Protein-coding genes are epigenetically regulated in *Arabidopsis* polyploids

Hyeon-Se Lee and Z. Jeffrey Chen*

Genetics Program and Department of Soil and Crop Sciences, Texas A&M University, College Station, TX 77843-2474

Edited by S. J. Peloquin, University of Wisconsin, Madison, WI, and approved April 9, 2001 (received for review February 7, 2001)

The fate of redundant genes resulting from genome duplication is poorly understood. Previous studies indicated that ribosomal RNA genes from one parental origin are epigenetically silenced during interspecific hybridization or polyploidization. Regulatory mechanisms for protein-coding genes in polyploid genomes are unknown, partly because of difficulty in studying expression patterns of homologous genes. Here we apply amplified fragment length polymorphism (AFLP)–cDNA display to perform a genome-wide screen for orthologous genes silenced in *Arabidopsis suecica*, an allotetraploid derived from *Arabidopsis thaliana* and *Cardaminopsis arenosa*. We identified ten genes that are silenced from either *A. thaliana* or *C. arenosa* origin in *A. suecica* and located in four of the five *A. thaliana* chromosomes. These genes represent a variety of RNA and predicted proteins including four transcription factors such as TCP3. The silenced genes in the vicinity of *TCP3* are hypermethylated and reactivated by blocking DNA methylation, suggesting epigenetic regulation is involved in the expression of orthologous genes in polyploid genomes. Compared with classic genetic mutations, epigenetic regulation may be advantageous for selection and adaptation of polyploid species during evolution and development.

genome duplication | epigenetics | gene silencing | evolution

Polyploidy results from duplication of a whole genome (autopolyploid) or from combining two or more distinct but related genomes (allopolyploid). It occurs in many organisms, but predominantly in vertebrates and plants (1, 2), and estimates indicate that over 70% of flowering plants had at least one event of polyploidization in their evolutionary lineage (1, 3–7). Many important crops, including banana, canola, coffee, cotton, maize, potato, oat, soybean, sugarcane, and wheat, are polyploid (8). The common occurrence of polyploids in nature probably reflects an evolutionary advantage of having redundant genes, freeing some gene copies from certain constraints of natural selection, to allow accumulation of new mutations that improve fitness. Consistent with this notion, polyploid plants are more widely distributed over more habitats than their diploid progenitors (1).

The most common form of polyploids is allopolyploidy, in which two or more different but related (homoeologous) genomes are brought together in a single-cell nucleus. It is unclear how the expression of homologous genes is regulated in the hybrid cell. The vast majority of genes retain their function during evolution (7, 9–12), whereas some duplicate genes are mutated or silenced within a few million years (13). However, ''genomic shock'' as predicted by McClintock (14) occurs rapidly, resulting in sequence elimination and rearrangement (15, 16), demethylation of retroelements (17, 18), and relaxation of imprinting genes (19) in polyploid genomes. Furthermore, in interspecific hybrids or allopolyploids of vertebrates, invertebrates, and plants, one parental set of rRNA genes is subjected to silencing (20, 21). The genes are stochastically silenced within two generations (22) after polyploid formation and reactivated by blocking DNA methylation and histone deacetylation (23) and in flower tissues and organs (24), suggesting roles of developmental programs and chromatin modification in the expression of orthologous genes.

Arabidopsis suecica ($2n = 4x = 26$) is a natural allotetraploid that is formed through interspecific hybridization between *Arabidopsis thaliana* ($2n = 2x = 10$) and *Cardaminopsis arenosa* ($2n = 4x = 32$) (25). Although *A. thaliana* and *C. arenosa* are classified in different genera, they are very closely related. Based on systematic and DNA sequence analyses, O'Kane *et al.* proposed to reclassify *C. arenosa* as *Arabidopsis arenosa* (25). Moreover, the orthologous genomes of *C. arenosa* and *A. thaliana* share 90–100% homology between the coding sequences studied (H.-S.L. and Z.J.C., unpublished observation). *C. arenosa* is an open pollinated autotetraploid; thus, it is likely that natural *A. suecica* lines are formed through pollination of *C. arenosa* with 2n gametes from diploid *A. thaliana*. Moreover, synthetic *A. suecica* hybrids can be made by crossing *C. arenosa* with *A. thaliana* autotetraploids (22). The natural strains of *A. suecica* have 13 pairs of chromosomes and behave as a disomic polyploid or functional diploid (1) during meiosis (pairing occurs only between homologous chromosomes).

Study of gene expression in polyploid genomes is impeded by the high degree of homology of the duplicate genes derived from orthologous genomes. Here we investigate regulation of protein-coding genes in polyploid genomes by using amplified fragment length polymorphism (AFLP)–cDNA analysis (26). The technique employs use of restriction polymorphisms present in orthologous genes to discriminate transcripts among polyploid and its progenitor species. We have successfully identified a variety of genes that are differentially expressed in *A. suecica*, some of which are epigenetically silenced and maintained by hypermethylation. Blocking cytosine methylation by 5-aza-2′-deoxycytidine (aza-dC), an inhibitor of DNA methyltransferases, induces genomic demethylation and reactivates silenced genes. Thus, epigenetic regulation plays an important role in the expression of the progenitor's genes during polyploid formation and speciation.

