a pungent sensation when mammals eat the fruits, which makes peppers an important spice crop. Capsaicinoids also have various health-related functions in humans (Aza-Gonzalez et al., 2011; Luo et al., 2011). A high level of capsaicinoids is preferred when using peppers for spicy food and for medicinal purposes, whereas the absence or only a small amount of capsaicinoids is favorable for their use as a vegetable. Therefore, various cultivars with different capsaicinoid levels have been developed, ranging from non-pungent to extremely pungent.

Capsaicinoids are synthesized by the condensation of vanillylamine and branched fatty acids in the placental septum of chili pepper fruit (Mazourek et al., 2009; Aza-Gonzalez et al., 2011). Vanillylamine is produced from phenylalanine in the phenylpropanoid pathway, and branched fatty acids are derived from valine or leucine. In the final step, capsaicinoids are formed through the acylation of vanillylamine and branched fatty acids. Loss-of-function

Pun1 mutations have been cloned as a qualitative genetic factor responsible for non-pungency (Stewart et al., 2005). *Pun1* encodes an acyltransferase that acylates vanillylamine with a fatty acid to form a capsaicinoid. Loss-of-function *pun1* alleles lead to the complete shutdown of capsaicinoid production (Stewart et al., 2005, 2007; Stellari et al., 2010; Wyatt et al., 2012; Kirii. et al., 2017). In addition to *pun1*, loss-of-function mutation in putative aminotransferase (*pAMT*) and putative ketoacyl-ACP reductase (*CaKR1*) were reported as genetic factors leading to loss of pungency (Lang et al., 2009; Koeda et al., 2019). Recently, a MYB transcription factor (*CaMYB31*) regulating capsaicinoid biosynthesis was proposed, and a genetic study demonstrated that a new stop codon in *CaMYB31* leads to no pungency (Arce-Rodriguez and Ochoa-Alejo, 2017; Han et al., 2019).

Although all mutations reported so far have qualitative effects that lead to loss of capsaicinoids, a wide variation in capsaicinoid level among pungent accessions has been attributed to numerous quantitative trait loci (QTLs). Previous genetic studies identified several dozen QTLs affecting capsaicinoid content, but only few have been characterized in detail (Blum et al., 2003; Ben-Chaim et al., 2006; Yarnes et al., 2012; Han et al., 2018). Thus, little is known about the molecular mechanism underlying quantitative variation in capsaicinoid level among pungent accessions.

A series of our studies have demonstrated that loss-of-function of *pAMT* caused a drastic reduction in capsaicinoid content (Lang et al., 2009; Tanaka et al., 2010a). *pAMT* encodes an aminotransferase, which produces vanillylamine, a precursor of capsaicinoids, from vanillin (Curry et al., 1999). Mutations in *pAMT* decrease the content of vanillylamine and capsaicinoids, simultaneously causing

the accumulation of low-pungency capsaicinoid analogs named capsinoids, instead of capsaicinoids. Capsinoids share a similar structure to capsaicinoids, but they exhibit considerably lower pungency because of the presence of an ester group instead of an amide group (Yazawa et al., 1989; Kobata et al., 1998; Lang et al., 2009). Therefore, *pamt* is a useful genetic factor to achieve low pungency with altered proportions of capsaicinoids and capsinoids (Jang et al., 2015; Jeong et al., 2015). In total, nine loss-of-function *pamt* alleles have been found in the *Capsicum* genus (Lang et al., 2009; Tanaka et al., 2010a, b, 2015, 2018; Koeda et al., 2014; Park et al., 2015), and 6 alleles were isolated in *C. chinense*. *C. chinense* is one of the domesticated species in the *Capsicum* genus that originated from the Amazon and is widely consumed in the USA, Caribbean, and South America (Bosland and Votava, 2000; Moses et al., 2014). Our previous study showed that each low-pungency *C. chinense* accession possesses a different loss-of-function *pamt* allele, but mutations in these alleles are commonly defined by *Tcc* transposon insertion or its footprint (Tanaka et al., 2015, 2018). A common characteristic of *C. chinense pamt* alleles raises an expectation that the mobilization of a *Tcc* family transposon could be involved in further allelic variations contributing to the quantitative variation in pungency levels in this species.

In the present study, we report two mutant alleles of *pAMT* (*pamt*^{L1} and *pamt*^{L2}), which harbor a *Tcc* transposon in different positions in intron 3. Both alleles exhibited leaky pAMT activities, but *pamt*^{L2} had a more severe negative effect on capsaicinoid content. RNA-seq analysis with the two *pamt* mutant alleles showed intronic transposon-derived change in splicing pattern, which

resulted in simultaneous expression of functional *pAMT* mRNA and non-functional splice variants. Of note, non-functional splice variants were dominant in *pamt*^{L2} compared with *pamt*^{L1}. This study demonstrates an interesting example where differences in the position of an intronic transposon determine quantitative variation in capsaicinoid content. The two leaky *pAMT* alleles can provide insights into the domestication process of chili peppers and may also offer a breeding tool for the quantitative reduction of pungency.

Results

Identification of unique mutant accessions with altered capsaicinoid/capsinoid ratios

When we investigated *Capsicum* genetic bio-resources by evaluating the ratio of capsinoid to capsaicinoid content, we identified that one pungent accession, Devil's Yellow (DY), had a higher capsinoid ratio compared with other pungent accessions (Fig. 1). We calculated the capsinoid ratio in each accession using the following formula : capsinoid ratio = {capsinoid content/(capsaicinoid + capsinoid contents)} × 100 (%). A pungent accession, Red Habanero (RH), preferentially produced capsaicinoids over capsinoids. The capsaicinoid and capsinoid contents in RH were 14264 μg/gDW and 385 μg/gDW, respectively, in the mature green stage (Fig. 1a, b), and the capsinoid ratio was 2.7% (Fig. 1c). DY also mainly produced capsaicinoids over capsinoids, but its capsinoid ratio was 7.8%, which was significantly higher than that of RH (Fig. 1c).

We also found one unique low-pungency accession, LP13, which exhibited a lower capsinoid ratio than those in loss-of-function *pamt* accessions. Loss-of-function *pamt* accessions preferentially produced capsinoids over capsaicinoids, and they exhibited a capsinoid ratio >90%. The capsaicinoid and capsinoid content in LP13 was 162 μg/gDW and 450 μg/gDW, respectively, in the mature green stage (Fig. 1a, b), and the capsinoid ratio was 73.4% (Fig. 1c).

Vanillylamine and vanillyl alcohol contents, and pAMT activity in placental tissue

The levels of vanillylamine and vanillyl alcohol, precursors of capsaicinoids and capsinoids, were measured in each accession. The vanillylamine content in the placenta of RH was approximately four times higher than that of DY (Table 1). No vanillylamine was detected in the placentas of LP13 and AD2. In contrast, the vanillyl alcohol content in the placenta was highest in AD2, followed by in LP13. The ability of placental tissue to transform vanillin to vanillylamine was evaluated as pAMT activity using the previously reported method with some modifications (Lang et al., 2009). The pAMT activity of RH was the highest among the chili peppers we investigated, and DY had moderate pAMT activity compared with RH (Table 1). Although LP13 also had slight activity, AD2 did not have any activity of pAMT.

DY-type and LP13-type *pAMT* alleles contain *Tcc* transposon insertions at different positions in intron 3

The alternation in capsaicinoid/capsinoid ratio and the reduction in pAMT activity led to an expectation that DY and LP13 have mutations in the *pAMT* locus. In this study, the complete genomic sequences of *pAMT* in the two accessions were determined and compared with the functional *pAMT* allele. The functional *pAMT* is composed of 17 exons and 16 introns and has a DNA coding sequence (CDS) of 1360 bp encoding a predicted protein of 459 amino acids (Fig. 2a). The *pAMT* alleles derived from DY and LP13 also had 17 complete exons and encoded 459 amino acids (Fig. 2b, c). The CDS was identical between the DY and LP13 sequences. Sequence alignment with the RH-type identified seven SNPs at positions 373, 727, 742, 848, 1230, 1281, and 1375 in the CDS of DY-type and LP13-type. Six mutations were non-synonymous, which led to amino acid changes in the pAMT protein (Fig. S2). In order to judge whether the amino acid change had an effect on capsaicinoid/capsinoid production, the amino acid sequence derived from another pungent accession (CAP1035_T1) was included in the alignment. Even though CAP1035_T1-type *pAMT* harbored identical amino acid residues to the DY-type and LP13-type, it exhibited a low capsinoid ratio (Fig. S2; Table S1). Therefore, the mutations in CDS region did not explain the increase in capsinoid ratio in DY or LP13.

It was found that both DY-type and LP13-type alleles carried an identical 2.3 kb insertion in intron 3 (Fig. 2b, c). The insertion shared structural characteristics of an *hAT* family transposon, and its sequence was identical to the *Tcc* transposon reported previously (Tanaka et al., 2011b). Notably, the

locations of the *Tcc* transposon were different within intron 3 in the two alleles. Whereas the *Tcc* transposon was located 591 bp away from exon 4 in the DY-type, the insertion was 342 bp away from exon 4 in the LP13-type (Fig. 2b, c). It has been reported that one loss of function *pamt* allele, *pamt*^{5}, carried a *Tcc* transposon insertion in intron 3 and an 8 bp insertion in exon 6 (Tanaka et al., 2011b). In *pamt*^{5}, the short insertion in exon 6 caused a frameshift mutation, and the location of the *Tcc* transposon was 542 bp away from exon 4 (Fig. 2d).

Thus, we designed primers for genomic PCR to confirm the three different positions of the intronic transposons. The result of genomic PCR indicated that the position of *Tcc* transposon differed in each allele (Fig. S3).

