ORIGINAL PAPER

Pollen transcriptome analysis of *Solanum tuberosum* ($2n = 4x = 48$), *S. demissum* ($2n = 6x = 72$), and their reciprocal F$_1$ hybrids

Rena Sanetomo · Kazuyoshi Hosaka

Abstract

Key message **Pollen mRNAs were different in reciprocal F$_1$ hybrids, which were probably caused by a cytoplasm–nuclear chromosomal genes interaction.**

Abstract We have found reciprocal differences in crossability between F$_1$ hybrids of *Solanum tuberosum* (T) and a Mexican wild potato species *S. demissum* (D). When the reciprocal hybrids were crossed as pollen parents with *S. demissum,* a significantly higher berry-setting rate was obtained in TD compared with DT. In this study, we performed a whole-genome transcript analysis of the pollen mRNA using a high-throughput sequencer. We obtained 12.6 billion bases that were aligned into 13,020 transcripts with 9,366 loci. All possible genetic modes were observed between the parents and their progeny, where over-dominance and under-recessive types were relatively frequent (15.7 and 15.3 %, respectively). We found that 59.1 % of transcripts were more abundant in TD and over fourfold higher transcription levels were found in 66 TD transcripts and three DT transcripts. A higher proportion of over-dominance and a lower proportion of under-recessive transcription types were also observed in TD. The percentage contributions of multiple transcripts at the same locus varied greatly and were transcribed differently between species. In the new allelic combinations created by hybridization, approximately three-fourth of the transcripts had intermediate percentage contributions between the parents but no differential transcription patterns were apparent between the reciprocal hybrids. A broad spectrum of functionally different nuclear genes was over-represented in TD pollen, some of which were directly related to pollen behavior. Since TD and DT pollen had the same composition of nuclear genes, a cytoplasm–nuclear chromosomal genes interaction is suggested for the cause of transcriptional and phenotypic differences between reciprocal hybrids.

Keywords Pollen transcriptome · Interspecific F$_1$ hybrids · Reciprocal difference · Allelic variation · Cytoplasm–nuclear chromosomal genes interaction

R. Sanetomo · K. Hosaka (✉)
NARO Hokkaido Agricultural Research Center,
Shinsei, Memuro, Hokkaido 082-0081, Japan
e-mail: spudman@affrc.go.jp

Introduction

An F$_1$ hybrid and the reciprocal F$_1$ hybrid share the same nuclear genomes but they often exhibit differences in the germination rate, growth rate, yield, environmental resistance, fecundity, crossability, and other characteristics. These phenomena are known as reciprocal differences, and they are common in various plant species. One cause of these differences is considered to be the cytoplasmic genome, which is inherited almost exclusively from the female parent (Roach and Wulff 1987). However, little is known about mechanism by the cytoplasmic genome interacts with nuclear chromosomal genes at the molecular level.

When reciprocal F$_1$ hybrids are produced from interploidy crosses, the first changes are observed during the early seed development stage (Pennington et al. 2008). Imprinted genes, in which only the male or female alleles are expressed exclusively in a parent-of-origin manner (Köhler and Makarevich 2006; Huh et al. 2007), play an

important role during endosperm development in hybrid seeds. In interspecific and inter-ploidy crosses of *Arabidopsis*, endosperm growth is controlled by a balance between maternally contributed Polycomb repressive complex proteins and paternally contributed AGAMOUS-LIKE Type-1 MADS domain transcription factors in a dose-dependent manner (Dilkes and Comai 2004; Josefsson et al. 2006; Walia et al. 2009; Köhler et al. 2010; Gehring et al. 2011). This balance controls whether the seed is normally formed or aborted in reciprocal crosses (Alleman and Doctor 2000; Gehring et al. 2004). In maize endosperm, zein storage protein accumulation begins earlier in balanced crosses and it is continued as the seed matures (Li and Dickinson 2010). Therefore, gene expression during the early stages of embryonic growth can affect the characteristics of the resulting reciprocal F_1 hybrids. Similarly, inconsistent results have been reported between reciprocal hybrids in terms of gene expression, as well as DNA methylation and histone modifications in transcribed regions (Vaughn et al. 2007; Andorf et al. 2010; He et al. 2010).

In our previous study (Sanetomo et al. 2011), we found reciprocal differences in crossability between F_1 hybrids of *Solanum tuberosum* L. ($2n = 4x = 48$) and a Mexican wild potato species *S. demissum* Lindl. ($2n = 6x = 72$). When the *S. tuberosum* (female) × *S. demissum* (male) hybrid (TD hybrid) and the reciprocal hybrid (DT hybrid) were crossed as pollen parents with *S. demissum,* a significantly higher berry-setting rate was obtained in TD (64.9 %) compared with DT (24.2 %). To understand this reciprocal difference, we compared the DNA sequences and methylation status of pollen DNA in TD and DT using methylation-sensitive amplified polymorphism (MSAP) analysis (Sanetomo and Hosaka 2011). Six distinct DNA bands were detected that indicated reciprocal differences between TD and DT, where two of the bands apparently originated from the chloroplast and mitochondrial DNA, while another three bands were shown to be DNA methylation level differences. Since DNA methylation is generally recognized to suppress gene expression by functioning as regulatory factors (Jacobsen and Meyerowitz 1997; Jones and Takai 2001), the methylated DNA regions are speculated to result in differential transcription in male gametes of reciprocal F_1 hybrids.

Recent advances in high-throughput microarray and sequencing technologies have facilitated genome-wide transcription analyses. The transcription levels in seedling stages of reciprocal hybrids were investigated via a genome-wide survey on *A. thaliana* (Andorf et al. 2010), rice (He et al. 2010), and maize (Guo et al. 2004). In the reproductive tissues, there was a dramatic change in transcription patterns throughout the embryo developmental time course in a wild diploid potato species *S. chacoense*

(Tebbji et al. 2010). A significantly lower complexity of pollen mRNA populations has been reported in *Tradescantia* (Willing and Mascarenhas 1984) and maize (Willing et al. 1988). In the mature pollen, only 27.5 % genes are expressed in soybean (Haerizadeh et al. 2009), and only 13–33 % genes, in *Arabidopsis* (Honys and Twell 2003; Pina et al. 2005; Borges et al. 2008), which are much lower than those in other vegetative tissues such as the flowers (68 %), leaves (62 %), and seedlings (68 %) of *Arabidopsis* (Pina et al. 2005). The lower transcription percentages in pollen reflect the fact that vegetative tissues contain variety of differentiated tissues, whereas the pollen transcripts originate from single cells, or they might simply reflect the specialization of pollen transcripts in preparation for dramatic changes in the pattern of cell growth during pollen germination and pollen tube growth (Honys and Twell 2003, 2004; Pina et al. 2005; Hafidh et al. 2012). The number of genes and the transcription levels increase from the desiccated mature pollen grains to hydrated pollen grains and then to pollen tubes of *Arabidopsis* (Wang et al. 2008). Compared with representative sporophytic tissues and pollen, sperm cells have a distinct but diverse transcriptional profile (Borges et al. 2008).

Based on cytological observations of chromosome pairing behavior during metaphase I, *S. tuberosum* (AAA^{t}A^t) and *S. demissum* (AADDD^{d}D^d) had different genome compositions (Matsubayashi 1991), which resulted in a complex allelic composition at each locus in the F_1 hybrid (AAA^{t}DD^d). Compared with the parents, the gene expression may be affected by the genetic mode in a hybrid. In a maize F_1 hybrid and its inbred parents, all possible modes of gene action were observed, including additivity, high- and low-parent dominance, under-dominance (referred to as under-recessive in this paper), and overdominance (Swanson-Wagner et al. 2006). Furthermore, two alleles are often unequally transcribed in a diploid species. This type of differential expression between the alleles of an autosomal non-imprinted gene is relatively common, and allelic differences are heritable (Knight 2004; Guo et al. 2004, 2006; Vuylsteke et al. 2005; Zhuang and Adams 2007). Guo et al. (2004) found that 11/15 genes analyzed exhibited allele-specific expression in maize hybrids and that the differential expression changed in different tissue types, environments, and stress conditions, which may play important roles in heterosis (Guo et al. 2004, 2006). Silencing or unequal expression of homoeologous alleles and the remodeling of DNA methylation were reported in *Arabidopsis* (Comai et al. 2000; Madlung et al. 2002), and *Gossypium* (Adams et al. 2003), when two different genomes were joined to form allotetraploids. However, allele-specific expression variation in a diploid/polyploid hybrid plant where two or more alleles are compared in the same genetic context is poorly understood.