Materials and Methods

Plant Materials. *A. suecica* and *A. thaliana* (Landsberg strain) were self-pollinated for at least four generations after they were obtained from original collections (25). *C. arenosa* (pink flower) was maintained by cross-pollination among a few plants in each generation. Autotetraploid *A. thaliana* lines (Columbia) and (Landsberg) were obtained as described (22). The plants were grown in vermiculite mixed with 10% soil in an environment-controlled chamber with 14/10 h (day/night) light and 24/20°C (day/night) temperature. Treatment of *A. suecica* seedlings by using aza-dC was adapted from a published protocol (23).

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: AFLP, amplified fragment length polymorphism; RT-PCR, reverse transcriptase–PCR; aza-dC, 5-aza-2′-deoxycytidine.

*To whom reprint requests should be addressed. E-mail: zjchen@acs.tamu.edu.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

GENETICS

DNA and RNA Analysis. Total genomic DNA and RNA isolation and DNA blot analysis were performed as described (22). AFLP-cDNA display was modified from a described method (26). Messenger RNA (≈2 μg) was purified by using biotinylated-oligo(dT) and streptavidin-paramagnetic particles (Promega). The mRNA was eluted in water for cDNA synthesis using AMV reverse transcriptase and RNase H and DNA Polymerase I (GIBCO/BRL). AFLP-cDNA display was performed according to the manufacturer's recommendations (GIBCO/BRL). In brief, the double-stranded cDNA (≈500 ng) was digested with *Eco*RI and *Mse*I and ligated to respective oligo primers (adaptors). The ligated products were subjected to preselective amplification (20 cycles) by using primers with one restrictive base at the 3′ end of the primers and selective amplification by using primers with two restrictive bases at the 3′ end. One of the primers was end-labeled with [γ-^{32}P]dATP. The generic primers are *Eco*RI: 5′-GACTGCGTACCAATTCNN-3′ and *Mse*I: 5′-GATGATTCCTGAGTAANNN-3′. The PCR cycles used for selective amplification consist of 12 cycles of 94°C for 30 s, 65°C for 30 s, and 72°C for 1 min and 20 cycles of 94°C for 30 s, 56°C for 30 s, and 72°C for 1 min. The annealing temperature between the first and 13th cycles was linked by a touch-down phase of decreasing 0.7° in each cycle. After amplification, the products were resolved in a 6% polyacrylamide gel as described (23).

To determine the identity of the cDNA fragments, the fragments of interest were excised and eluted in TE (10 mM Tris/1 mM EDTA, pH 8.0) and reamplified by using the same primer pairs as in display analysis. Sequencing reactions were performed by using the purified PCR products as template in the dideoxy chain termination method (27) in an Applied Biosystems 377 sequencer or manually. For duplicate alleles, individual cDNA fragments were cloned in pGEM vector (Promega) and sequenced. About 10–20 inserts from each gene were sequenced from each locus. Sequence alignment was analyzed by using DNASTAR or CLUSTAL W software. Reverse transcriptase (RT)–PCR was performed according to the SuperScript One-Step RT-PCR kit (GIBCO/BRL) with 40 cycles of 94°C for 15 s, 50–65°C for 30 s, and 72°C for 1 min. Four primer pairs for the genes of interest were listed below: (*i*) *RFP*, forward 5′-GGCCAGAATGCCACATTATATC-3′ and reverse 5′-CTAGACCAGTGCCTCTGATTTAC-3′; (*ii*) *TCP3*, forward 5′-GGTCCACCTTTTCCTAATCAAAC-3′ and reverse 5′-TAGCTTCAAGTGGGGTTAAAGGT-3′; (*iii*) *HYP1*, forward 5′-AGAAGATCCGGATTGCTACAGA-3′ and reverse 5′-TAACCGAACTCTCGTCCGTAGAT-3′; and (*iv*) *Mdh*, forward 5′-GTTGGTCACATCAACACCAGAT-3′ and reverse 5′-AAAGTCTGAGAGTGGTCCCAAGT-3′. The amplified products were analyzed by using cleaved amplified polymorphic sequence (CAPS; ref. 28).