DY-type and LP13-type *pAMT* alleles have different reductions in capsaicinoid content

To test the effect of different *pAMT* allelic types on capsaicinoid content, F_2 populations from a cross between RH and accessions harboring the three-types mutant *pAMT* alleles were constructed. DY, LP13, and AD2 were used as paternal parents for crossing with RH. The *pAMT* genotype in each individual was determined by genomic PCR to check for intronic *Tcc* insertion.

The relationships between DY-type *pAMT* and capsaicinoid contents were investigated in 122 RH×DY F_2 individuals. Capsaicinoid contents decreased significantly in DY-type homozygotes up to a half of the level in RH-type homozygotes (Fig. 3a). In contrast, the capsinoid content was three times higher in DY-type/DY-type, compared with that in RH-type genotypes (Fig. 3b). The DY-type homozygotes exhibited higher capsinoid ratios up to 11.0% (Fig. 3c). Next,

the association between LP13-type *pAMT* and capsaicinoid content was investigated in 117 RH×LP13 F_2 individuals. Median capsaicinoid contents in LP13-type/LP13-type were about one-tenth of than that in RH-type genotypes (Fig. 3d). Homozygotes with LP13-type had higher capsinoid contents, which resulted in high capsinoid ratios up to 48.4% (Fig. 3e, f). Of 67 RH×AD2 F_2 individuals, AD2-type/AD2-type plants exhibited drastic reductions in capsaicinoid contents. The median value of capsaicinoid content was 67.8 μg/gDW in the AD2-type/AD2-type, whereas that in RH-type homozygotes was > 10000 μg/gDW (Fig. 3g). The capsinoid content increased 10-fold in the AD2-type/AD2-type, thereby leading to a very high capsinoid ratio up to 96.4% (Fig. 3h, i).

Given that DY-type and LP13-type exhibited intermediate capsinoid ratios between RH-type and AD2-type, the DY-type and LP-13-type *pAMT* alleles can be regarded as leaky alleles. Moreover, in RH×DY or RH×LP13 F_2 populations, there were no significant differences in capsinoid ratio between homozygotes of RH-type and heterozygotes, suggesting that DY-type and LP13-type are recessive alleles (Fig. 3c, f). Hereafter, the two mutant *pAMT* alleles harboring intronic transposons in DY and LP13 are named *pamt*^{L1} and *pamt*^{L2}, respectively.

pamt^{L1} and *pamt*^{L2} exhibited low expression levels of *pAMT*

To elucidate the molecular mechanism underlying the stepwise reduction in capsaicinoid content in the two leaky *pAMT* alleles, we investigated gene expression levels of *pAMT*. The mRNA expression levels of five capsaicinoid biosynthesis-related genes were determined in the mature green fruit stage.

pAMT exhibited low expression levels in DY (*pamt*^{L1}) and LP13 (*pamt*^{L2}), although other genes showed similar expression levels among all accessions (Fig. 4a). Peak transcript levels in the three accessions were observed at the mature green stage (Fig. 4b). This was consistent with the pattern that capsaicinoid contents increased at the mature green stage in all accessions. RH (functional *pAMT*) exhibited the highest expression level, and the expression levels of *pamt*^{L1} and *pamt*^{L2} were 7.1% and 3.1% of that in RH. To determine whether the promoter region was associated with different transcription levels, the 5' flanking sequences of *pamt*^{L1} and *pamt*^{L2} were determined and compared with functional *pAMT*. There were no significant nucleotide polymorphisms in the promoter region among the three alleles, suggesting that the promoter regions were not involved in the reduction in transcript levels in the leaky alleles (Fig. S4).

Alternative splicing occurred in both leaky alleles, but more frequently in *pamt*^{L2} than in *pamt*^{L1}

RT-PCR was conducted to determine the full-length *pAMT* cDNA. The *actin* and full-length *Pun1* cDNA sequences were amplified from all accession tested (Fig. 5a). RT-PCR for full-length *pAMT* cDNA amplified 1.5 kb fragments in RH and DY, but a larger cDNA fragment was amplified in LP13. Fig. 5b-d show a schematic representation of the genomic structure of *pAMT* and a graphic summary of the distinct cDNA sequences derived from RT-PCR. In RH and DY, the *pAMT* cDNA sequences contained a complete ORF of 1377 bp encoding a functional pAMT (Fig. 5b, c). In LP13, non-functional *pAMT* cDNA as well as

functional *pAMT* were detected. The non-functional *pAMT* cDNAs of LP13 contained partial sequences of *Tcc* between exons 3 and 4, and the partial sequence was 403 bp in length (Fig. 5d). In addition, a second intron sequence (535 bp) was inserted in the cDNA of LP13. These insertions in the *pAMT* cDNA sequence led to frameshift mutations, resulting in non-functional pAMT.

RT-PCR indicated that alternative splicing occurred in $pamt^{L2}$, but $pamt^{L1}$ did not show any difference compared with functional *pAMT*. To understand the splicing pattern in *pAMT* in more detail, RNA-seq analysis was performed in RH, DY, and LP13. RNA-seq data were obtained from placental tissue, and the reads were mapped to each *pAMT* allele. RH (functional *pAMT*) exhibited the highest FPKM value, and those of $pamt^{L1}$ and $pamt^{L2}$ were 7.3% and 2.7% of that in RH. The number of reads mapped to each location of the reference sequence were determined. The read coverage map in functional *pAMT* derived from RH exhibited 17 peaks corresponding to exons, and the splicing efficiency was almost 100% in all introns, suggesting that *pAMT* was spliced normally (Fig. 6a, d). The coverage map in $pamt^{L1}$ and $pamt^{L2}$ also showed 17 peaks corresponding to exons, but they additionally had broad peaks in intron 3 (Fig. 6b, c). Although some reads in broad peaks may come from repetitive transposons in other genomic regions, junction reads spanning from exons to the *Tcc* transposon occupied more than 60% of total depth in the broad peaks. We calculated the splicing efficiency in each intron based on junction reads using the following formula : Splicing efficiency = (correctly-spliced reads/total reads mapped in exon-intron junction) × 100 (%) (Fig. S5 and S6; Table S3). The splicing efficiency in intron 3 decreased to 82.4% (Fig. 6d). The splicing efficiency in intron 2 and intron 3 declined to

44.4%, and 16.6%, respectively (Fig. 6d). RNA-seq analysis suggested that alternative splicing occurred in intron 3 in both leaky alleles, and non-functional splice variants could be dominant in *pamt*^{L2}.

qPCR analysis was conducted to evaluate the amounts of functional *pAMT* and non-functional splice variants. When exon 16 specific primers were used to evaluate the total amount of *pAMT* transcripts, both leaky alleles showed similar reductions in the total amounts of *pAMT* transcripts (Fig. 6e). As for functional *pAMT* transcripts, *pamt*^{L2} exhibited much lower expression levels (Fig. 6f). In spite of the lower expression of functional transcripts, *pamt*^{L2} exhibited higher expression levels of non-functional variants, compared with *pamt*^{L1} (Fig. 6g). This supported the notion that the non-functional splice variants are minor in *pamt*^{L1} but dominant in *pamt*^{L2}.

Discussion

Two leaky *pamt* alleles have different effects on pungency

Various loss-of-function *pAMT* alleles have been reported in *C. chinense,* but mutations in these alleles are commonly defined by *Tcc* transposon insertions or footprints (Tanaka et al., 2010, 2015, 2018). Thus we expected that unknown mutant *pamt* alleles still exist in *C. chinense*, and we conducted screening of *C. chinense* accessions. In the process of screening, we found two pungent accessions (DY and LP13), which exhibited a unique capsaicinoid/capsinoid composition (Fig. 1; Table S1). *pamt*^{L1} and *pamt*^{L2} derived from DY and LP13 exhibited intermediate activities between functional-type and loss-of-function types, although *pamt*^{L2} had lower activity (Table 1; Fig. S1). Both leaky alleles

shared identical *Tcc* transposons inserted in intron 3, but it was notable that the positions of the intronic transposons differed in the two alleles (Fig. 2). Although both leaky alleles had common negative effects on capsaicinoid contents and positive effects on increased capsinoid contents, they had differing effects on the composition of capsaicinoid and its analogs (Fig. 3). The reduction in the level of capsaicinoid was reflected in an increase in capsinoid content and an elevated capsinoid ratio (Fig. 3).

In total, nine mutant *pamt* alleles have been found previously in the *Capsicum* genus, but all were loss-of-function types (Lang et al., 2009; Tanaka et al., 2010a, b, 2015, 2018; Koeda et al., 2014; Park et al., 2015). Known allelic variants of the *pAMT* were either functional or extreme loss-of-function, but intermediates between these extremes were unknown. In this study, two leaky-type alleles of *pAMT* were identified, $pamt^{L1}$ and $pamt^{L2}$ (Fig. 2). Even though wide quantitative variation in capsaicinoid content is associated with numerous QTLs (Blum et al., 2003; Ben-Chaim et al., 2006; Yarnes et al., 2012; Han et al., 2018), different *pAMT* genotypes had significant effects on capsaicinoid content in this study (Fig. 3). If breeders use a specific allele among *pAMT* allelic variants (functional *pAMT*, $pamt^{L1}$, $pamt^{L2}$, loss-of-function *pamt*) according to their demand for pungency level, they can adjust the pungency level by design, at least roughly. Our findings on allelic variants of *pAMT* could provide a useful genetic factor to decrease pungency level in a stepwise manner, which will contribute to chili pepper breeding for its associated food and pharmaceutical industries.