In this study, we performed a whole-genome transcription analysis for the pollen mRNA of *S. tuberosum*, *S. demissum*, and their reciprocal F$_1$ hybrids TD and DT using a high-throughput sequencer. We aimed to identify transcriptional differences between the parents and their progeny, and between the reciprocal hybrids by comparing the abundance of each transcript and the proportional contribution of transcripts at each locus. For the first time, we described allele-specific transcripts in the pollen of highly polyploid hybrids (pentaploid) while the possible causes of reciprocal differences are discussed.

Materials and methods

Plant materials

The parents used in this study were a *S. tuberosum* breeding line Saikai 35 (referred to as T) and seedlings of the 10H3 family, which were derived by selfing from one *S. demissum* PI 186551 plant (referred to as D). The *S. demissum* accession was obtained from the US Potato Genebank (NRSP-6; Sturgeon Bay, Wisconsin). The interspecific hybrid family 6H37 was obtained by a cross between *S. demissum* as the female and *S. tuberosum* as the male (DT), while the 6H38 family was obtained via the reciprocal cross (TD). One T, 14 TD, and 11 DT plants were propagated clonally and grown in the same field. Seedlings of D were grown in a greenhouse during the same season as field-grown plants.

RNA extraction

Pollen was collected, quick-frozen in liquid nitrogen, and ground to a powder using a mortar and pestle. Total RNA was extracted using an RNeasy® Plant Mini Kit (QIAGEN, ME, USA). To eliminate the possibility of contamination with genomic DNA, all RNA samples were treated with DNase (TURBO DNA-*free*™, Ambion, TX, USA) for 20 min at 37 °C. The RNA samples were quantified using an ND-1000 NanoDrop spectrophotometer (Thermo Scientific, DE, USA), and the quality was checked using an Agilent 2100 Bioanalyzer (Agilent Technologies, CA, USA). Equal amounts of RNA from the 14 TD and 11 DT samples were combined as TD and DT, respectively. Four cDNA libraries were generated from T, D, TD, and DT using an mRNA-Seq Sample Prep Kit (Illumina, CA, USA).

Illumina sequencing

Four cDNA libraries were sequenced by a commercial service provider (Hokkaido System Science Co., Ltd.,

Sapporo, Japan). A 75-base single-end run was performed on the Illumina GA IIx platform using one sample per lane in a flow cell. All 75 bases in each read were filtered to ensure the sequence quality and complexity and assembled using the Velvet program (Zerbino and Birney 2008) for de novo assembly. The hash length in base pairs (*k*-mer length) was determined using a series of *k*-mers to optimize the Velvet assembly for higher transcript contiguity (longer transcript length) and specificity (fewer spurious overlaps). Overlapping contigs were grouped and regarded as a transcript using the program Oases (version 0.1.8, http://www.ebi.ac.uk/~zerbino/oases/). Sequence reads of each sample were mapped with a maximum of two nucleotide mismatches in the assembled transcripts to quantify the abundance of the transcripts using the program Bowtie (Langmead et al. 2009). The abundance, or the coverage of each transcript was determined by read counts and normalized using the number of reads per kilo base exon per million mapped reads (RPKM) (Mortazavi et al. 2008). The RPKM value of the read density reflected the molar concentration of a transcript in the starting sample after normalizing for the RNA length and total read number in the measurements. This facilitated a transparent comparison of transcript levels within and between samples.

Identification of chloroplast and mitochondrion sequences

Chloroplast and mitochondrial genome sequences were identified using reference genome sequences of the potato chloroplast (Chung et al. 2006, DQ231562.1) and mitochondrion (The Potato Genome Sequencing Consortium 2011, *S. tuberosum* Group Tuberosum RH89-039-16 mitochondrion sequences).

Single nucleotide polymorphism (SNP) genotyping

SNPs were searched using SAMtools (Li et al. 2009), and mapped reads harboring SNPs were visualized using the graphical viewer program Tablet (Milne et al. 2010), which allowed us to further separate one transcript into multiple transcript variants. The parental origins of transcript variants were determined by comparison with the parental transcripts.

Homology search and functional annotation

Transcript sequences were subjected to BLASTX and nucleotide BLAST analysis (provided by the National Center for Biotechnology Information) to deduce their putative functions. The results were extracted only for the best hit, and general functional classification of genes was performed based on the corresponding gene ontology (GO)

terms (http://www.geneontology.org/GO.current.annotations.html). The GO Slim term was searched using public databases in the Arabidopsis Information Resource (TAIR; http://www.arabidopsis.org/tools/bulk/go/index.jsp), Sol Genomics Network (http://solgenomics.net/search/template) and The European Bioinformatics Institute (EBI; http://www.ebi.ac.uk/). For double-category assignments, the biologically most informative category was chosen, or a choice was made based on additional biological information. The GO functional classes were assigned using GO Terms Classifications Counter (http://www.animalgenome.org/tools/catego/) and classified in terms of the main function and children functions in molecular function and biological process categories.

Results

Data summary

After filtering the low-quality sequence data, a total of 3,035,230–3,204,206 kb (72.5–80.3 %) of sequences per sample were generated by a single run of 75 cycles using Illumina GA IIx (Table 1). The trimmed high-quality sequence reads were assembled using the Velvet program with different k-mer lengths of 51, 61, 69, and 73. The best assembly was achieved using $k = 69$, with which all reads were assembled, which yielded a total of 37,238 contigs measuring at least 100 bp in length. These contigs were grouped and assembled into transcripts using the Oases program with the default parameters, thereby generating 13,020 transcripts with 9,366 loci.

High-quality reads of each sample were mapped using a maximum of two nucleotide mismatches in the transcript sequences. Some transcripts had no reads in a sample, so the actual number of mapped transcripts varied from 12,595 in T to 13,018 in TD. The number of mapped reads for each transcript was normalized using the RPKM values. The RPKM values varied greatly among loci from 0.0 to 13,547.2, but 70.3–77.3 % of transcripts had <1,000 RPKM values (Table 1; Fig. 1). The mean RPKM value of each sample ranged from 151.2 in D to 156.5 in T.

Parental differences in the RPKM values

Of the 13,020 transcripts, 425 transcripts were not transcribed in T, while 174 transcripts were not transcribed in D. We found that 12,421 transcripts were transcribed in both T and D, although the transcription levels, particularly those with lower RPKM values, differed between T and D (Fig. 1a). The RPKM values of 7,117 transcripts (57.3 %) were higher in D than in T.

Differences in the RPKM values of TD and DT

In contrast to the large difference between T and D, the RPKM values of the respective transcripts in TD were mostly similar to those in DT (Fig. 1b). Four transcripts were transcribed in either TD or DT, whereas one transcript was D-specific. The remaining 13,015 transcripts were transcribed in both TD and DT. The transcription levels of 7,697 transcripts (59.1 %) were higher in TD than DT (Table 2). Furthermore, TD generated more transcripts with larger TD:DT ratios (Table 2). For example, over fourfold higher transcription levels were found for three transcripts in DT compared with 66 transcripts in TD. These resulted in TD–DT plots that shifted slightly toward the upper side of the y-axis in Fig. 1b.