Results

Identification of Silenced Genes in *Arabidopsis* Polyploids by AFLP-cDNA Display. To discriminate between orthologous transcripts in polyploid genomes, we applied AFLP-cDNA technology (26). This technique employs the use of restriction polymorphism in analyzing cDNA samples. After the cDNA fragments are restricted, they are ligated onto primers with compatible restriction ends. Ligated products can be amplified by PCR and resolved in a sequencing gel. Any fragments that are present in one sample but not the other are isolated and sequenced. The AFLP-cDNA display procedure was optimized to produce constant results in a diploid *A. thaliana* and its isogenic autotetraploid ($2n = 4x = 20$) line. Using a given primer combination in four separate experiments, we were able to obtain consistent results between the two lines in each experiment (Fig. 1*a*). As expected, the majority of genes in diploid and autotetraploid lines were expressed at similar levels (Fig. 1*a*). However, differential expression patterns of a few genes were observed and verified by RT-PCR (H.-S.L. and Z.J.C., unpublished results). After optimization, we applied the technique to *A. suecica* and its diploid progenitors, *A. thaliana* and *C. arenosa*. A typical display analysis resulted in three expression patterns among the allotetraploid and its progenitors (Fig. 1*b*). First, no restriction polymorphism was detected between the same-size alleles (open circles). Second, some alleles displayed clear polymorphism and coexpression patterns. For example, sequence analysis of two individual alleles (arrows) confirmed that the orthologous alleles of the gene encoding acylthioesterase (GenBank accession no. Z36911) from *A. thaliana* and *C. arenosa* were coexpressed in *A. suecica* (data not shown). Third, some alleles exhibited polymorphism, but no obvious patterns of coexpression (open squares). Of the 4,428 cDNA fragments assayed, 450 (11%) cDNAs fell into this category. We further sequenced and ana-

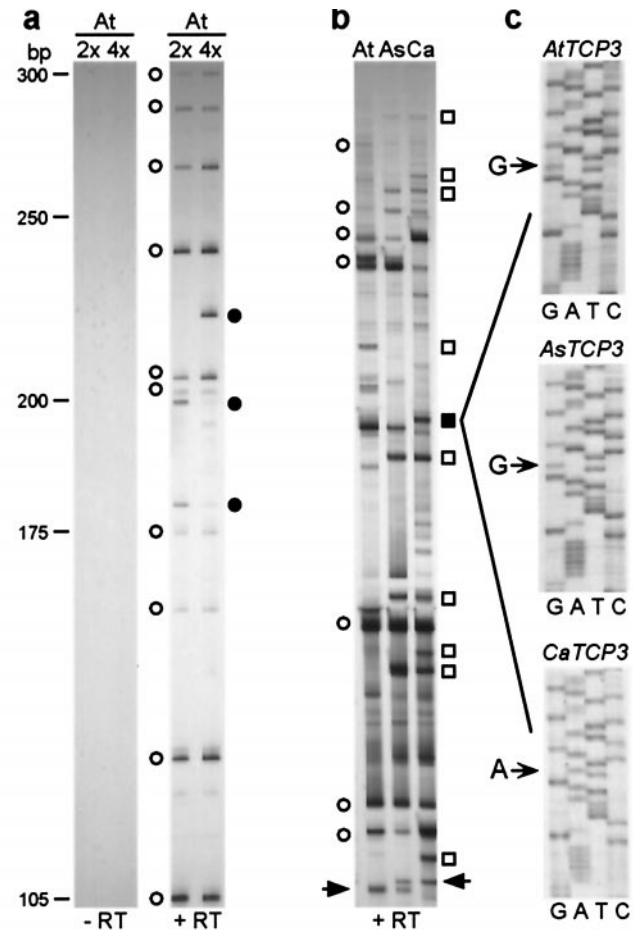

Fig. 1. (*a*) AFLP-cDNA display of gene expression in diploid (2x) and autotetraploid (4x) *Arabidopsis*. The cDNA was synthesized with or without reverse transcriptase (+RT or −RT), respectively, and subjected to AFLP analysis. The primer pairs used in PCR were *Eco*RI-TA and *Mse*I-CAT. Circles indicate the cDNA fragments present in both diploid and autotetraploid (open) and either the diploid or autotetraploid (filled). (*b*) AFLP-cDNA analysis in an allotetraploid, *A. suecica* (As), and its two diploid progenitors, *A. thaliana* (At) and *C. arenosa* (Ca). The primer pairs used in PCR were *Eco*RI-AC and *Mse*I-CAC. Symbols indicate nonpolymorphic cDNA fragments (open circles), coexpression of orthologous alleles (arrows), and the cDNA alleles present only in one of the diploids and/or allotetraploid (open squares). Three alleles from one locus (filled square) were cloned and sequenced. (*c*) Sequencing gel results show a region of *TCP3* from *A. thaliana* (*AtTCP3*), *A. suecica* (*AsTCP3*), and *C. arenosa* (*CaTCP3*). A base transition (G to A) was identified among the orthologous alleles (indicated by arrows). The size markers for *a* and *b* are shown to the left.