The different positions of intronic transposons affected alternative splicing, resulting in different pAMT activity

The difference in gene structure between the two leaky alleles was the location of the *Tcc* transposon insertion (Fig. 2). Therefore, different positions of the *Tcc* intronic transposon seemed to be involved in different pAMT activities. In order to reveal the mechanism underlying the different pAMT activities between *pamt*^{L1} and *pamt*^{L2}, we conducted RNA-seq analysis. The obtained results indicated that the transposon insertion in intron 3 had an influence on the splicing of the *pAMT* transcript, resulting in simultaneous expression of the functional transcript and a non-functional splice variant in both leaky alleles (Fig. 6). Abnormal splicing resulted in low expression levels in both alleles compared with the functional type allele (Fig. 6f). This low expression level in the leaky alleles could be due to elimination of the non-functional variant by a nonsense-mediated mRNA decay (NMD) mechanism. NMD is a surveillance pathway that removes nonsense mRNAs to prevent toxic effects from truncated proteins (Conti and Izaurralde, 2005).

It was notable that splicing efficiency was different in intron 3 between *pamt*^{L1} and *pamt*^{L2} (Fig. 6d). Whereas splicing efficiency in intron 3 decreased up to 82.4% in *pamt*^{L1}, it declined to 16.6% in *pamt*^{L2}. In *pamt*^{L2}, splicing efficiency in intron 2 also decreased to 44.4%. This suggested that non-functional splice variants were more dominant in *pamt*^{L2} compared with in *pamt*^{L1}. This was consistent with transcripts detected in RT-PCR (Fig. 5). Although a non-functional splice variant was not detected in *pamt*^{L1} using RT-PCR, RNA-seq and qRT-PCR analyses supported that non-functional splice variants existed as

minor molecules (Fig. 6g). In order to capture nonfunctional splice variant in DY, RT-PCR analysis was conducted with primers in exon 3 and exon 4. The RT-PCR analysis detected non-functional variants as a minor band in DY (Fig. S7). Taken together, comparative analysis between *pamt*^{L1} and *pamt*^{L2} suggested that the different positions of the transposon altered alternative splicing efficiency in intron 3, resulting in different pAMT activity levels.

It has been reported that intronic transposons can alter splicing pattern in various plants (Varagona et al., 1992; Lisch, 2013; Wei and Cao, 2016). In soybean (*Glycine max*), a CACTA family transposon (Tgm-Express1) in intron 2 of flavanone 3-hydroxylase gene causes multiple chimeric transcripts containing the transposon sequence, resulting in lighter-colored seed and flowers (Zabala and Vodkin, 2007). Genome-wide transcriptome analysis of *Gossypium raimondii* suggested that intronic transposons play important roles in intron retention during alternative splicing (Li et al., 2014). Our previous study also showed that a *Tcc* transposon in intron 3 led to abnormal splicing in *pamt*^{5} of AD2 (Tanaka et al. 2010b ; Fig. S8). However, it was difficult to evaluate the effect of intronic transposons on pAMT activity, because *pamt*^{5} also has a short insertion in exon 6 leading to a frame-shift mutation and loss-of-function (Fig. 2d). In this study, we identified *pamt*^{L1} and *pamt*^{L2}. Both alleles encoded functional pAMT proteins and had *Tcc* transposons in intron 3 (Fig. 2 b, c). The finding of *pamt*^{L1} and *pamt*^{L2} gave us opportunity to investigate how the intronic *Tcc* transposons influence pAMT activity. In this study, we demonstrated that intronic *Tcc* transposons decrease the expression level of *pAMT* but leave some functional *pAMT* mRNAs, resulting in the leaky activity of pAMT (Table 1; Fig. 6)

Differences in pAMT activity between the two leaky alleles can be explained by different splicing efficiencies, which may be determined by the approximately 200 bp difference in transposon insertion position. Although it has been reported that intronic transposon affects splicing and phenotype in various plants, little attention has been given to the effect of different insertion positions on splicing pattern. Comparison between the two leaky *pamt* alleles provided a striking case to examine how the different positions of an intronic transposon can change the splicing efficiency, leading to phenotypic variation.

Tcc transposon contains splicing signals

In both leaky alleles, non-functional splice variants commonly contained 403 bp insertions between exons 3 and 4 (Fig. 5). The sequence of the insertion was identical to part of the *Tcc* transposon. The 403 bp sequence was also observed in a splice variant derived from *pamt*^{5}, which also has an intronic *Tcc* in intron 3 (Tanaka et al., 2011b). Conserved signals for splicing in higher plants have been proposed (Lorkovic et al., 2000). Sequence analysis of *Tcc* suggested that it contains splicing signal sequences (Fig. 7; Fig. S9). A branch point (CTCAT) and 3' splicing site (TGCATAG) exist in the 5' upstream region of the 403 bp sequence, and a 5' splicing site (GTAAGT) is located in the 3' downstream region (Fig. 7). These signal sequences in *Tcc* work together with the original signal sequences of intron 3 of *pAMT*, which leads to the excision of 403 bp region of *Tcc*. Therefore, the *Tcc* transposon acts as a new splice site, thereby interfering with normal splicing and decreasing the amount of functional mRNA.

The *Tcc* transposon contributes to allelic variation in *pAMT* and produces a spectrum of pungency in the domestication process of *C. chinense*

Crop domestication has led to changes in genomes, resulting in drastic changes in plant traits (Gross and Olsen, 2010). In cereal crops and tomato, molecular genetic research has uncovered genetic changes involved in advantageous traits for human use, including grain and fruit size, seed dormancy, plant structure, and starch composition (Frary et al., 2000; Muños et al., 2011; Chakrabarti et al., 2013; Lenser et al., 2013; Lin et al., 2014). However, genetic changes involved in domestication related to taste and flavor are still poorly understood.

As for chili pepper, humans have selected individuals with good pungency levels according to their preference and culture. As a result, a number of accessions with various levels of pungency have been developed around the world. Therefore, genes involving capsaicinoid production are a good case study for domestication and crop evolution concerning taste. Our previous studies showed that low-pungency in *C. chinense* is mainly caused by loss-of-function *pamt* alleles (Tanaka et al., 2015, 2018). Two leaky *pamt* alleles have been identified in this study. These leaky alleles provide evidence of how natural variation in *pAMT* had contributed not only to low pungency but also to quantitative differences in pungency level in *C. chinense*. It is notable that all mutant *pAMT* alleles could occur through the insertion of *Tcc* or its footprints (Fig. S10). It is an attractive idea that mobilization of *Tcc* transposon family members has contributed to the wide spectrum of pungency in *C. chinense* accessions via allelic variation of *pAMT*. There are examples where transposon

activity has become a primary source of allelic variation of a single gene (Vitte et al., 2014). The most striking example comes from foxtail millet (*Setaria italica*), where the amylose content is determined by multiple weak or null alleles of the granule-bound starch synthase (GBSS1) gene, and these allelic variations were caused by several types of transposon insertions (Kawase et al., 2005). Our study on *pAMT* will be an interesting example where allelic variation has been caused by common *hAT* family transposon, resulting in both leaky-type and loss-of-function alleles.

Phylogenic analysis of *pAMT* suggested that the mutational event involved in *Tcc* insertions has occurred multiple times in a specific phylogenic cluster (Fig. S11). This led to an idea that the *Tcc* family was active in the cluster, and humans have unconsciously utilized it to create allelic variation in *pAMT*. This pattern of *pAMT* mutation was different from that in *C. annuum,* the most popular domesticated species of *Capsicum*, where *pAMT* mutations affecting capsaicinoid production have occurred only few times, but transposable elements were not associated with these mutations in *C. annuum* (Lang et al., 2009; Tanaka et al., 2010a). The *Tcc* transposon seems to be involved in genetic mutations other than *pAMT* in *C. chinense*. Very recently, it has been reported that a similar *hAT* family transposon was inserted into ketoacyl-ACP reductase (*CaKR1*), which stops capsaicinoid production in a non-pungent *C. chinense* accession (Koeda et al., 2019). It will be interesting to check whether *Tcc* has contributed to other agricultural traits in *C. chinense*.

This study provides an example where allelic variation characterized by different positions of an intronic transposon have contributed to quantitative

differences in capsaicinoid content via abnormal splicing. The two leaky alleles will be useful to decrease pungency level in a stepwise manner. It is open question why *Tcc* transposon is not related to allelic variation in *pAMT* in other *Capsicum*. It also remains unknown whether the transposon has contributed to phenotypes other than pungency. Further surveys of *pAMT* and genome-wide analysis of *Tcc* transposon will provide clearer insights into the role of transposons in phenotypic variation in *C. chinense*.

EXPERIMENTAL PROCEDURES

Plant material

Four chili pepper accessions were used in this study: 'Red Habanero' (RH), 'Devil's Yellow' (DY), LP13, and 'Aji Dulce strain2' (AD2). All accessions belong to *C. chinense.* Three F_2 populations derived from intraspecific crossing with RH as the female parent and with the different *pAMT* mutant accessions (DY, LP13, and AD2) as the male parent were used to investigate the effect of *pAMT* genotype on capsaicinoid content. In addition to the four accessions above, one pungent accession (CAP1035_T1) was used to check the association between SNPs at the *pAMT* locus with the capsinoid ratio. All plants were grown at the experimental farm of Okayama University (lat. 34 °68′ N, long. 133° 91′ E) in 2016 and 2017. Three seedlings per accession were transplanted to the field in May and June. For the F_2 population, 122 individuals of RH×DY F_2, 117 individuals of RH×LP13 F_2 and 64 individuals of RH ×AD2 F_2 were transplanted to the field.

High-performance liquid chromatography analysis for capsaicinoid and capsinoid contents

Three fruits were collected at different developmental stages for each accession. As for F_2 individuals, three fruits at the mature green stage were used to measure the capsaicinoid composition. Samples were lyophilized using a freeze drier for 3 days and then ground in a blender. Capsaicinoids and capsinoids were extracted from 200 mg of dry fruit powder using 4 mL acetone. The supernatant was transferred to a 1.5-mL tube through a filter (DISMIC 13HP, ADVANTEC) and then used for high-performance liquid chromatography (HPLC) analysis. The capsaicinoid composition was determined in accordance with the HPLC method of Tanaka et al. (2015). The capsaicinoids and capsinoids were separated on a Inertsil ODS-3 column (250 mm × 4.6 mm i.d. GL Science) with a mixture of methanol/water (75:25) at a flow rate of 1.0 mL/min, and detected using a UV-Vis detector at 280 nm. The total capsaicinoid content was calculated as the sum of capsaicin and dihydrocapsaicin, whereas the total capsinoid content was considered as the sum of capsiate and dihydrocapsiate Data for each fruit part and developmental stage were averaged using three biological replicates.