Comparison of the RPKM values of parents and their progeny

Based on comparisons with their parental and mid-parental RPKM values, that is, (T + D)/2 RPKM values, the transcription levels of TD and DT transcripts were classified into seven categories (Fig. 2): (1) the RPKM values of 4,081 transcripts (15.7 %) were higher than those of the high-parent +12.5 %; (2) 1,385 (5.3 %) were within the range of the high-parent ± 12.5 %; (3) 2,789 (10.7 %) were within the range between the high-parent −12.5 % and the mid-parent +12.5 %; (4) 5,972 (22.9 %) were within the range of the mid-parent ± 12.5 %; (5) 5,529 (21.2 %) were within the range between the mid-parent −12.5 % and the low-parent +12.5 %; (6) 2,297 (8.8 %)

Table 1 Summary of obtained data

Sample	T	D	TD	DT
No. of bases (kb)	3,175,152	3,204,206	3,146,108	3,035,230
No. of mapped reads	21,355,449	23,318,726	21,475,383	17,246,376
No. of transcripts	12,595	12,846	13,018	13,016
No. of transcripts with less than 1,000 RPKM values	9,463	9,037	9,349	10,064
Total RPKM values	2,037,300.4	1,968,924.6	2,005,482.3	2,003,634.7
Mean RPKM value (Standard deviation)	156.5 (592.99)	151.2 (523.65)	154.0 (564.65)	153.9 (568.49)

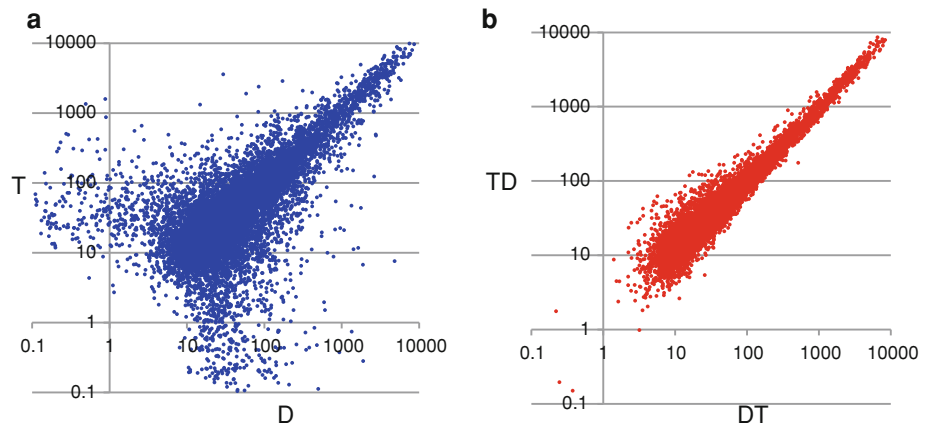**Fig. 1** Comparisons of the RPKM values of 13,020 transcripts in T and D (**a**), and TD and DT (**b**)

Table 2 Comparisons of the RPKM values in 13,019 transcripts in TD and DT

TD : DT	No. of transcripts	RPKM value in TD					RPKM value in DT				
		<10	<100	<1,000	>1,000	Mean	<10	<100	<1,000	>1,000	Mean
TD < DT											
0:10	1	1	0	0	0	0	1	0	0	0	0.27
0:10–1:9	0	–	–	–	–	–	–	–	–	–	–
1:9–2:8	2	0	2	0	0	4.02	2	0	0	0	18.57
2:8–3:7	26	4	21	1	0	14.02	18	7	1	0	38.06
	mt (1)	(1)							(1)		
3:7–4:6	412	39	330	42	1	34.71	135	252	24	1	56.73
	mt (12)	(2)	(9)	(1)			(1)	(9)	(2)		
	ct (2)	(2)						(2)			
4:6–5:5	4,881	220	3,117	1,303	241	193.73	332	3,155	1,182	212	219.97
	mt (48)	(4)	(33)	(9)	(2)		(1)	(34)	(11)	(2)	
	ct (20)	(2)	(16)		(2)		(1)	(17)		(2)	
Subtotal	5,322										
TD > DT											
5:5–6:4	6,133	397	4,732	838	166	158.83	192	4,783	979	179	141.69
	mt (43)	(1)	(35)	(7)			(3)	(33)	(7)		
	ct (12)	(2)	(9)	(1)			(2)	(9)	(1)		
6:4–7:3	1,230	313	888	29	0	44.30	50	1,100	77	3	25.35
	mt (26)		(22)	(4)			(3)	(20)	(3)		
	ct (4)			(4)				(1)	(3)		
7:3–8:2	268	116	148	4	0	49.91	9	236	23	0	17.90
	mt (7)		(5)	(2)			(5)		(2)		
	ct (2)			(2)					(2)		
8:2–9:1	53	31	22	0	0	53.54	2	44	7	0	10.46
	mt (4)		(4)				(4)				
9:1–10:0	10	10	0	0	0	39.77	0	10	0	0	3.91
10 : 0	3	3	0	0	0	0.70	3	0	0	0	0
Subtotal	7,697										
Total	13,019	1,134	9,260	2,217	408	154.04	744	9,587	2,293	395	153.90
Mitochondrial	(141)	(8)	(108)	(23)	(2)	135.29	(17)	(97)	(25)	(2)	127.58
Chloroplast	(40)	(6)	(25)	(7)	(2)	219.34	(3)	(29)	(6)	(2)	192.35

The number of transcripts possibly derived from chloroplast (ct) and mitochondrial (mt) genes is indicated in parentheses

were within the range of the low-parent ± 12.5 %; (7) 3,987 (15.3 %) were lower than the low-parent –12.5 %. The categorization scale is shown in Fig. 2.

We found that 2,986 (22.9 %) of TD and 2,986 (22.9 %) of DT transcripts were within the range of the mid-parent ± 12.5 %, while 6,975 (53.6 %) of TD, and 7,315 (56.2 %) of DT transcripts were within the range of mid-parent ± 37.5 % (Fig. 2). Furthermore, 780 (6.0 %) of TD and 605 (4.6 %) of DT transcripts were within the range of the high-parent ± 12.5 %, while 977 (7.5 %) of TD and 1,320 (10.1 %) of DT transcripts were within the range of the low-parent ± 12.5 %. Thus, 67.1 % of TD and 71.0 % of DT transcripts were almost similar to the parental values or between the parental values. Interestingly, 2,571 (19.7 %) of TD and 1,510 (11.6 %) of DT transcripts had higher transcription levels than those of the high-parent (over-transcription), while 1,717 (13.2 %) of TD and 2,270 (17.4 %) of DT transcripts had lower transcription levels than those of the low-parent (under-transcription) (Fig. 2). Thus, TD had a higher number of over-transcribed and a smaller number of under-transcribed transcripts than DT, which contributed to a higher proportion of TD transcripts skewed toward higher transcription levels compared with DT.

Chloroplast and mitochondrial transcripts

Based on homology searches using the complete chloroplast and mitochondrial genome sequences, we detected 40 (30 loci) chloroplast and 141 (99 loci) mitochondrial transcripts (Table 3). Of these, 21 transcripts (17 loci) were only found in the chloroplast genome and 39 transcripts (34 loci) were exclusive to mitochondrial genome. The remaining transcripts were shared between the chloroplast and mitochondrial genomes (eight transcripts at eight loci) or the nuclear chromosomes (113 transcripts at 70 loci). Seven and two mitochondrial transcripts were found only in D and T, respectively, and four chloroplast transcripts only in D. However, all of these transcripts were transcribed in both TD and DT (Table 2), which did not support their maternal inheritance.

The RPKM values of 18 chloroplast transcripts were higher in TD, whereas those of 22 chloroplast transcripts were higher in DT. However, the RPKM values of 80 mitochondrial transcripts were higher in TD, whereas those of 61 mitochondrial transcripts were higher in DT. There was a greater than fourfold difference between TD and DT in only four mitochondrial transcripts (the ratios of TD:DT were 8:2–9:1 for locus nos. 4069, 5276, 7742, and 9231;

Fig. 2 Comparisons of the RPKM values of the parents and the TD and DT progeny. Of the 13,020 transcripts, T was > D for 5,478 transcripts (**a**), whereas D was > T for 7,542 transcripts (**b**). The transcription levels of the progeny were categorized using the scale shown on the *left. Parentheses* show the numbers of mitochondrial and chloroplast transcripts, respectively
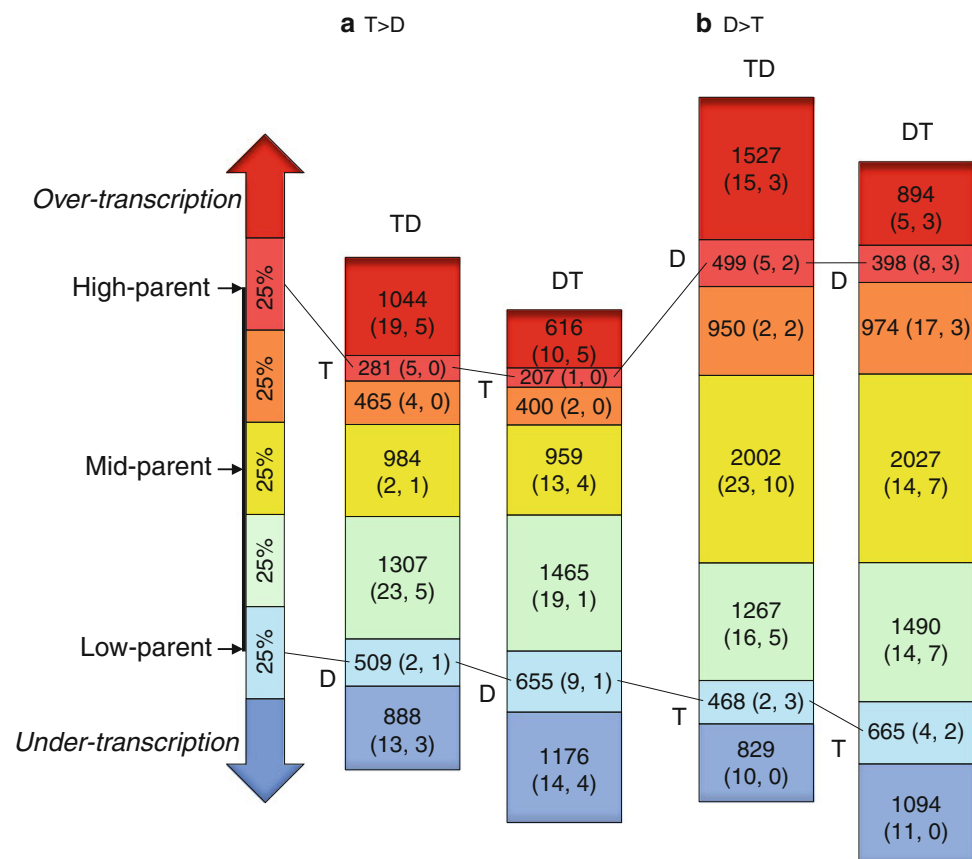
Table 2). Loci 5276 and 9231 had partial homology to the nuclear genome, whereas the remaining two transcripts originated from the mitochondrial genome only.