Table 1. A set of *A. thaliana* (A.t.) or *C. arenosa* (C.a.) genes is differentially expressed in *A. suecica*

Clone	Predicted transcript or protein	Sequence identity, %*	Parental origin determined by AFLP-cDNA and sequencing	Chromosomal location in *A. thaliana*
A122	TCP3 family	100	A.t.	1
L113	Transcription factor IIB	100	C.a.	4
S822	NAC protein	97	C.a.	5
S823	Bell homeodomain 2	100	A.t.	4
S882	Putative transposase	100	A.t.	1
S862	Untranslated RNA	97	C.a.	4
S813	ATP/GTP nucleotide-binding protein	100	A.t.	5
S842	Glutathione-regulated potassium eflux system protein	100	A.t.	4
S814	Putative protein	87	C.a.	3
S815	Unknown	98	A.t.	4

*The percentage of sequence identity is determined by alignment analyses using the sequences (in a range of 150 to 300 bp) of the cloned cDNA fragments and orthologous genes and/or expressed sequence tags in the AtDB or GenBank.

lyzed individual fragments from these alleles. Sequencing results for three cDNA fragments indicated that they were orthologous alleles encoding *TCP3* (29, 30). The single nucleotide substitution (Fig. 1*c*) allows us to determine that the expressed allele in *A. suecica* was from *A. thaliana*. Twenty-five (22.7%) of the 110 cDNA fragments that were sequenced exhibited differential expression patterns of parental genes in the allotetraploid *A. suecica*. The remaining 85 cDNAs were coexpressed as determined by RT-PCR analysis (data not shown), although they exhibited restriction polymorphism between the orthologous alleles.

We identified and confirmed a set of ten different genes that were differentially expressed in *A. suecica* and its progenitors. Based on predicted protein sequences, the functions of the ten genes fell into several categories (Table 1), including four transcriptional factors for RNA Polymerase II, a putative transposase, an untranslated RNA, and an unknown gene. The gene was classified as unknown because it is located in an intron region based on current annotation. However, the transcripts were detected by cDNA display and matched ESTs in the databases. We did not observe genes that were expressed only in the allopolyploid but not in the diploids. The data suggest that genes involved in various biological functions are subjected to differential expression in an allotetraploid, contrary to the notion that silenced genes in polyploids contain mainly transposons (17, 18).

Expression Patterns of Progenitor Genes in the *TCP3* Vicinity. We discovered that silencing involves genes encoding several classes of transcriptional factors, including Bell homoeodomain 2, NAC protein, transcription factor IIB, and TCP3 family (29). TCP3 [*T*eosinte branched 1, *C*ycloidea 1, and proliferating cell nuclear antigen (PCNA) factors, *P*CF1 and 2] are a newly discovered family of transcriptional factors that share a common motif in DNA binding domains and bHLH structure, presumably involved in the activation of PCNA during DNA replication and cell division (29). Genetic mutation of *Teosinte Branched 1* causes axillary meristem development in maize (31), resulting in

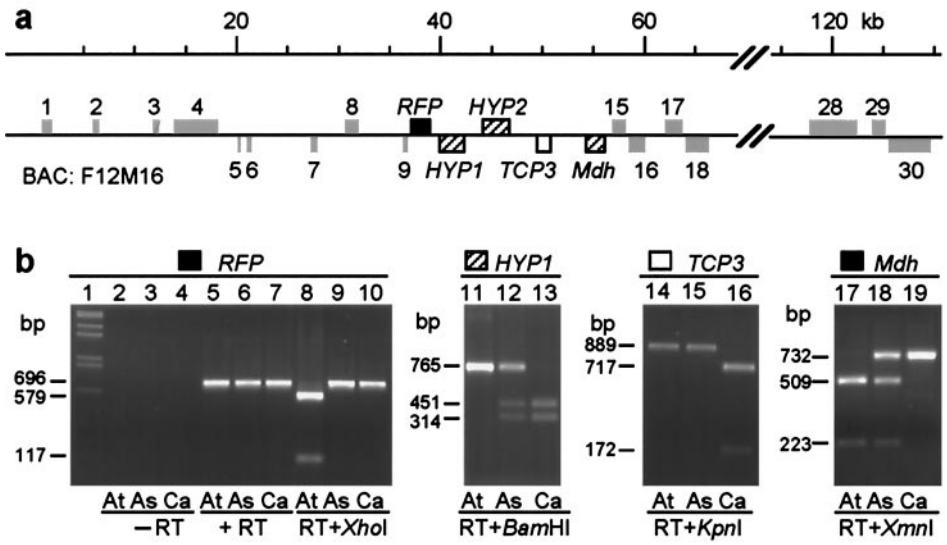

Fig. 2. Differential expression patterns of parental genes in *A. suecica*. (*a*) Diagram of a *TCP3*-containing BAC clone (F12 M16). The BAC contains 30 predicted genes (gray boxes), some of which are not shown and indicated by double-slash lines. Five genes in the vicinity of *TCP3* include *RFP*, *HYP1*, *HYP2*, *TCP3*, and *MDH*. The boxes indicate that expression patterns of parental genes in *A. suecica* are from *C. arenosa* only (black), *A. thaliana* only (open), or both (slash lines) origins (see *b*). (*b*) Parental gene expression patterns in *A. suecica* and its diploid progenitors. Agarose gels show RT-PCR results of four genes in *A. thaliana* (At), *A. suecica* (As), and *C. arenosa* (Ca). For every gene (e.g., *RFP*) the amplification was performed by addition of mock (lanes 2–4) or reverse transcriptase (lanes 5–10). The RT-PCR products of the *RFP* were then digested with *Xho*I (lanes 8–10). The products were resolved in a 1.5% agarose gel. The RT-PCR products amplified from the *HYP1* (lanes 11–13), *TCP3* (lanes 14–16), and *MDH* (lanes 17–19) were digested with *Bam*HI, *Kpn*I, and *Hin*dIII, respectively. The controls omitted for *HYP1*, *TCP3*, and *MDH* are RT-PCR without the reverse transcriptase and RT-PCR products without restriction digestion. *Eco*RI- and *Hin*dIII-digested lambda DNA is shown in lane 1 as size markers.