Assay for pAMT enzymatic activity

The placenta was excised from the fruits of pepper and homogenized with 50 mM potassium phosphate buffer, pH 6.8 (PKB) (0.1 g placenta/1 mL PKB). The homogenate was centrifugated at 2000 g for 2 min to produce a supernatant. One hundred microliters of the supernatant was added into 200 μL of a PKB solution

of vanillin, 4-aminobutyric acid (GABA), pyridoxal-5-phosphate (PLP), and methyl vanillate as an internal standard (final concentrations: 4 mM, 16 mM, 1 mM, and 40 μM, respectively). The mixture was incubated at 30°C. An aliquot (10 μL) of the mixture was added into 30 μL of ethanol to terminate the reaction at 0 and 24 hours. The vanillylamine content in the ethanol solution was measured using an HPLC system (Shimadzu Prominence LC-20A series, Shimadzu Co., Kyoto, Japan) with the following specifications: column, Wakopak® Fluofix-II 120E, 250 mm × 4.6 mm i.d. (Wako Pure Chem. Ind. Ltd., Osaka, Japan); solvent, A, 0.05% trifluoroacetic acid, B, methanol, 10–15% B (0–10 min), 15–40% B (10–20 min), 40–100% B (20–30 min), 100% B (30–35 min); flow rate, 1 mL/min; fluorescence detection, Ex 280 nm and Em 320 nm. The increase in vanillylamine content from 0 to 24 hours per gram of placenta weight was indicated as pAMT activity. The vanillyl alcohol content measured using the above-mentioned HPLC conditions and the vanillylamine content at 0 hour were considered as their intact contents in the placentas.

Genomic sequence analysis of *pAMT* alleles

The genomic sequences covering the full-length of *pAMT* were determined for DY and LP13 in accordance with the method presented by Tanaka et al. (2015). Genomic DNA was isolated from leaf tissues using Nucleon PhytoPure (GE Healthcare Japan, Tokyo, Japan). Genomic PCR was performed using KOD FX neo (TOYOBO) with the primer sets described in Table S2. The PCR products were purified using an Exo Star kit (GE Healthcare Japan) and nucleotide sequencing was performed by the Eurofins sequencing service (Eurofins

Genomics, Tokyo, Japan). ATGC (GENETYX, Tokyo, Japan) was applied to analyze nucleotide sequences.

The different locations of intronic *Tcc* transposon were confirmed by genomic PCR (Fig. S3). The primers were designed based on the position of the *Tcc* transposon in each mutant *pAMT* allele, and primer sequences are shown in Table S2.

RNA extraction

Fruits at different developmental stages were harvested, and the placental septum was separated for RNA extraction. Total RNA was extracted using Sepasol®-RNA I Super G (Nacalai, Japan), and purified using an RNeasy Mini spin column (Qiagen). RNA concentrations were measured using a spectrophotometer. The integrity of isolated RNA was checked based on ribosomal RNA bands separated using agarose gel electrophoresis. All RNAs used for RT-PCR were treated with DNase I prior to cDNA synthesis in order to remove contaminating DNA. RNA extraction was conducted with three biological repeats.

Quantitative reverse transcription polymerase chain reaction

A quantitative reverse transcription polymerase chain reaction (qRT-PCR) analysis was performed to determine the expression levels of structural genes related to the capsaicinoid biosynthetic pathway. Synthesis of first strand cDNA was performed by reverse transcription of 500 ng of DNA-free RNA using PrimeScript™ RT Master Mix (Takara). Quantitative PCR was performed using

KOD SYBR® qRT-PCR Mix (Toyobo) with a Roche Light Cycler Nano Real-Time PCR System in accordance with the manufacturer's instructions.

The PCR conditions were as follows: initial denaturation at 98°C for 2 min, followed by 45 cycles of 10 s at 98°C, 10 s at 60°C, and 30 s at 68°C. *Actin* was used as an internal control gene for normalization. The comparative CT method (2-ΔΔCT method) was used to determine expression levels. Primer sequences for qRT-PCR are listed in Table S2.

Transcript analysis by RT-PCR and RNA-seq and evaluation of splicing efficiency in *pAMT* alleles

The full-length *pAMT* cDNA sequence was amplified by RT-PCR. The RT reaction was performed using ReverTra Ace (TOYOBO, Japan). RT-PCR was performed using KOD-FX neo with primer sets (pAMT F1 and pAMT R1481). The PCR procedure was as follows: 1 cycle of 2 min at 96°C; 35 cycles of 10 s at 98°C, 30 s at 55°C, and 2 min at 68°C; and a final extension of 5 min at 68°C. PCR products were separated on a 1.0% agarose gel and stained with Gel Red (Biotium, Hayward, CA, USA) to check the amplicon sizes. The nucleotide sequences of the RT-PCR amplicons was determined as described above.

For RNA-seq analysis, total RNA samples from three independent extractions with a total six mature green fruits were pooled. After selection for mRNA, a strand-specific mRNA library was constructed. Pair-end (2 × 125 bp) sequencing were performed using a Hiseq 2500 (Illumina, San Diego, CA, USA). The raw reads were cleaned using flex-bar to remove adaptor sequences and low quality reads. Clean reads were mapped onto complete genomic sequences of each *pAMT*

allele by using TopHat v. 2.0.6 and Bowtie v.2.0.5. The resulting bam files were sorted and indexed using samtools. The read depth at each nucleotide position was determined from the bam file using IGVtools count function. Read depth map for each *pAMT* allele was created after calculating RPM (reads per million) at each nucleotide position.

The splicing efficiency in each intron of *pAMT* was also determined from the resulting sam file. Reads lying in the junction between an exon and an intron (total junction reads) were collected from the sam file using customized perl program. Then we counted correctly spliced reads from the total junction reads. When a read is split with the exact length of the intron, the read was considered as a correctly spliced read (Fig. S5). Finally, splicing efficiency was calculated using following formula. Splicing efficiency = (correctly-spliced reads / total junction reads) × 100 (%).

Data availability statements

The complete genomic sequences of two leaky *pamt* alleles (*pamt*^{L1} and *pamt*^{L2}) have been deposited in GenBank (accession no. LC483648 and LC483649). The dataset of RNA-seq reads has been submitted to the DNA Data Bank of Japan (DDBJ) Sequence Read Archive under the accession number DRA008429.

Author contribution statement

YT designed the research, wrote the manuscript, and conducted the experiments on sequencing, gene expression, and RNA-seq. TA and YK cultivated plant materials and performed HPLC analysis of capsaicinoids. TG, YY, and KY assisted with cultivation and data analysis. YM, SN, and KK conducted HPLC analyses for precursors and enzyme assays. All authors read and approved the final manuscript.

Acknowledgment

The authors thank Sho Oono (Kyoto University) for his valuable advice. We would also like to thank Yuji Muraga and Shiho Fukuta (Okayama University) for their assistance with experiments. This study was supported in part by a Grant-in-Aid for Scientific Research (C) (18K05618), Japan, and The Sanyo Broadcasting Foundation.

Conflict of interest

The authors declare that they have no conflict of interest.

Supporting Information

Fig. S1. Loss-of-function of pAMT causes capsinoid biosynthesis

Fig. S2. Amino acid alignment of pAMT

Fig. S3. Genomic PCR analysis to distinguish different positions of *Tcc* insertion

Fig. S4. Nucleotide alignment in the promoter region of *pAMT*

Fig. S5. Method to calculate splicing efficiency based on RNA-seq data

Fig. S6. Mapping situations of RNA-seq reads in the intron 3 region

Fig. S7. The RT-PCR analysis to capture exon3- exon4 region of *pAMT* cDNA sequence.

Fig. S8. RNA-seq coverage plot in loss-of-function *pamt*^{5} derived from AD2

Fig. S9. Sequence analysis from exons 3 to 4 in functional *pAMT* and leaky *pamt* alleles

Fig. S10. Structural variation of loss-of-function *pamt* alleles

Fig. S11. Phylogenetic tree of *pAMT* alleles

Table S1. Capsaicinoid and capsinoid content including additional *Capsicum* accessions.

Table S2. Primers for genomic PCR, RT-PCR, and qRT-PCR in this study.

Table S3. The number of RNA-seq reads mapping to the pAMT locus and evaluation of splicing efficiency.

REFERENCES

Arce-Rodriguez M. L., Ochoa-Alejo, N. (2017) An R2R3-MYB transcription factor regulates capsaicinoid biosynthesis. ***Plant Physiol.*** **174**, 1359-1370.

Aza-Gonzalez, C., Nunez-Palenius, H.G. and Ochoa-Alejo, N. (2011) Molecular biology of capsaicinoid biosynthesis in chili pepper (*Capsicum* spp.). *Plant Cell Rep.* **30**, 695–706.

Ben-Chaim, A., Borovsky, Y., Falise, M., Mazourek, M., Kang, B.C., Paran, I. and Jahn, M. (2006) QTL analysis for capsaicinoid content in *Capsicum*. *Theor. Appl. Genet.* **113**, 1481–1490.

Blum, E., Mazourek, M., O'Connell, M., Curry, J., Thorup, T., Liu, K.D., Jahn, M., and Paran, I. (2003) Molecular mapping of capsaicinoid biosynthesis genes and quantitative trait loci analysis for capsaicinoid content in *Capsicum. Theor. Appl. Genet.* **108**, 79–86.