Intra-locus percentage contributions

Of the 13,020 transcripts, 7,474 were single transcripts from 7,474 loci, whereas 5,546 transcripts were multiple transcripts from 1,718 loci. We found that 763 and 433 loci produced two and three transcripts, respectively, with up to 13 transcripts/locus. The percentage contributions of each transcript to the locus were calculated as $100 \times$ (RPKM value of a transcript/total RPKM value of all transcripts within a locus). If two or three transcripts were transcribed equally from a locus, percentage contributions of 50 and 33.3 % were expected, respectively. However, the percentages varied widely both among transcripts and between T and D ($r = 0.656$) (Fig. 3a). By contrast, very similar percentage contributions for the respective transcripts were found between TD and DT ($r = 0.991$), although the percentages varied widely among transcripts from 0 to 100 % (Fig. 3b).

Comparison of the intra-locus percentage contributions of the parents and their progeny

To investigate the genetic mode by which alleles of the same locus from two species were expressed in the complex composition of the F_1 hybrid, the intra-locus percentage contributions of 5,546 transcripts from 1,718 loci in TD and DT were compared with those of the corresponding parental percentages, which were categorized as shown in Fig. 4. We found that 2,056 (37.1 %) of TD and 1,957 (35.3 %) of DT transcripts were within the range of mid-parental percentages $\pm$ 12.5 %, while 4,124 (74.4 %) of TD and 4,034 (72.7 %) of DT transcripts were within the range of the mid-parental percentages $\pm$ 37.5 %. Thus, approximately three-fourths of the transcripts had intermediate percentages between the parental percentages in

their progeny. The remaining transcripts were either within the range of the D-parent percentages $\pm$ 12.5 % (853 transcripts) or the T-parent percentages $\pm$ 12.5 % (421 transcripts), or they exceeded the D-parent percentages +12.5 % (938 transcripts) and the T-parent percentages (722 transcripts). Thus, the progeny percentages tended to be intermediate between the parental percentages or fairly similar to the D-parent percentages (Fig. 4).

SNP genotyping of transcripts

Each transcript could be further distinguished into transcript variants. One hundred loci were selected from those that exhibited the largest differences in TD and DT (referred to as the Top 100 loci). The Top 100 loci included three loci from TD < DT and 97 loci from TD > DT based on the transcripts in Table 2, which contained two mitochondrial loci (loci 7742 and 9231). These loci were analyzed for SNPs to distinguish the transcript variants. For 17 loci, including the mitochondrial locus 9231, no SNPs were found in the transcripts, whereas SNPs distinguished 260 transcript variants for 83 loci. The sequences were compared with those of their parents to determine the parental origins (Table 4). One transcript variant in TD was not found in both parents. We found that 78 and 111 transcript variants were T-derived and D-derived, respectively, while 70 were detected in both parents. Thus, each progeny locus was composed of an average of 0.78 (ranging from 0 to 3) T-derived, 1.11 (ranging from 0 to 3) D-derived, and 0.87 (ranging from 0 to 3) common transcript variants. The former two transcript variants were designated as parent-specific transcripts.

The mitochondrial locus 7742 was composed of two D-derived and one T-derived transcripts, all of which were transcribed in both TD and DT, thereby indicating its nuclear chromosome origin.

Intra-locus percentage contributions of parent-specific transcripts and their differential transcription in TD and DT

The abundance of parent-specific transcripts was quantified based on their observed read numbers. As the Top 100 mainly comprised loci where TD > DT, 93.7 % (177/189) of the parent-specific transcripts had higher read numbers in TD than DT. The intra-locus percentage contributions were positively correlated between TD and DT for T-derived ($r = 0.739$) and D-derived transcripts ($r = 0.577$), but 16 (20.5 %) of 78 T-derived and 26 (23.4 %) of 111 D-derived transcripts were transcribed only in TD, which indicated that, irrespective of their parental origins, one-fifth of the parent-specific transcripts were not transcribed in DT (Fig. 5). Of the Top 100 loci, 50 loci were composed of both T- and D-derived transcripts. These were analyzed further

Table 3 Putative mitochondrial and chloroplast genes among 13,020 transcripts at 9,366 loci

Primary origin	Secondary origin	No. of loci	No. of transcripts
Mitochondria	–	34	39
Mitochondria	Chloroplast	4	4
Mitochondria	Nuclear chromosome	61	98
Chloroplast	–	17	21
Chloroplast	Mitochondria	4	4
Chloroplast	Nuclear chromosome	9	15

Fig. 3 Intra-locus percentage contributions of each transcript from 1,718 loci with multiple transcripts (total of 5,546 transcripts) for T and D (**a**), and TD and DT (**b**)
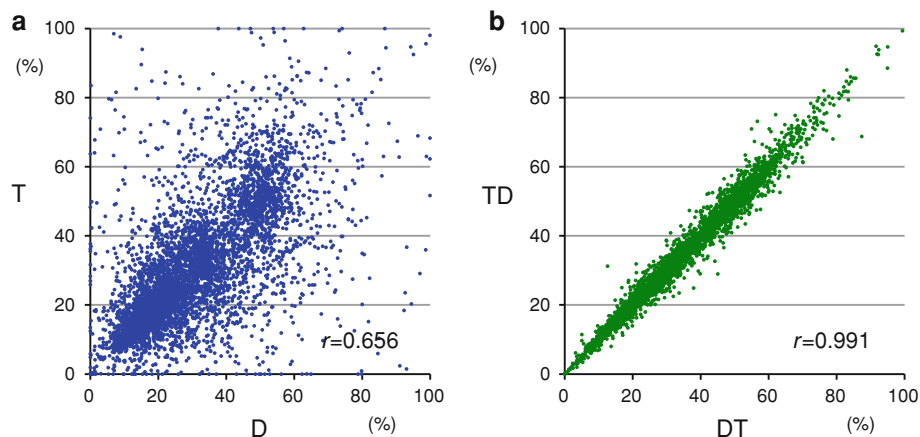

Fig. 4 Intra-locus percentage contributions of 5,546 transcripts from 1,718 loci for the parents and TD and DT progeny. It can be seen that 2,781 transcripts had higher intra-locus percentage contributions in T compared with D (**a**), while those of 2,765 transcripts had higher percentages in D than T (**b**). The percentage contributions of the progeny were categorized using the scale shown below
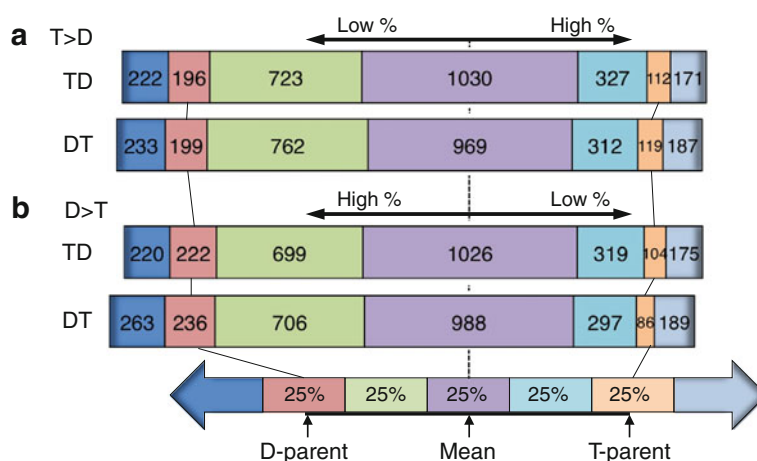