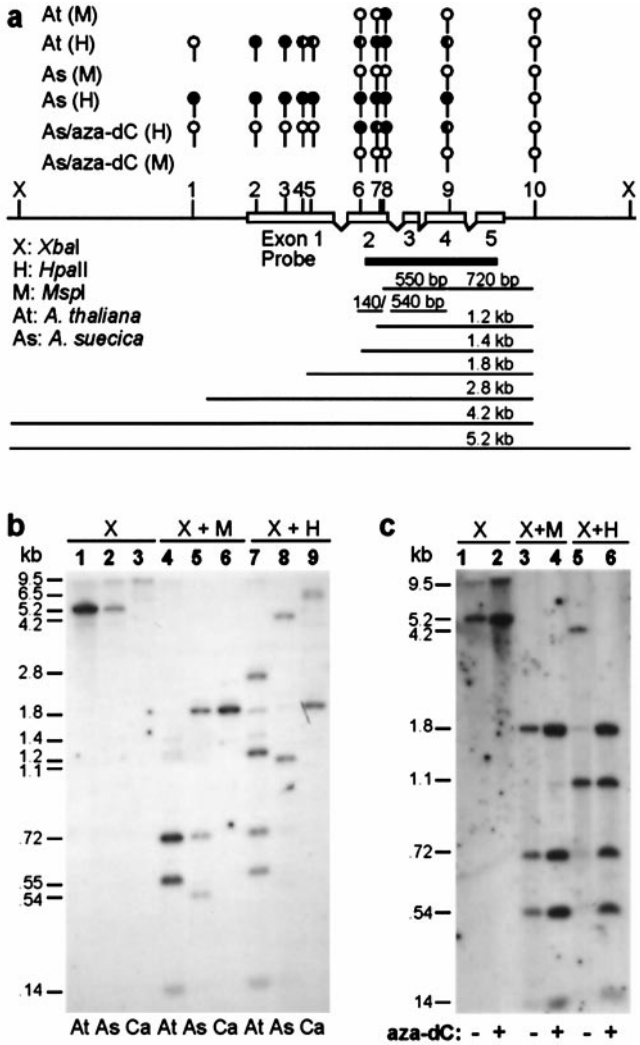

Fig. 3. Silenced *RFP* is hypermethylated and demethylated by aza-dC in *A. suecica*. (*a*) Restriction map of the *RFP* genomic sequence (*Xba*I fragment) showing ten *Msp*I/*Hpa*II sites (labeled from 1 to 10) from the start codon to the end of the coding sequence. Five exons (open boxes) are joined by four introns (stippled lines). The probe used for DNA blot analysis is shown below the diagram. Sizes of the fragments resulting from complete digestion with *Xba*I (X) and *Msp*I (X + M) or *Hap*II (X + H) are shown. Lollipops indicate complete (filled), partial (half-filled), or no methylation (open) at the sites shown. (*b*) Genomic DNA isolated from young leaves of *A. suecica* (As) and its diploid progenitors, *A. thaliana* (At) and *C. arenosa* (Ca), was digested with *Xba*I (lanes 1–3). Each sample was divided equally into three parts; two of the samples were digested again with another restriction enzyme, *Msp*I or *Hap*II. The DNA samples were then separated in agarose gel, blotted, and hybridized with the probe indicated above. *A. suecica* has two *Xba*I fragments (lane 2), one from *A. thaliana* (lane 1) and another from *arenosa* (lane 3). Eight of nine *Hpa*II sites are partially methylated in *A. thaliana* (compare lanes 4 and 7), although completely methylated in *A. suecica* (compare lanes 8 and 7). Only a few *Hpa*II/*Msp*I sites were detected in the *C. arenosa* gene (lanes 6 and 9). (*c*) Demethylation of the *AtRFP* genomic sequence after aza-dC treatment. Genomic DNA was isolated from seedlings of the *A. suecica* treated with (+) or without (−) addition of aza-dC (10 mg/L), a chemical inhibitor of DNA methyltransferases. The DNA was digested by *Xba*I and *Msp*I or *Hpa*II, as mentioned, and subjected to DNA blot analysis by using the same probe.

a stem structure like that of its ancestor, teosinte. DNA-methylation-dependent epigenetic mutation in *Cycloidea 1* (*CYC1*) is correlated with asymmetrical flower development in *Antirrhinum* (30), suggesting an important role of epigenetic regulation of gene expression in plant morphology and evolution.