Bosland, P.W., and Votava, E.J. (2000) Peppers:vegetable and spice capsicums CABI Publishing, New York

Chakrabarti, M., Zhang, N., Sauvage, C. et al. (2013) A cytochrome P450 regulates a domestication trait in cultivated tomato. *Proc. Natl Acad. Sci. USA*, **110**, 17125–17130.

Conti, E. and Izaurralde, E. (2005) Nonsense-mediated mRNA decay: molecular insights and mechanistic variations across species. *Curr. Opin. Cell Biol.* **17**, 316–325.

Curry, J., Aluru, M., Mendoza, M., Nevarez, J., Melendrez, M. and O'Connell, M.A. (1999) Transcripts for possible capsaicinoid biosynthetic genes are differentially accumulated in pungent and non-pungent *Capsicum* spp. *Plant Sci.* **148**, 47–57.

Frary, A., Nesbitt, T.C., Grandillo, S., Knaap, E., Cong, B., Liu, J., Meller, J., Elber, R., Alpert, K.B. and Tanksley, S.D. (2000) fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. *Science*, **289**, 85–88.

Gross, B.L. and Olsen, K.M. (2010) Genetic perspectives on crop domestication. *Trends Plant Sci.* **15**, 529–537.

Han, K., Lee, H.-Y., Ro, N.-Y., Hur, O.-S., Lee, J.-H., Kwon, J.-K. and Kang, B.C. (2018) QTL mapping and GWAS reveal candidate genes controlling capsaicinoid content in *Capsicum*. *Plant Biotechnol. J.* **16**, 1546–1558.

Han, K., Jang, S., Lee, J. -H., Lee, D.-G., Kwon, J.-K. and Kang, B.C. (2019) A MYB transcription factor is a candidate to control pungency in *Capsicum annuum*. *Theor. Appl. Genet.* 132, 1235–1246.

Jang, S., Koeun, H., Jo, Y. D., Jeong, H.-J., Siddique, M. I. and Kang, B. C. (2015) Substitution of a dysfunctional pAMT allele results in low-pungency but high levels of capsinoid in *Capsicum chinense* 'Habanero'. *Plant Breed. Biotech.* **3**, 119–128.

Jeong, H.S., Jang, S., Han, K., Kwon, J.K. and Kang, B.C. (2015) Marker-assisted backcross breeding for development of pepper varieties (*Capsicum annuum*) containing capsinoids. *Mol. Breed.* **35**, 226.

Kawase, M., Fukunaga, K. and Kato, K. (2005). Diverse origins of waxy foxtail millet crops in East and Southeast Asia mediated by multiple transposable element insertions. *Mol. Genet. Genomics* **274**, 131–140.

Kirii, E., Goto, T., Yoshida, Y., Yasuba, K. and Tanaka, Y. (2017). Non-pungency in a Japanese chili pepper landrace (*Capsicum annuum*) is caused by a novel loss-of-function *Pun1* allele. *Hort. J.* **86**, 61–69.

Koeda, S., Sato, K., Tomi, K., Tanaka, Y., Takisawa, R., Hosokawa, M., Doi, M., Nakazaki, T. and Kitajima, A. (2014) Analysis of non-pungency, aroma, and origin of a *Capsicum chinense* cultivar from a Caribbean island. *J. Japan Soc. Hort. Sci.* **83**, 244-251.

Koeda, S., Sato, K., Saito, H., Nagano, A. J., Yasugi, M., Kudoh, H., and Tanaka, Y. (2019) Mutation in the putative ketoacyl-ACP reductase CaKR1induces loss of pungency in Capsicum. *Theor. Appl. Genet.* **132**, 65-80.

Kobata, K., Todo, T., Yazawa, S., Iwai, K. and Watanabe, T. (1998) Novel capsaicinoid-like substances, capsiate and dihydrocapsiate, from the fruits of a nonpungent cultivar, CH-19 sweet, of pepper (*Capsicum annuum* L.). *J. Agric. Food Chem.* **46**, 1695–1697.

Lang, Y.Q., Kisaka, H., Sugiyama, R., Nomura, K., Morita, A., Watanabe, T., Tanaka, Y., Yazawa, S. and Miwa, T. (2009) Functional loss of pamt results in biosynthesis of capsinoids, capsaicinoid analogs, in *Capsicum annuum* cv. CH-19 Sweet. *Plant J.* **59**, 953–961.

Lenser, T. and Theissen, G. (2013). Molecular mechanisms involved in convergent crop domestication. *Trends Plant Sci.* **18**, 704–714.

Li, Q., Xiao, G. and Zhu, Y.X. (2014) Single-nucleotide resolution map- ping of the *Gossypium raimondii* transcriptome reveals a new mechanism for alternative splicing of introns. *Mol. Plant* **7**, 829–840.

Lin, T., Zhu, G.T., Zhang, J.H. et al. (2014) Genomic analyses provideinsights into the history of tomato breeding. *Nat. Genet.* **46**, 1220–1226

Lisch, D. (2013) How important are transposons for plant evolution? *Nat. Rev. Genet.* **14**, 49-61.

Lorkovic, Z.J., Kirk, D.A.W., Lambermon, M.H.L. and Filipowicz, W. (2000) Pre-mRNA splicing in higher plants. *Trends Plant Sci.* **5**,160–167.

Luo, X.J., Peng, J. and Li, Y.J. (2011) Recent advances in the study on capsaicinoids and capsinoids. *Eur. J. Pharmacol.* **650**, 1–7.

Mazourek, M., Pujar, A., Borovsky, Y., Paran, I., Mueller, L. and Jahn, M.M. (2009) A dynamic interface for capsaicinoid systems biology. *Plant Physiol.* **150**, 1806–1821.

Moses, M., Umaharan, P. and Dayanandan, S. (2014) Microsatellite based

analysis of the genetic structure and diversity of *Capsicum chinense* in the Neotropics. *Genet. Resour. Crop Evol.* **61**, 741-755

Muños, S., Ranc, N., Botton, E., et al. (2011) Increase in tomato locule number is controlled by two single-nucleotide polymorphisms located near WUSCHEL. *Plant Physiol.*, **156**, 2244–2254.

Park, Y.-J., Nishikawa, T., Minami, M., Iwasaki, T. and Matsushima, K. (2015) A low-pungency S3212 genotype of Capsicum frutescens caused by a mutation in the putative aminotransferase (p-AMT) gene. *Mol. Genet. Genomics* **290**, 2217–2224.

Stellari, G.M., Mazourek, M. and Jahn, M.M. (2010) Contrasting modes for loss of pungency between cultivated and wild species of *Capsicum*. *Heredity* **104**, 460–471

Stewart, C., Kang, B.C., Liu, K., Mazourek, M., Moore, S.L., Yoo, E.Y., Kim, B.D., Paran, I. and Jahn, M.M. (2005) The Pun1 gene for pungency in pepper encodes a putative acyltransferase. *Plant J.* **42**, 675–688.

Stewart, C., Mazourek, M., Stellari, G.M., O'Connell, M. and Jahn, M. (2007) Genetic control of pungency in *C. chinense* via the *Pun1* locus. *J. Exp. Bot.* **58**, 979–991.

Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T. and Yazawa, S. (2010a) Newly mutated putative-aminotransferase in nonpungent pepper (*Capsicum annuum*) results in biosynthesis of capsinoids, capsaicinoid analogues. *J. Agric. Food Chem.* **58**, 1761–1767.

Tanaka, Y., Hosokawa, M., Miwa, T., Watanabe, T. and Yazawa, S. (2010b) Novel loss-of-function putative aminotransferase alleles cause biosynthesis of capsinoids, non-pungent capsaicinoid analogues, in mildly pungent chili peppers (*Capsicum chinense*). *J. Agric. Food Chem.* **58**, 11762–11767.

Tanaka, Y., Sonoyama, T., Muraga, Y., Koeda, S., Goto, T., Yoshida, Y. and Yasuba, K. (2015) Multiple loss-of-function putative aminotransferase alleles contribute to low pungency and capsinoid biosynthesis in *Capsicum chinense*. *Mol. Breed.* **35**, 142.

Tanaka, Y., Fukuta, S., Koeda, S., Goto, T., Yoshida, Y., Yasuba, K. (2018) Identification of a novel mutant pAMT allele responsible for low-pungency and capsinoid production in chili pepper: accession 'No. 4034' (*Capsicum chinense*) *Hort. J.* **87**, 222–228.

Varagona, M. J., Purugganan, M., Wessler, S.R. 1992. Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* **4**, 811-820.

Vitte, C., Fustier, M.A., Alix, K. and Tenaillon, M.I. (2014) The bright side of transposons in crop evolution. *Brief. Funct. Genomics*, doi:10.1093/bfgp/elu002.

Wei, L. and Cao, X. (2016) The effect of transposable elements on phenotypic variation: insights from plants to humans. *Sci. China Life Sci.* **59**, 24-37.

Wink, M. (1988) Plant breeding: importance of plant secondary metabolites for protection against pathogens and herbivores. *Theor. Appl. Genet.* **75**, 225-233.

Wink, M. (2003). Evolution of secondary metabolites from an ecological and molecular phylogenetic perspective. *Phytochemistry* **64**, 3-19.

Wyatt, L.E., Eannetta, N.T., Stellari, G.M. and Mazourek, M. (2012) Development and application of a suite of non-pungency markers for the Pun1 gene in pepper (*Capsicum* spp.). Molecular Breeding **30**, 1525-1529.

Yarnes, S.C., Ashrafi, H., Reyes-Chin-Wo, S., Hill, T.A., Stoffel, K.M., Van Deynze, A. and Gulick, P. (2012) Identification of QTLs for capsaicinoids, fruit quality, and plant architecture-related traits in an interspecific *Capsicum* RIL population. *Genome* **56**, 61–74.