Table 4 The number of transcript variants in the Top 100 loci distinguished by SNP genotyping

Parental origin	No. of transcript variants per locus	No. of transcript variants
T-derived	1	35
	2	34
	3	9
	Mean	0.78
D-derived	1	41
	2	58
	3	12
	Mean	1.11
Common	1	48
	2	16
	3	6
(No SNP)	1	17
	Mean	0.87
Total		277[a]

[a] One transcript variant was detected only in TD, which was included in the total number

to determine how often they were transcribed in a parent-of-origin manner in TD and DT. With the exception of four loci (loci 3794 and 4967, 6582 and 6805), the total read numbers of D-derived transcripts for a locus were larger or smaller than those of T-derived transcripts in both TD and DT. In the loci 3794 and 4967 and 6582, the total read numbers of D-derived transcripts in TD were higher than those of T-derived transcripts, whereas the former were lower than the latter in DT (Table 5). By contrast, the read number was lower in the D-derived transcript than the T-derived transcript at locus 6805 in TD, whereas the reverse was true in DT. Only these four parent-specific transcripts were differentially transcribed in TD and DT. However, the differences in read numbers between T- and D-derived transcripts in DT and the read numbers themselves were very small (Table 5). Thus, the differential transcription in TD and DT might be negligible.

Homology search and functional annotation

Homology searches were performed only for the Top 100 loci using BLASTX and nucleotide BLAST. Known

sequences were hit by 95, of which 90 had putative functions (Supplementary Table). Transcription of the top three loci [cold-stress-inducible protein C17, chloroplast ribulose-1,5-bisphosphate carboxylase/oxygenase (Rubisco) activase, and Rubisco small subunit] was limited to D (RPKM values of 47.4–84.7), and only traces of their transcription were detected in DT (RPKM values of 0.2–1.7). Since Rubisco is generally not expressed in pollen, these transcripts might be resulted from contaminating sporophytic tissues. The next five loci shared homology with coding sequences of SKP1, putative vicilin, putative methyltransferase PMT27, SLF-interacting SKP1, and beta-tubulin in that order.

Six loci matched the same or related genes for late embryogenesis abundant (LEA) protein. Five loci matched the same gene for putative methyltransferase PMT27. Likewise, multiple loci shared homology with single genes, that is, inorganic phosphate transporter 1 (three loci), pre-pro-cysteine proteinase (three loci), adenosine kinase isoform 2T (two loci), chloroplast Rubisco activase (two loci), putative vicilin (two loci), S-adenosyl-L-homocysteine hydrolase-like (three loci), cyclophilin ROC7-like (two loci), glyceraldehyde 3-phosphate dehydrogenase-like (two loci), phosphoglycerate kinase-like (two loci), elongation factor 1-alpha-like (two loci), and 34 kDa outer mitochondrial membrane protein porin-like (two loci) proteins. The remaining known loci (54 loci) were assigned to single genes. Thus, the Top 100 loci were actually reduced to 77 loci, which included 10 loci with unknown functions.

The functional roles were assigned as GO terms (Fig. 6). Five main GO categories were found for biological processes: metabolic processes (30 loci), responses to stress (nine loci), transport (eight loci), cellular processes (eight loci), and reproduction (three loci). The biological functions of 19 loci were unknown. Furthermore, each category was subdivided using the children terms into 2–7 sub-categories. Seven main GO categories were found for the

Table 5 Four loci showing imprinting-like transcription, which were found from the 50 loci, consisted of both T- and D-derived transcript variants, among the Top 100 loci

Rank	Locus	Transcript variant	No. of reads			
			TD	DT	T	D
8	Locus 6805	T-derived	23	0	11	0
		D-derived	15	1	0	6
	RPKM value		23.39	2.24	5.65	2.90
52	Locus 6582	T-derived	5	4	4	0
		T-derived	5	0	8	0
		D-derived	27	3	0	17
		Common	12	3	5	1
	RPKM value		33.25	7.43	9.60	7.70
75	Locus 3794	T-derived	14	4	11	0
		T-derived	4	3	7	0
		D-derived	24	2	0	10
		D-derived	35	1	0	13
		D-derived	0	0	0	1
		Common	15	4	6	6
		Common	44	9	13	11
	RPKM value		82.33	21.85	20.50	28.35
85	Locus 4967	T-derived	1	1	10	0
		D-derived	2	0	0	2
		Common	26	4	9	2
	RPKM value		22.30	6.20	13.72	5.78

molecular functions, that is, catalytic activity (21 loci), binding (13 loci), structural molecule activity (nine loci), transporter activity (seven loci), transferase activity (two loci), enzyme regulator activity (one locus), and transcription factor activity (one locus). The molecular functions of 23 loci were unknown. Each category of catalytic

Fig. 5 Intra-locus percentage contributions of SNP-based parent-specific transcripts in the Top 100 loci. We compared 78 T-derived transcripts (**a**) and 111 D-derived transcripts (**b**) in TD and DT
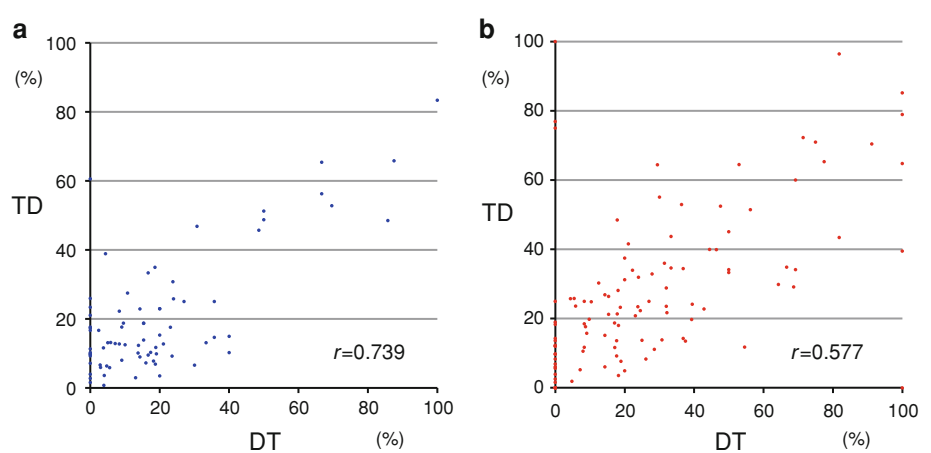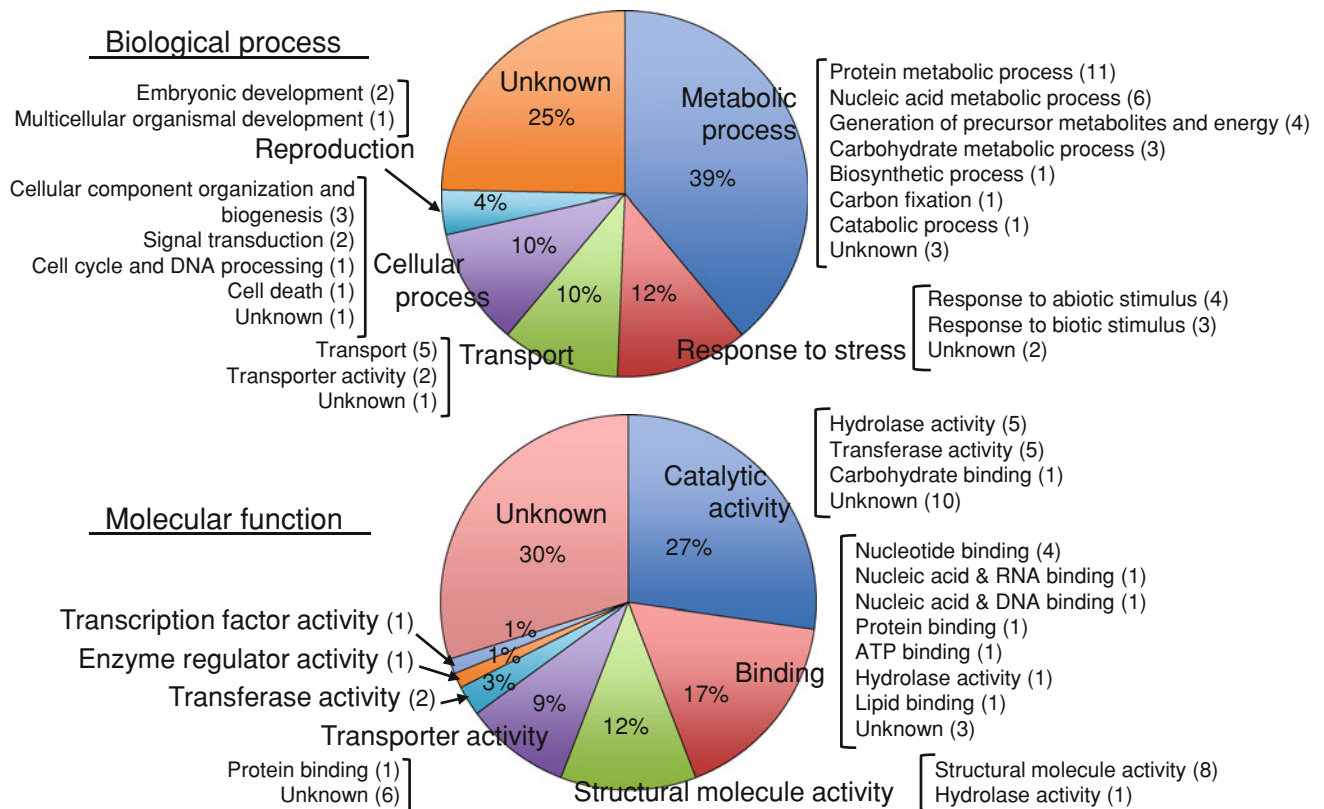
Fig. 6 Gene ontology (GO) analysis of the Top 100 loci, where 77 loci were actually analyzed because multiple loci originated from the same genes. These loci were classified into five main biological process GO categories and seven molecular function categories