To gain a molecular understanding of chromosomal organization and gene silencing in polyploids, we analyzed genome structure of *TCP3*. There are two related *TCP3* genes in *Arabidopsis*, one in chromosome 1, another in chromosome 3 (29). Two other ESTs with low homology were also identified in chromosomes 2 and 4, respectively. The cDNA fragment we cloned and sequenced was *A. thaliana TCP3* (*AtTCP3*) in chromosome 1, which has the highest homology to other members in the family (29). *AtTCP3* is located in the BAC (F12 M16) that has been sequenced and annotated (32). The BAC has a 130,235-bp insert containing 30 genes (Fig. 2*a*), 11 of them on the top strand and 19 on the bottom. *TCP3* is located on the bottom strand in the middle of the BAC insert. Five genes in the vicinity of *TCP3*, including three genes on the bottom strand and two genes on the top strand, were further analyzed. Ring finger protein (RFP) is a large family of transcription factors involved in various pathways of gene regulation. One member of this family is COP1, a protein that interacts antagonistically with HY5, which in turn binds to the promoters of light-inducible genes (33). The other three genes in the vicinity of *TCP3* encode two hypothetical proteins (HYP1 and HYP2) and a mitochondrial NAD-dependent malate dehydrogenase (MDH), respectively.

The expression profiles of genes in the vicinity of *TCP3* are shown in Fig. 2*b*. By using gene-specific primers in RT-PCR assays, the *RFP* was amplified in all three lines dependent on reverse transcription (Fig. 2*b*, compare lanes 2–4 and 5–7). To discriminate between the progenitor transcripts, the amplified cDNA fragments were analyzed by using cleaved amplified polymorphic sequence (CAPS) analysis (28). The *A. thaliana RFP* was cleaved by *Xho*I (lane 8), whereas the *C. arenosa RFP* fragment was uncut (lane 10). *A. suecica*, like *C. arenosa*, possessed only the uncut fragment (lane 9), suggesting that the *A. thaliana RFP* was not expressed in *A. suecica*. A trivial explanation is the mutation or loss of the *A. thaliana RFP* in *A. suecica*. Results obtained from DNA blot analysis (see below) and PCR amplification using genomic template DNA (data not shown) ruled out this possibility.

Our sequencing results indicated that *AtTCP3* was expressed in *A. suecica*, whereas *CaTCP3* transcripts were absent (Fig. 1*c*). This result was confirmed by RT-PCR analysis. Based on *Kpn*I-digested patterns of the cDNA fragments, only *A. thaliana*-like transcripts were present in *A. suecica* (Fig. 2*b*, lanes 14–16). A trivial explanation is heterozygosity of the *CaTCP3* alleles in *A. suecica*, because natural *C. arenosa* is an out-crossing autotetraploid. However, reactivation of the silenced *CaTCP3* in *A. suecica* (see below) ruled out his possibility.

Our data indicate that the silencing direction of *RFP* is different from *TCP3*, even though the two genes are closely located and separated by only two other genes, *HYP1* and *HYP2*. RT-PCR analysis indicated that *HYP1* along with *HYP2* (data not shown) from both parental origins were expressed in *A. suecica* (lanes 11 and 12). In addition, *MDH*, located upstream of *TCP3*, was coexpressed in *A. suecica* (lanes 17–19), although the adjacent *TCP3* is expressed from *A. thaliana* origin. The data suggest that expression patterns of parental genes in polyploid genomes are complicated even in a small chromosomal domain.

Methylation, Gene Silencing, and Reactivation of Silenced Genes. It has been shown that the silenced rRNA genes in an allotetraploid are maintained by DNA methylation and histone deacetylation (23). Because chromatin modification is a general mechanism for gene regulation, we predict that silenced genes are associated with DNA methylation. To test this hypothesis, we examined the DNA methylation status of the *RFP* in *A. suecica* and its two progenitor species. There are 10 *Hpa*II/*Msp*I (CCGG) sites in the promoter and coding sequences of *AtRFP* (Fig. 3*a*). The locus-specific methylation status was examined by