Yazawa, S., Suetome, N., Okamoto, K. and Namiki, T. (1989) Content of capsaicinoids and capsaicinoid-like substances in fruit of pepper (*Capsicum annuum* L.) hybrids made with "CH-19 Sweet" as a parent. *J. Japan Soc. Hort. Sci.* **58**, 601–607.

Zabala, G. and Vodkin, L. (2007) Novel exon combinations generated by alternative splicing of gene fragments mobilized by a CACTA transposon in *Glycine max*. *BMC Plant. Biol.* **7**, 38.

Table Legend

Table 1. Vanillylamine and vanillyl alcohol contents, and pAMT activity in placental tissue.

Figure Captions

Fig. 1. Capsaicinoid and capsinoid contents in *C. chinense* accessions

a) Features of chili pepper fruits in different developmental stages. Four *C. chinense* accessions (RH: Red Habanero, DY: Devil's Yellow, LP13, AD2: Aji Dulce strain2) contain different alleles of *pAMT* (functional *pAMT*, $pamt^{L1}$, $pamt^{L2}$, loss-of-function $pamt^{5}$). b) Capsaicinoid and capsinoid content. c) Capsinoid ratio in mature green stages IM: immature (approximately 10 days post anthesis [dpa]), MG: mature green (30 dpa), M: mature (40 dpa). Bar indicates standard error (n=3) Capsinoid ratio was determined using following formula : capsinoid ratio = {capsinoid content/ (capsaicinoid + capsinoid contents)} × 100 (%). Student's t-test indicated a significant difference in capsinoid ratio between RH and DY and LP13 and AD2, respectively.

Fig. 2. Comparison of the gene structures in four different *pAMT* alleles

a) RH-type (functional-type). b) DY-type ($pamt^{L1}$). c) LP13-type ($pamt^{L2}$). d) AD2-type (loss-of-function type $pamt^{5}$).

Black boxes indicate the 17 exons, and grey boxes indicates *Tcc* transposons. Three *pAMT* alleles (b–d) have *Tcc* transposons in intron 3, but their positions were different. *Tcc* transposons were located 591 bp, 342 bp, and 542 bp away from exon 4 in $pamt^{L1}$, $pamt^{L2}$, and loss-of-function type $pamt^{5}$, respectively. In the loss-of-function type $pamt^{5}$, an 8 bp insertion in exon 6 causes a frame-shift mutation.

Fig. 3. The effect of different *pAMT* allele on capsaicinoid and capsinoid compositions

RH (functional *pAMT*/ functional *pAMT*)×DY (*pamt*$^{L.1}$/ *pamt*$^{L.1}$) population (n=122) (a) capsaicinoid; (b) capsinoid; (c) capsinoid ratio. RH (functional *pAMT*/ functional *pAMT*)×LP13 (*pamt*$^{L.2}$/ *pamt*$^{L.2}$) population (n=117); (d) capsaicinoid; (e) capsinoid; (f) capsinoid ratio. RH (functional *pAMT*/functional *pAMT*)×AD2 (loss-of-function *pamt*^{5}/loss-of-function *pamt*^{5}) population (n=67) (g) capsaicinoid; (h) capsinoid; (i) capsinoid ratio. *pAMT* genotypes are represented on the horizontal axis. The vertical axis shows the contents of capsaicinoid and capsinoid, and capsinoid ratio. Capsinoid ratio = {capsinoid content/ (capsaicinoid + capsinoid contents)} × 100 (%). Data in box plots are displayed as median values (horizontal line), lower and upper quartiles (boundary box), minimum and maximum values (whiskers). Different letters indicates significant differences between different *pAMT* genotypes (Tukey's HSD test $p<0.05$).

Fig. 4. Expression levels of *pAMT* and other capsaicinoid biosynthesis-related genes

a) Expression levels of structural genes in the capsaicinoid biosynthesis pathway at the mature green stage. Three accessions (RH: Red Habanero, DY: Devil's Yellow, LP13) were used. *Pal: phenylalanine ammonia-lyase, pAMT: putative aminotransferase, Fat: Acyl-ACP thioesterase, KAS: β-ketoacyl ACP synthase, and Pun1: acyltransferase.* b) Expression levels of *pAMT* at different stages of fruit development. IM: immature (approximately 10 days post anthesis [dpa]), MG: mature green (30 dpa), M: mature (40 dpa). Total RNA was extracted from

the placental septum tissue of fruits. The expression level of genes in RH at mature green stage (MG) was set to 1.0. All data are presented as means of three replicates. Error bars shows standard errors.

Fig. 5. RT-PCR analysis for full-length *pAMT*

(a) RT-PCR analysis for *Actin*, full-length *Pun1* and full-length *pAMT*. Three accessions (RH: Red Habanero, DY: Devil's Yellow, LP13) were used. RNA was extracted from the placental septum tissue of mature green fruit. Primers for *Pun1* and *pAMT* were designed to amplify the full-length coding sequences (CDS). b–d) Schematic diagrams of the genomic structure of *pAMT* alleles, and RT-PCR-derived cDNAs in *C. chinense* accessions. Colored boxes indicate exons and horizontal lines indicate introns. *pamt*^{L1} and *pamt*^{L2} contain *Tcc* transposons in different positions in intron 3. The slanting lines indicate splicing sites in transcripts. (b) Functional *pAMT* derived from RH, (c) *pamt*^{L1} derived from DY. Functional cDNAs corresponding the full-length CDS were detected in RH and DY. (d) *pamt*^{L2} derived from LP13. In LP13, functional cDNA and non-functional splice variants containing the partial *Tcc* sequence (403 bp) between exons 3 and 4 were detected.

Fig. 6. RNA-seq analysis reveals that intronic *Tcc* transposons change the splicing efficiency in *pAMT*, resulting in non-functional splice variants

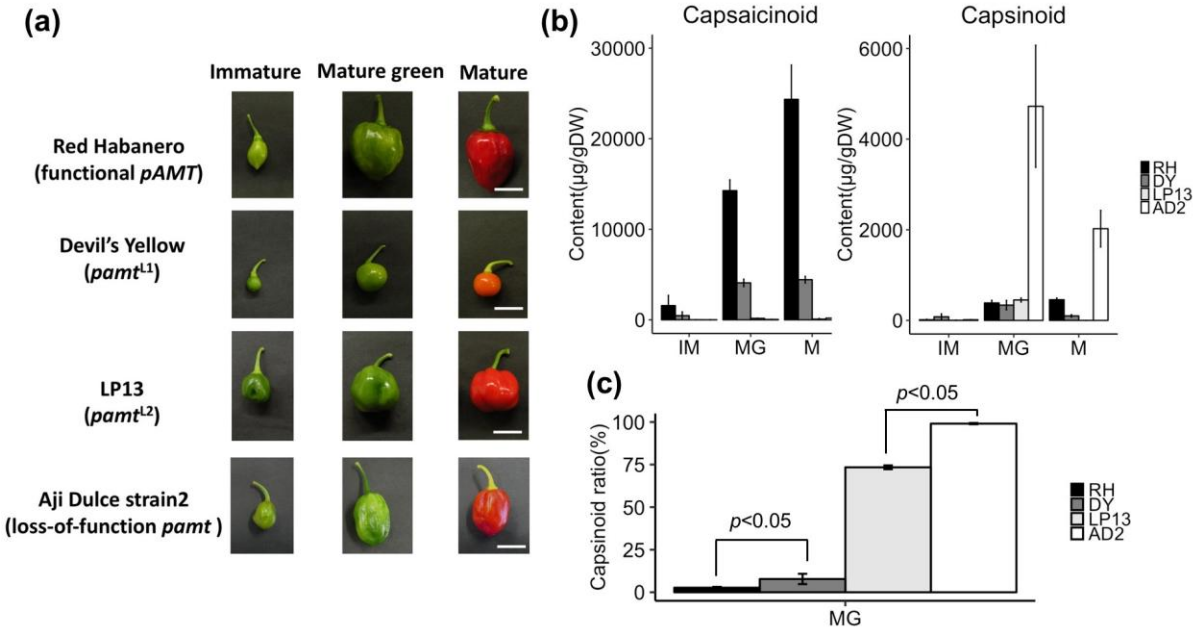
(a–c) RNA-seq coverage plot in *pAMT* alleles: (a) functional *pAMT*, (b) *pamt*^{L1}, and (c) *pamt*^{L2}. RNA-seq reads were mapped to the genomic sequence of the corresponding *pAMT* alleles using TopHat. The x axis shows the nucleotide position in the *pAMT* locus, and the y