activity, binding, and structural molecule activity was subdivided into three, seven and two sub-categories, respectively.

Discussion

The numbers of comparable transcripts and loci

In this study, a high-throughput sequencer generated 12.6 billion bases from pollen mRNA, which were assembled into 13,020 transcripts with 9,366 loci. The number of loci appeared to be overestimated, because the homology search indicated that a maximum of 77 loci were actually involved in the Top 100 loci. At present, 39,031 protein-coding genes have been recognized in the potato genome (The Potato Genome Sequencing Consortium 2011). Therefore, $\leq$18.5 % (9,366 × 0.77/39,031) of the total genes was transcribed in pollen. Low transcript percentages in mature pollen have been reported in various species (Willing and Mascarenhas 1984; Willing et al. 1988; Honys and Twell 2003; Pina et al. 2005; Borges et al. 2008; Haerizadeh et al. 2009) because pollen are highly specialized single cells (Honys and Twell 2003, 2004; Pina et al. 2005).

However, the number of transcripts appeared to be underestimated. We found that there was an average of 0.78 T-derived, 1.11 D-derived, and 0.87 common transcript variants per locus for the Top 100 loci. These result predicted 25,850 (2.76 × 9,366) transcripts, which is twice the number detected in this study. This underestimation was partly due to the assembling of contigs. Similarly, the number of transcripts may have been overestimated, because a maximum of 77 loci were actually identified in the Top 100 loci. If we apply the same reduction rate for the loci obtained from the Top 100 loci, the number of transcripts would be 10,025 (0.77 × 13,020). Therefore, although the actual number of transcripts was uncertain, over 10,000 transcripts were comparable in the parents and the progeny and the reciprocal F_1 hybrids.

Genetic mode of the transcription levels

The transcription levels in hybrids often deviate from the mid-parental levels of their homozygous parents in a genome-wide surveys (Vuylsteke et al. 2005; Guo et al. 2006; Swanson-Wagner et al. 2006; Zhuang and Adams 2007; Guo et al. 2008; Wei et al. 2009; Andorf et al. 2010; He et al. 2010; Riddle et al. 2010).

As with a maize F_1 hybrid and its inbred parents (Swanson-Wagner et al. 2006), all possible modes of gene action were observed in our study. We found that 54.9 % of the transcripts in the F_1 hybrids were intermediate between, or close to, the parental RPKM values, which were probably due to additive or partial dominance/recessive genetic effects. The similar transcription levels of the progeny and those of either one of the parents probably reflected complete dominance or recessive modes, although these were found rarely (5.3 and 8.8 %, respectively) (Fig. 2). Over-transcription above the high-parent transcription level or under-transcription below the low-parent transcription level probably indicated over-dominance or under-recessive modes, respectively, which were found in a relatively high percentage of cases (15.7 and 15.3 %, respectively). Among 20,638 genes analyzed in reciprocal inter-subspecies rice hybrids, He et al. (2010) also found that 3,261 of the Nipponbare × 93-11 hybrid and 3,229 of the reciprocal hybrid exhibited non-additive transcription patterns, where 20.2–39.5 % indicated over-transcription and 16.4–34.6 % indicated under-transcription. It is easy to see that over-transcription could be sources of heterotic effects in hybrids.

Allelic variation and the genetic mode

Multiple transcripts were found for loci in TD and DT, where each locus had an average of 0.78 T-derived, 1.11 D-derived, and 0.87 common transcript variants. This was expected because of the heterozygosity and polyploidy (a genome composition of AAA^tDD^d) present in our materials. We found that the member transcripts at a locus had different levels of transcription, which were represented by different percentage contributions and they were transcribed differently between species (Fig. 3a). Differential transcription among the alleles at a locus is a commonly reported observation (Knight 2004; Guo et al. 2004, 2006; Vuylsteke et al. 2005; Zhuang and Adams 2007). This suggests that alleles function and are organized competitively and/or harmoniously within a species and species-specifically.

Allelic variation is often altered by the new allelic combinations created by hybridization (Guo et al. 2006; Zhuang and Adams 2007). To the best of our knowledge, our study is the first to describe how the member transcripts of parents behave in immediate interspecific polyploid hybrids ($2n = 5x = 60$). We found that the species-specific and/or allele-specific percentage contributions of the parental species were mostly intermediate between those of the parents in the F_1 hybrids, suggesting that genetic transmission primarily occurred via the additive mode (Fig. 4). For example, assuming two alleles with intra-locus contribution percentages of 50 and 50 % in one

species and two alleles with intra-locus percentages of 30 and 70 % in the other species, the resulting interspecific hybrid would have intra-locus percentages of 25, 25, 15, and 35 %, respectively. In addition, the percentage contributions did not differ between reciprocal hybrids (Fig. 3b). This may suggest that immediate interspecific hybrids have not established a new gene expression network, which may be one explanation for the relatively high proportion of loci exhibiting over- or under-transcription. Furthermore, this unorganized gene expression may be associated with hybrid vigor. In this context, the parental species may have undergone a stabilization process when establishing a new gene expression network shaped by specialized allelic contributions (Adams et al. 2003).

Reciprocal difference

Most of the parental pollen transcripts were transcribed in both TD and DT pollen, but the abundance of transcripts differed reciprocally (Table 2). We found that 59.1 % of transcripts were more abundant in TD and greater than fourfold higher transcription levels were found in 66 transcripts in TD, but only three in DT. Inconsistent percentages of over-transcription and under-transcription patterns were also observed in TD and DT (Fig. 2). Recent genome-wide transcription analyses have also detected significant numbers of reciprocal differences in the seedlings of inter-varietal hybrids between two homozygous *A. thaliana* lines C24 and Columbia (Andorf et al. 2010), in the seedlings of inter-subspecies hybrids between the homozygous *Oryza sativa* lines Nipponbare (ssp. *japonica*) and 93-11 (ssp. *indica*) (He et al. 2010), and in the developing endosperm and embryo of the same rice reciprocal hybrids (Luo et al. 2011).

Causal factors affecting differential transcription in TD and DT pollen

Differential transcription in TD and DT pollen might be due to a direct maternal effect (non-genetic), cytoplasmic genomic effect, and/or imprinting (differential expression of maternal or paternal alleles in a preferentially or exclusively uniparental manner). The possibility of a direct maternal effect, however, must have been minimal, and it can be ignored because the plants were grown in nearby rows in the same field and the pollen itself was a combined collection from the progeny population.