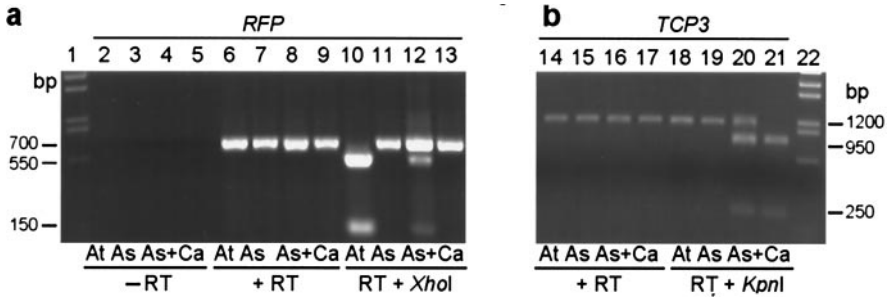

Fig. 4. Reactivation of the silenced *RFP* and *TCP3* in *A. suecica*. (*a*) RNA was isolated from young leaves of *A. thaliana* (At), *C. arenosa* (Ca), and *A. suecica* treated with (As+) or without (As) addition of 10 mg/l aza-dC. The mRNA was subjected to RT-PCR and cleaved amplified polymorphic sequence (CAPS) analyses (28). Negative controls (−RT, lanes 2–4) indicate PCR amplification is completely dependent on reverse transcriptase. The amplified products (lanes 6–9) were digested by *Xho*I (lanes 10–13). The *thaliana*-like *RFP* transcripts were absent in *A. suecica* before aza-dC treatment (lane 11), but reactivated after the treatment (lane 12). (*b*) The silenced *C. arenosa-TCP3* is reactivated by aza-dC. The RT-PCR products were amplified by using *TCP3*-specific primers (lanes 14–17) and digested with *Kpn*I (lanes 18–21). The *C. arenosa*-like *TCP3* transcripts were absent in *A. suecica* before aza-dC treatment (lane 19), but reactivated after the treatment (lane 20). *Eco*RI- and *Hin*dIII-digested lambda DNA is shown lanes 1 and 22 as a size marker.

using a double digestion of the genomic DNA with *Xba*I and *Hpa*II or *Msp*I. *Hpa*II and *Msp*I are isoschizimers that recognize CCGG sites. However, *Hpa*II will not cut if the inner cytosine is methylated. As predicted, *Xba*I digestion alone produced a 5.2-kb fragment in *A. thaliana* (Fig. 3*b*, lane 1) and a 9.5-kb fragment in *C. arenosa* (lane 3). Both fragments were present in *A. suecica* (lane 2), ruling out a possibility of gene loss during polyploid evolution. After the second digestion with *Msp*I, three small fragments were detected in *A. thaliana* (lane 4). A single fragment detected in *C. arenosa* (lane 6) indicates fewer *Msp*I sites present in *C. arenosa* than in *A. thaliana*. Again, *A. suecica-RFP* profiles matched a combined pattern of the *A. thaliana* and *C. arenosa* genes. A 540-bp, instead of 550-bp, fragment was detected in *A. suecica* (lane 5), suggesting that one site is accessible to *Msp*I but not *Hpa*II, because of demethylation in the allotetraploid. *Hpa*II-digestion generated a series of fragments in diploids, indicating that *RFP* is partially methylated in *A. thaliana* (compare lanes 4 and 7) and *C. arenosa* (compare lanes 6 and 9). The 4.2-kb fragment detected in *A. suecica* (lane 8) is derived from *A. thaliana* (see below), suggesting that, except for the last *Hpa*II site, *AtRFP* is completely methylated in *A. suecica*. The *C. arenosa-RFP*, however, was less methylated (lane 8). Moreover, a 1.1-kb fragment (lane 8) detected in *A. suecica* is derived from the *C. arenosa*. A simple explanation is that some CCGG sites of the *C. arenosa* locus are demethylated in *A. suecica* compared with the orthologous loci in the diploid progenitors.

To further investigate whether methylation is correlated with *RFP* activity, we examined *RFP* methylation patterns in *A. suecica* treated with aza-dC, a chemical inhibitor for DNA methyltransferases (Fig. 3*c*). After the treatment, methylation of the *Hap*II sites in the *thaliana-RFP* locus was eliminated, resulting in several small fragments, as observed after *Msp*I-digestion (compare lanes 3–6).

As a result of demethylation by aza-dC, the silenced *A. thaliana-RFP* was reactivated in *A. suecica* (Fig. 4*a*, lane 12). Moreover, the silenced *TCP3* was also reactivated by aza-dC (Fig. 4*b*, lane 20), and associated with hypermethylation (data not shown). Parental origins of the silenced genes in the two loci are different. In *RFP* locus, the *A. thaliana*-like allele is silenced, whereas in *TCP3* the *C. arenosa* allele is silenced. The above data indicate that one of the orthologous alleles of *RFP* or *TCP3* is methylated and silenced in *A. suecica* allotetraploid, whereas the active genes are hypomethylated and expressed. We conclude that the reactivation of the silenced genes is dependent on demethylation (Fig. 3), but independent of parental origin.

Discussion

Epigenetic Regulation and Polyploid Genome Evolution. AFLP-cDNA display is a powerful tool to screen for parental genes that are differentially expressed in *A. suecica*, an allotetraploid derived from *A. thaliana* and *C. arenosa*. Compared with DNA microarray, this method is relatively inexpensive and requires no prior sequence information. Using this method, we estimated that about 2.5% of genes (25 of 110 cDNA fragments sequenced that represent 11% of displayed cDNA) were subjected to silencing. The number may be underestimated because orthologous alleles without the *Eco*RI/*Mse*I polymorphism were not examined.