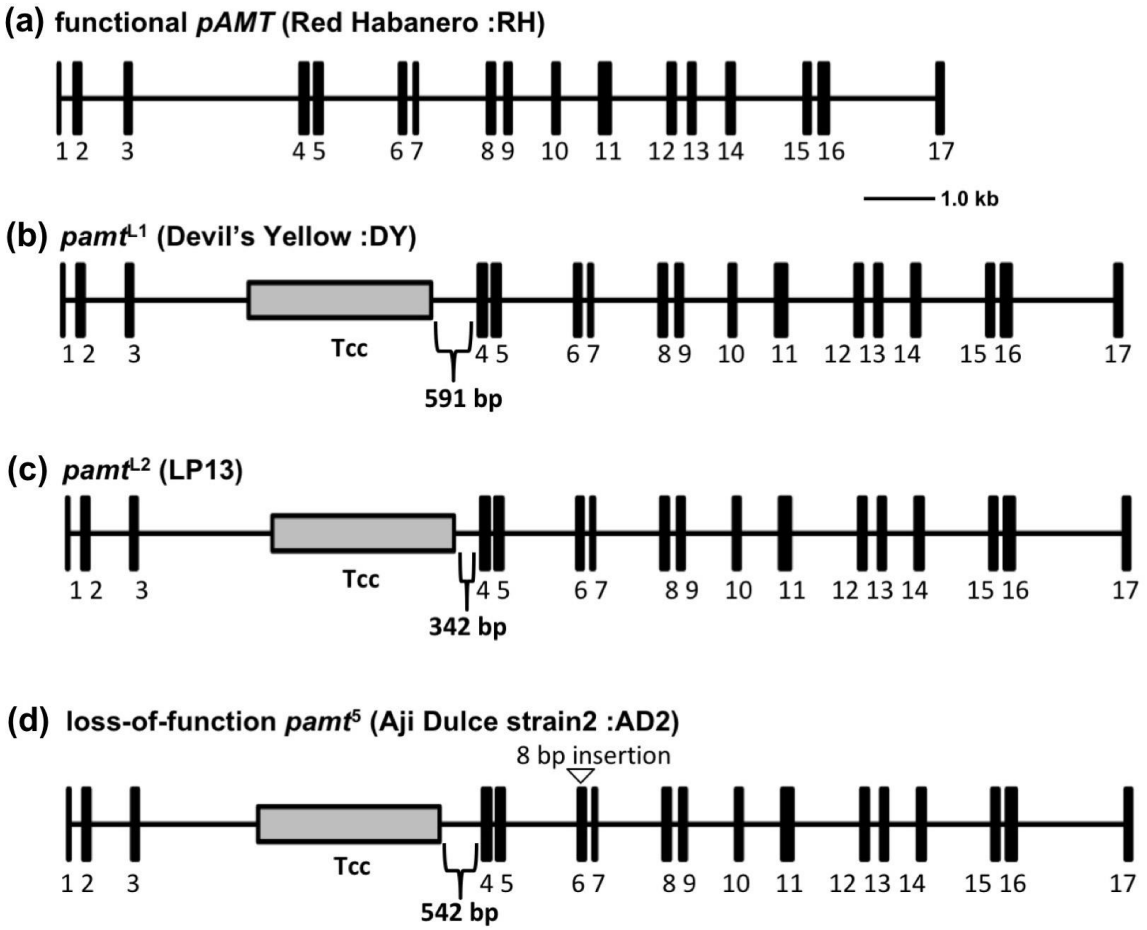
axis represents the read depth. Depth was normalized as reads per million (RPM). Reads were mapped to the 17 exons in all alleles, but overall depth was low in leaky alleles. In addition, broad transcript mountains were observed in intron 3 of the leaky alleles. (d) Splicing efficiency in each intron. Splicing efficiency = (correctly-spliced reads / total reads mapped in exon-intron junction) × 100 (%). (e–f) qRT-PCR analysis to validate splicing disruption in intron 3. (e) Quantification of total *pAMT* transcripts. Primers will bind to exon 16. (f) Specific quantification for normally spliced *pAMT* transcripts. Primers bind to the junctions of exon 3 and exon 4. (g) Detection of abnormal splice variants containing a partial *Tcc* sequence. A partial *Tcc* sequence was inserted between exons 3 and 4, and primer will bind to the insertion region. Total RNA was extracted from the placental septum tissue of mature green fruits. The expression level of genes in RH was set to 1.0. All data are presented as means of five replicates. Error bars shows standard errors.

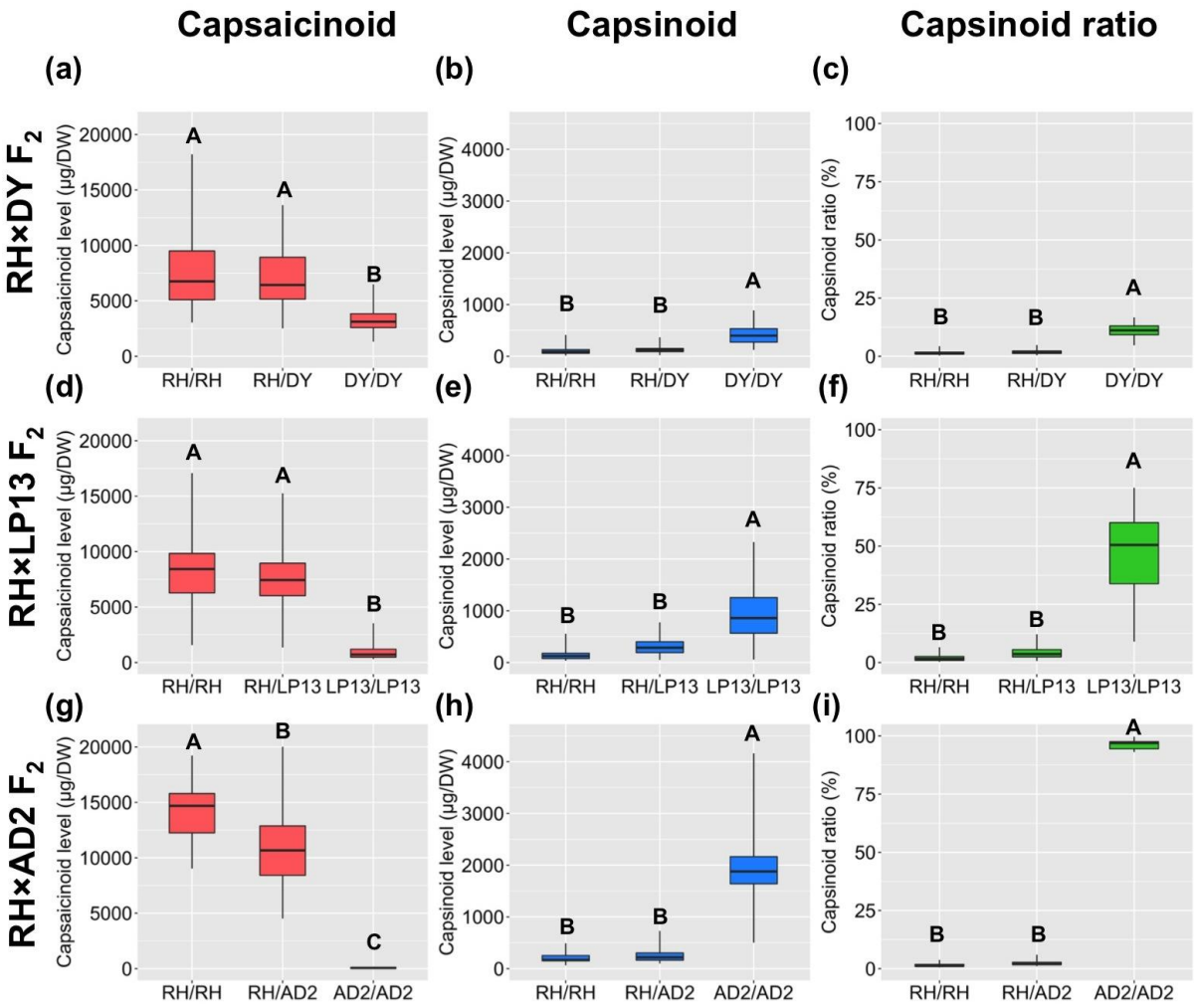
Fig. 7. The intronic transposon in *pAMT* acts as a new exon, leading to simultaneous expression of functional *pAMT* mRNA and non-functional splice variants

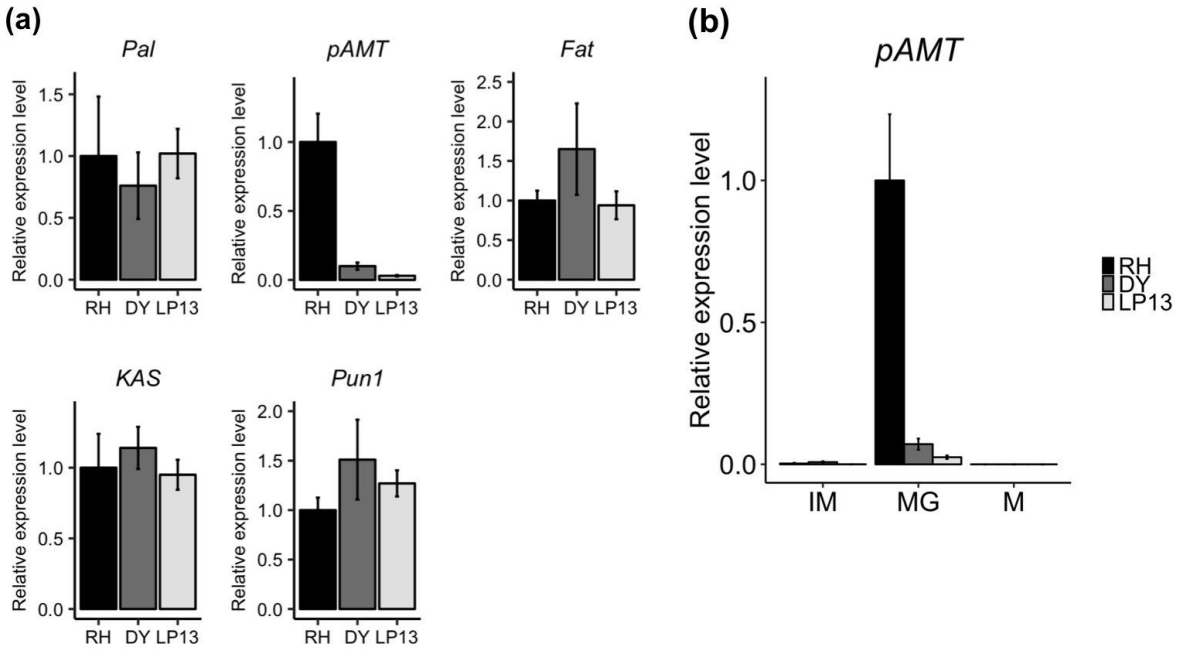
Boxes and lines indicates exons and introns, respectively. 5'-SS: 5' splicing signals. 3'-SS: 3' splicing signals. BP: branch point. a) Functional *pAMT* allele. Intron between exons 3 and 4 is normally spliced to produce functional mRNA. b) leaky *pAMT* allele. Intronic *Tcc* transposon contains splicing signals. The signals on *Tcc* work together with native splicing signals, resulting in simultaneous expression of functional *pAMT* mRNA and non-functional splice variants containing a partial sequence of *Tcc*.