We found 40 (30 loci) transcripts and 141 (100 loci) transcripts with homologous sequences to the chloroplast and mitochondrial genomes, respectively (Table 3). However, of these (1) only four transcripts with mitochondrial sequences exhibited fourfold reciprocal differences; (2) at least 61.2 % were orthologous to nuclear chromosomal

genes; (3) no transcripts supported maternal inheritance. In addition, mRNAs with poly(A) tails from the chloroplast and mitochondrial genes were generally found only during degradation so they were expected only to comprise a minor fraction of the steady-state pool (Forner et al. 2007; del Campo 2009). The present experimental procedure could only read poly(A)-tailed mRNA. Therefore, these transcripts were probably orthologs encoded in the nuclear genome so they did not contribute to the differential transcription patterns in TD and DT pollen mRNAs. In this study, therefore, we could not provide positive support for a direct cytoplasmic effect on reciprocal differences, although we found that the cytoplasmic difference unambiguously affected the differential crossability of TD and DT pollen (Sanetomo et al. 2011).

The intra-locus percentage contributions of the parent-specific transcripts were very similar between TD and DT (Fig. 3b), even for those in the Top 100 loci (Fig. 5). The SNP-based analysis of parent-specific transcripts in the Top 100 loci only detected four loci that exhibited imprinting-like transcription profiles. Even for these imprinting-like loci, however, the difference might not have been caused in a parent-of-origin manner but in an allelic bias manner, due to parental allelic variation at the locus (Table 5). Guo et al. (2004) described that a parent-of-origin effect was minimal in maize hybrids from reciprocal crosses because maternal or paternal transmission had little effect on the allele-specific transcript ratio. Most imprinted genes are expressed in various tissues in rice plants, while their expression in a parent-of-origin manner is limited to the endosperm and embryo (Gehring et al. 2011; Luo et al. 2011). Therefore, we conclude that the imprinted genes were not significantly involved in the reciprocal differences in TD and DT pollen.

Differential transcription of diverse genes in reciprocal hybrids

The pollen-expressed genes of *Arabidopsis* are over-represented in the Gene Ontology (GO) categories such as cell wall metabolism, signaling, the cytoskeleton, and membrane transport (Honys and Twell 2003, 2004; Pina et al. 2005). The up-regulation of transporter activities has been reported in pollen in various species such as soybean (Haerizadeh et al. 2009), *Arabidopsis* (Wang et al. 2008), and tobacco (Hafidh et al. 2012). The over-representation of these categories or some unique or pollen-specific transcripts selectively activated during pollen maturation might reflect the increasing functional specialization of mature pollen in preparation for a dramatic change in the pattern of cell growth during pollen germination and pollen tube growth (Honys and Twell 2004; Pina et al. 2005; Wang et al. 2008; Hafidh et al. 2012). In our study,

however, the over-representation of transcripts in these categories contributed to the reciprocal differences to a minor extent. Instead, the Top 100 loci were classified in a diverse range of GO categories (Fig. 6). We did not compare the pollen transcripts with other organ transcripts, so it remains unknown whether the reciprocally differential transcription was pollen-specific. However, it was notable that a broad spectrum of genes contributed to the reciprocal differences.

Among these genes, we found that several interesting genes were highly transcribed or more intensely accumulated in TD pollen compared with DT pollen, which contributed to the reciprocal differences. SSK1 is expressed specifically in pollen, and it acts as an adaptor in the SCF (Skp1-Cullin1-F-box)^{SLF} complex, which is required for cross-pollen compatibility in S-RNase-based self-incompatibility systems (Zhao et al. 2010). Putative methyltransferase PMT27 is a member of the *TUMOROUS SHOOT DEVELOPMENT2* (*TSD2*) gene family in *Arabidopsis* where it has an essential role in cell adhesion and coordinated plant development (Krupková et al. 2007). Late embryogenesis abundant (LEA) proteins are responsive to water deficits, and they are accumulated in dry seeds, before disappearing during germination, so they are relating to both hydration and dehydration (Colmenero-Flores et al. 1997). The latter two proteins may have dual functions in pollen where they promote rapid adhesion to the stigma surface and hydration before germination. S-adenosyl-L-homocysteine hydrolase is a key enzyme during methionine metabolism, and it is known to be a required cofactor as a methyl donor in a wide number of methylation reactions during pollen germination and pollen tube elongation (Ranocha et al. 2001; Moscatelli et al. 2005; Masuko et al. 2006). Based on the aforementioned biological functions in pollen, it can be said that these proteins might be associated with the differential crossability of TD and DT pollen. Further analysis of temporal changes in these proteins after pollination is needed to explore the relationships between their functions and pollen behavior.

Conclusion

Previously, we observed that, irrespective of being crossed as a male or female, the F_1 (namely, TD and DT) and BC_1 progenies always had higher berry-setting rates when they contained T cytoplasm compared with D cytoplasm (average = 2.04 times; Sanetomo et al. 2011). In addition, male and female chromosomal factors were found to be independently involved in successful crosses (Sanetomo et al. 2011). In this study, the differential transcription between TD and DT pollen was represented not by differences in the percentage contributions of transcripts

within a locus, but by differences in the overall transcription levels of a locus. In addition, a broad spectrum of nuclear genes contributed to reciprocal differences, although the composition of nuclear chromosomal genes was similar in TD and DT pollen. Therefore, we suggest that genetic interactions between cytoplasmic genome and the nuclear chromosomal genes contributed greatly to differences in the transcription levels of various genes and phenotypic differences between reciprocal hybrids.

References

Adams KL, Cronn R, Percifield R, Wendel JF (2003) Genes duplicated by polyploidy show unequal contributions to the transcriptome and organ-specific reciprocal silencing. Proc Natl Acad Sci USA 100:4649–4654

Alleman M, Doctor J (2000) Genomic imprinting in plants: observations and evolutionary implications. Plant Mol Biol 43:147–161

Andorf S, Selbig J, Altmann T, Poos K, Witucka-Wall H, Repsilber D (2010) Enriched partial correlations in genome-wide gene expression profiles of hybrids (A. thaliana): a systems biological approach towards the molecular basis of heterosis. Theor Appl Genet 120:249–259

Borges F, Gomes G, Gardner R, Moreno N, McCormick S, Feijó JA, Becker JD (2008) Comparative transcriptomics of *Arabidopsis* sperm cells. Plant Physiol 148:1168–1181

Chung HJ, Jung JD, Park HW, Kim JH, Cha HW, Min SR, Jeong WJ, Liu J (2006) The complete chloroplast genome sequences of *Solanum tuberosum* and comparative analysis with Solanaceae species identified the presence of a 241-bp deletion in cultivated potato chloroplast DNA sequence. Plant Cell Rep 25:1369–1379

Colmenero-Flores JM, Campos F, Garciarrubio A, Covarrubias AA (1997) Characterization of *Phaseolus vulgaris* cDNA clones responsive to water deficit: identification of a novel late embryogenesis abundant-like protein. Plant Mol Biol 35:393–405

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plant Cell 12:1551–1568

del Campo EM (2009) Post-transcriptional control of chloroplast gene expression. Gene Reg Syst Biol 3:31–47

Dilkes BP, Comai L (2004) A differential dosage hypothesis for parental effects in seed development. Plant Cell 16:3174–3180

Forner J, Weber B, Thuss S, Wildum S, Binder S (2007) Mapping of mitochondrial mRNA termini in *Arabidopsis thaliana*: t-elements contribute to 5′ and 3′ end formation. Nucleic Acids Res 35:3676–3692

Gehring M, Choi Y, Fischer RL (2004) Imprinting and seed development. Plant Cell 16:S203–S213

Gehring M, Missirian V, Henikoff S (2011) Genomic analysis of parent-of-origin allelic expression in *Arabidopsis thaliana* seeds. PLoS ONE 6:e23687

Guo M, Rupe MA, Zinselmeier C, Habben J, Bowen BA, Smith OS (2004) Allelic variation of gene expression in maize hybrids. Plant Cell 16:1707–1716

Guo M, Rupe MA, Yang X, Crasta O, Zinselmeier C, Smith OS, Bowen B (2006) Genome-wide transcript analysis of maize hybrids: allelic additive gene expression and yield heterosis. Theor Appl Genet 113:831–845