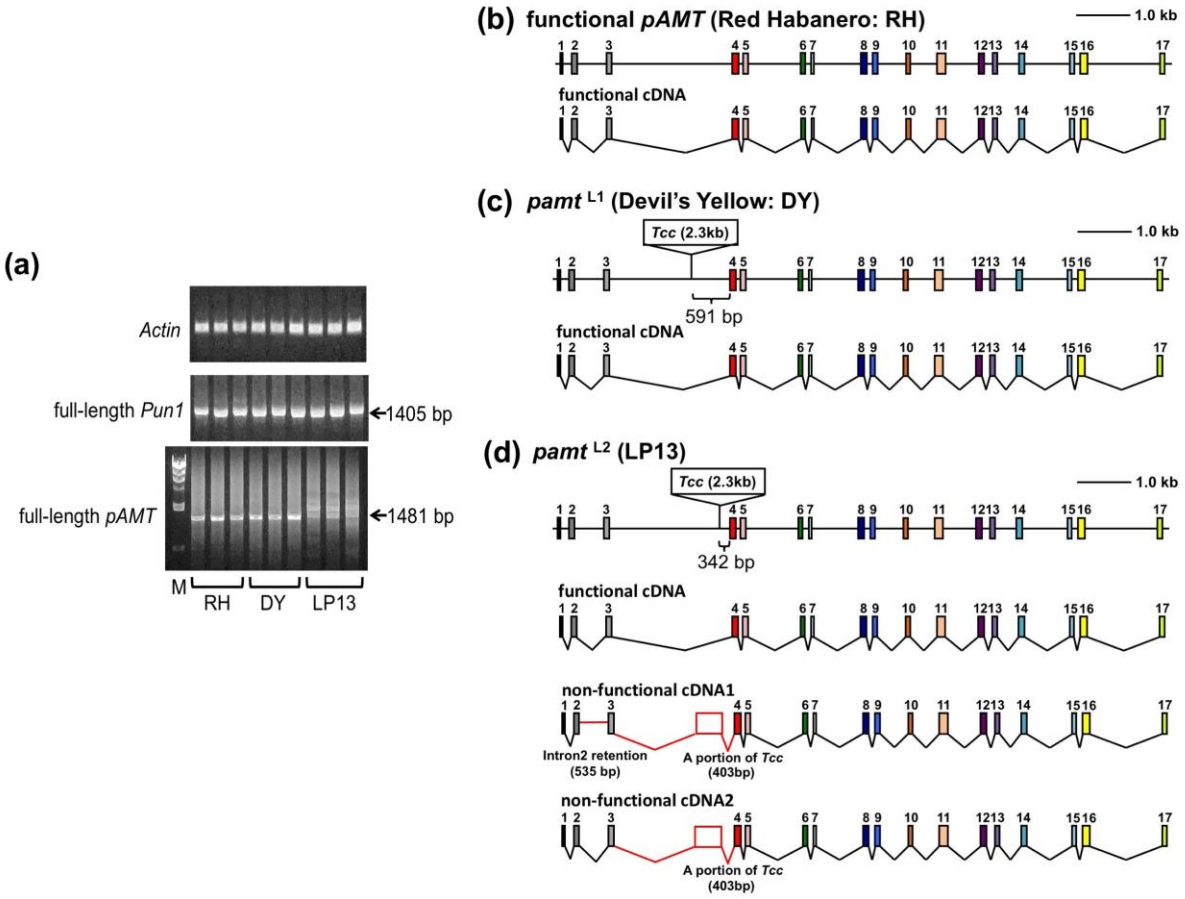
Table 1. Vanillylamine and vanillyl alcohol contents, and pAMT activity in placental tissue.					
Accession and pAMT genotypes [a]	µg/g (placenta)				pAMT enzymatic activity
	Vanillylamine		Vanillyl alcohol		
Red Habanero (functional *pAMT* /functional *pAMT*)	27.2		10.3		61.99
Devil's Yellow (*pamt*^{L1}/*l pamt*^{L1})	6.6		8.7		26.19
LP13 (*pamt*^{L2}/*pamt*^{L2})	nd.		42.4		7.59
Aji Dulce strain2 (loss-of-function *pamt*^{5}/loss-of-function *pamt*^{5})	nd.		96.1		nd.
Mean of two duplicates					
[a] *pAMT* alleles were referenced to Tanaka et al. (2015) and this study.					
[b] The ability of placental tissue to transform vanillin to vanillylamine in 24 h.					

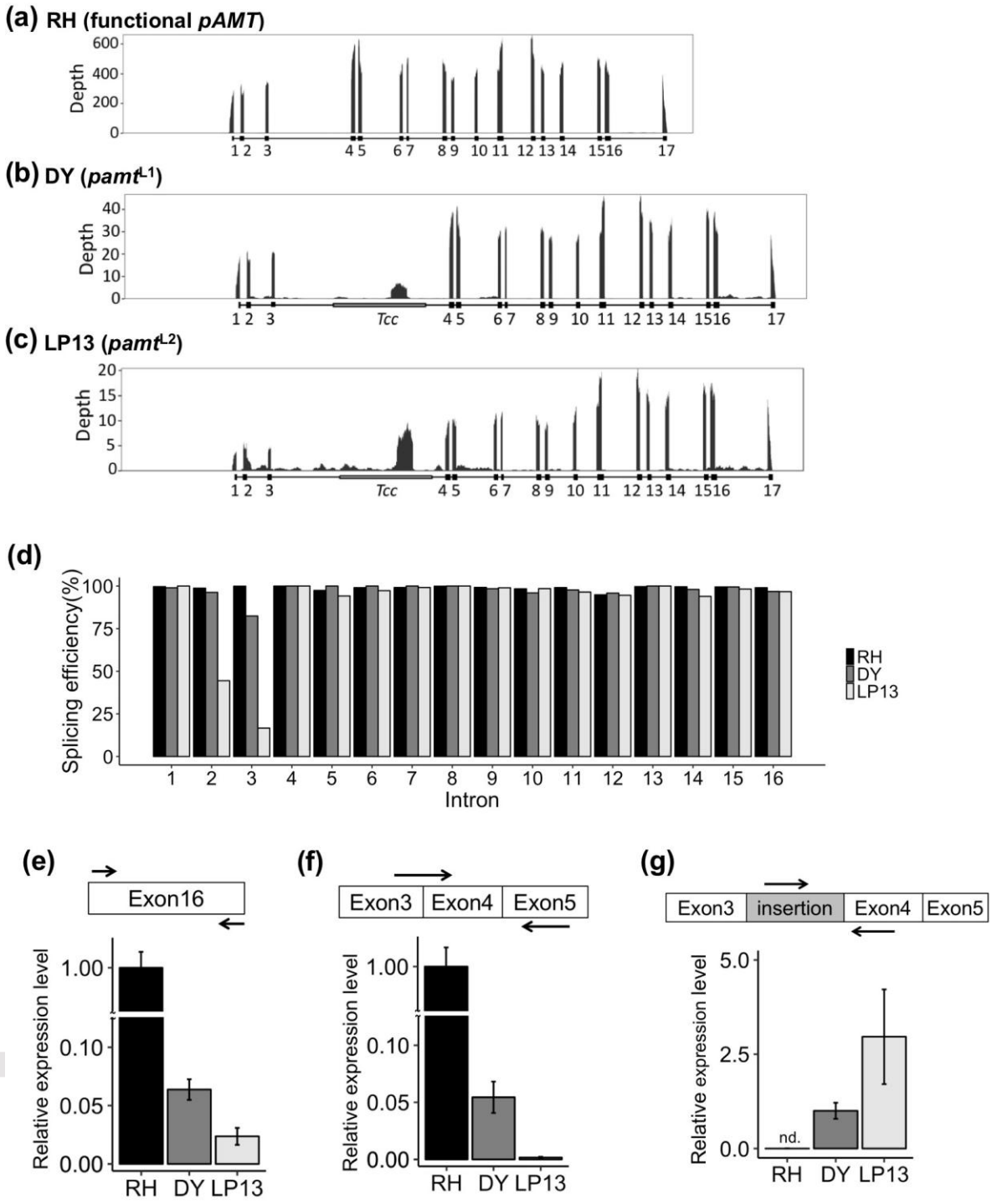




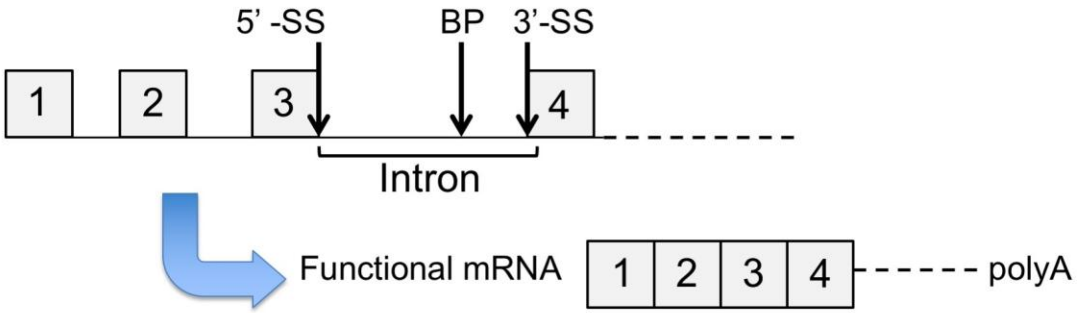








(a) Functional *pAMT* allele



(b) leaky *pAMT* allele