Guo M, Yang S, Rupe M, Hu B, Bickel DR, Arthur L, Smith O (2008) Genome-wide allele-specific expression analysis using Massively Parallel Signature Sequencing (MPSS^TM) reveals cis- and trans-effects on gene expression in maize hybrid meristem tissue. Plant Mol Biol 66:551–563

Haerizadeh F, Wong CE, Bhalla PL, Gresshoff PM, Singh MB (2009) Genomic expression profiling of mature soybean (*Glycine max*) pollen. BMC Plant Biol 9:25

Hafidh S, Breznenová K, Růžička P, Feciková J, Čapková V, Honys D (2012) Comprehensive analysis of tobacco pollen transcriptome unveils common pathways in polar cell expansion and underlying heterochronic shift during spermatogenesis. BMC Plant Biol 12:24

He G, Zhu X, Elling AA, Chen L, Wang X, Guo L, Liang M, He H, Zhang H, Chen F, Qi Y, Chen R, Deng XW (2010) Global epigenetic and transcriptional trends among two rice subspecies and their reciprocal hybrids. Plant Cell 22:17–33

Honys D, Twell D (2003) Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Physiol 132:640–652

Honys D, Twell D (2004) Transcriptome analysis of haploid male gametophyte development in *Arabidopsis*. Genome Biol 5:R85

Huh J, Bauer M, Hsieh T, Fischer R (2007) Endosperm gene imprinting and seed development. Curr Opin Genet Dev 17:480–485

Jacobsen SE, Meyerowitz EM (1997) Hypermethylated *SUPERMAN* epigenetic alleles in *Arabidopsis*. Science 277:1100–1103

Jones PA, Takai D (2001) The role of DNA methylation in mammalian epigenetics. Science 293:1068–1070

Josefsson C, Dilkes B, Comai L (2006) Parent-dependent loss of gene silencing during interspecies hybridization. Curr Biol 16:1322–1328

Knight JC (2004) Allele-specific gene expression uncovered. Trends Genet 20:113–116

Köhler C, Makarevich G (2006) Epigenetic mechanisms governing seed development in plants. EMBO Rep 7:1223–1227

Köhler C, Scheid OM, Erilova A (2010) The impact of the triploid block on the origin and evolution of polyploidy plants. Trends Genet 26:142–148

Krupková E, Immerzeel P, Pauly M, Schmülling T (2007) The *TUMOROUS SHOOT DEVELOPMENT2* gene of Arabidopsis encoding a putative methyltransferase is required for cell adhesion and co-ordinated plant development. Plant J 50:735–750

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory- efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25

Li N, Dickinson HG (2010) Balance between maternal and paternal alleles sets the timing of resource accumulation in the maize endosperm. Proc R Soc B 277:3–10

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup (2009) The sequence alignment/map (SAM) format and SAMtools. Bioinformatics 25:2078–2079

Luo M, Taylor JM, Spriggs A, Zhang H, Wu X, Russell S, Singh M, Koltunow A (2011) A genome-wide survey of imprinted genes in rice seeds reveals imprinting primarily occurs in the endosperm. PLoS Genet 7:e1002125

Madlung A, Masuelli RW, Watson B, Reynolds SH, Davison J, Comai L (2002) Remodeling of DNA methylation and

phenotypic and transcriptional changes in synthetic Arabidopsis allotetraploids. Plant Physiol 129:733–746

Masuko H, Endo M, Saito H, Hakozaki H, Park JI, Kawagishi-Kobayashi M, Takada Y, Okabe T, Kamada M, Takahashi H, Higashitani A, Watanabe M (2006) Anther-specific genes, which expressed through microsporogenesis, are temporally and spatially regulated in model legume, *Lotus japonicus*. Genes Genet Syst 81:57–62

Matsubayashi M (1991) Phylogenetic relationships in the potato and its related species. In: Tsuchiya T, Gupta PK (eds) Chromosome engineering in plants: genetics, breeding, evolution Part B. Elsevier, Amsterdam, pp 93–118

Milne I, Bayer M, Cardle L, Shaw P, Stephen G, Wright F, Marshall D (2010) Tablet-next generation sequence assembly visualization. Bioinformatics 26:401–402

Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods 5:621–628

Moscatelli A, Scali M, Prescianotto-Baschong C, Ferro M, Garin J, Vignani R, Ciampolini F, Cresti M (2005) A methionine synthase homolog is associated with secretory vesicles in tobacco pollen tubes. Planta 221:776–789

Pennington PD, Costa LM, Gutierrez-Marcos JF, Greenland AJ, Dickinson HG (2008) When genomes collide: aberrant seed development following maize interploidy crosses. Ann Bot 101:833–843

Pina C, Pinto F, Feijó JA, Becker JD (2005) Gene family analysis of the Arabidopsis pollen transcriptome reveals biological implications for cell growth, division control, and gene expression regulation. Plant Physiol 138:744–756

Ranocha P, McNeil SD, Ziemak MJ, Li CJ, Tarczynski MC, Hanson AD (2001) The *S*-methylmethionine cycle in angiosperms: ubiquity, antiquity and activity. Plant J 25:575–584

Riddle NC, Jiang H, An L, Doerge RW, Birchler JA (2010) Gene expression analysis at the intersection of ploidy and hybridity in maize. Theor Appl Genet 120:341–353

Roach DA, Wulff RD (1987) Maternal effects in plants. Annu Rev Ecol Syst 18:209–235

Sanetomo R, Hosaka K (2011) Reciprocal differences in DNA sequence and methylation status of the pollen DNA between F_1 hybrids of *Solanum tuberosum* × *S. demissum*. Euphytica 182:219–229

Sanetomo R, Ono S, Hosaka K (2011) Characterization of crossability in the crosses between *Solanum demissum* and *S. tuberosum*, and the F_1 and BC_1 progenies. Am J Potato Res 88:500–510

Swanson-Wagner RA, Jia Y, DeCook R, Borsuk LA, Nettleton D, Schnable PS (2006) All possible modes of gene action are observed in a global comparison of gene expression in a maize F_1 hybrid and its inbred parents. Proc Natl Acad Sci USA 103:6805–6810

Tebbji F, Nantel A, Matton DP (2010) Transcription profiling of fertilization and early seed development events in a solanaceous species using a 7.7 K cDNA microarray from *Solanum chacoense* ovules. BMC Plant Biol 10:174

The Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–195

Vaughn MW, Tanurdžić M, Lippman Z, Jiang H, Carrasquillo R, Rabinowicz PD, Dedhia N, McCombie WR, Agier N, Bulski A, Colot V, Doerge RW, Martienssen RA (2007) Epigenetic natural variation in *Arabidopsis thaliana*. PLoS Biol 5:e174

Vuylsteke M, van Eeuwijk F, Hummelen PV, Kuiper M, Zabeau M (2005) Genetic analysis of variation in gene expression in *Arabidopsis thaliana*. Genetics 171:1267–1275

Walia H, Josefsson C, Dilkes B, Kirkbride R, Harada J, Comai L (2009) Dosage-dependent deregulation of an AGAMOUS-LIKE gene cluster contributes to interspecific incompatibility. Curr Biol 19:1128–1132

Wang Y, Zhang WZ, Song LF, Zou JJ, Su Z, Wu WH (2008) Transcriptome analyses show changes in gene expression to accompany pollen germination and tube growth in Arabidopsis. Plant Physiol 148:1201–1211

Wei G, Tao Y, Liu G, Chen C, Luo R, Xia H, Gan Q, Zeng H, Lu Z, Han Y, Li X, Song G, Zhai H, Peng Y, Li D, Xu H, Wei X, Cao M, Deng H, Xin Y, Fu X, Yuan L, Yu J, Zhu Z, Zhu L (2009) A transcriptomic analysis of superhybrid rice *LYP9* and its parents. Proc Natl Acad Sci USA 106:7695–7701

Willing RP, Mascarenhas JP (1984) Analysis of the complexity and diversity of mRNAs from pollen and shoots of Tradescantia. Plant Physiol 175:865–868

Willing RP, Bashe D, Mascarenhas JP (1988) An analysis of the quantity and diversity of messenger RNAs from pollen and shoots of *Zea mays*. Theor Appl Genet 75:751–753

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829

Zhao L, Huang J, Zhao Z, Li Q, Sims TL, Xue Y (2010) The Skp1-like protein SSK1 is required for cross-pollen compatibility in *S*-RNase-based self-incompatibility. Plant J 62:52–63

Zhuang Y, Adams KL (2007) Extensive allelic variation in gene expression in Populous F_1 hybrids. Genetics 177:1987–1996