When two different genomes are combined into a single cell, they must respond to the consequence of genome duplication, especially multiple copies of genes with similar or redundant functions. The majority of orthologous genes are coexpressed, whereas some are silenced. There are two possible models to explain gene silencing in polyploid genomes. First, genes are silenced or lost because of mutations (including deletions, insertions, and rearrangements) of the DNA sequences during evolution (6, 7). Indeed, many isozyme loci are lost or down-regulated during polyploidization (34–37). It was estimated that in the salmonid and cyprinid fish, the loss of duplicate isozyme loci could be as high as 35–65%, suggesting that loss of duplicate gene function is common after polyploidization (10, 13), which occurred 50 million years ago. Second, expression of a progenitor's genes is epigenetically controlled. We hypothesize that during interspecific hybridization or the early process of polyploidization, epigenetic mechanisms reprogram expression of orthologous genes in biological pathways so that a polyploid cell can adjust properly during development. Compared with genetic mutations, epigenetic regulation provides an effective and flexible means for the cell to respond to genome and gene duplication or "genomic shock" (14), because it is established or erased relatively easily (38, 39). For example, one parental set of rRNA genes can be subject to silencing during vegetative growth by chromatin modification and then reactivated during flower development (24). Although the role of developmental regulation in silencing duplicate genes is speculative, this type of regulation has advantages for selection and adaptation of the newly formed polyploid plants. For instance, there is evidence that some protein heterodimers may not function as well as homodimers or vice versa (40, 41). Thus, a silencing strategy could balance the advantage and disadvantage of having multiple copies of orthologous genes or gene products (e.g., transcriptional factors) in a polyploid cell. It is notable that many of the genes we identified encode transcriptional factors important

for many cellular processes and morphological development. The epigenetic regulation of heterologous protein production might compete for fitness under various environmental cues or physiological and developmental conditions. It is likely that some of the silenced genes may be needed at a certain time during plant development. The flexibility of epigenetic regulation would allow polyploid genomes to "flip" some genes on and off in response to environmental cues and developmental changes. Indeed, the reactivation of silenced rRNA genes in allotetraploid *Brassica napus* during flower development (24) may be a signal for rapid cell division and protein synthesis needed for flowering. The epimutation in *CYC1* causes heritable mutation for developing dorsoventral asymmetry flowers in *Antirrhinum*, which is one of the old flower mutations found in *Linaria vulgaris* described by Linnaeus (30). It is notable that *TCP3*, a homolog of *CYC1*, from *C. arenosa* origin is silenced by hypermethylation in the allotetraploid, suggesting that epigenetic control is involved in the formation of plant form and function.

Epigenetic Regulation in Plants and X-Chromosome Inactivation and Imprinting in Mammals. Silencing of endogenous genes by chromatin modification during polyploid formation is reminiscent of X chromosomes inactivation in mammals, in which almost all of the genes from one of the two X chromosomes are silenced. Silencing is initiated by untranslated RNA at the locus of *Xist* (42) and *Tsix* (43) (transcribed in opposite directions) and maintained by DNA methylation and histone modifications. Apparently, the silenced X chromosome can be reactivated during reproductive stages, while every egg carries a viable X chromosome. The choice of which X chromosome is activated is determined by imprinting of the *Xist* from maternal origin in early embryo proper (43, 44). In the epiblast lineage, however, an X chromosome is randomly silenced. The silencing schemes observed in polyploid genomes are distinct from X chromosome inactivation, however. First, silencing does not occur at every locus in a particular chromosome or even in a small chromosome segment. Second, although many genes are coexpressed in the duplicate genomes, only 2.5% of genes we studied are subjected to differential expression. Third, although the choice of selecting which set of rRNA genes to silence is not random, imprinting was not involved in silencing the uniparental set of rRNA genes (24). Finally, the silencing mechanism itself may be different. If chromatin is involved, the chromatin state of every gene needs to be remodeled. How and why the polyploid genomes choose to "flip" on and off the expression of some parental genes after polyploid formation remains an interesting topic in polyploid biology.

This article is dedicated to Emeritus Professor Gary E. Hart, in honor of his many contributions to analysis and comprehension of polyploid plant (wheat) genomes. We thank the *Arabidopsis thaliana* genome center (ATGC) for providing an annotated sequence of the BAC, F12M16. We are grateful to Thomas C. Osborn, Dorothy E. Shippen, and Douglas E. Soltis for constructive suggestions on improving the manuscript and to members of the Chen lab for numerous discussions. The work was supported by grants from the National Science Foundation Plant Genome Program (NSF0077774) and the Texas Agricultural Experiment Station.

1. Lewis, W. H. (1980) *Polyploidy* (Plenum, New York).
2. Becak, M. L. & Becak, W. (1998) *Cytogenet. Cell Genet.* **80,** 28–33.
3. Stebbins, G. L. (1950) *Variation and Evolution in Plants* (Columbia Univ. Press, New York).
4. Stebbins, G. L. (1971) *Chromosomal Evolution in Higher Plants* (Edward Arnold, London).
5. Masterson, J. (1994) *Science* **264,** 421–424.
6. Soltis, P. S. & Soltis, D. E. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 7051–7057.
7. Wendel, J. F. (2000) *Plant Mol. Biol.* **42,** 225–249.
8. Hilu, K. W. (1993) *Am. J. Bot.* **80,** 2521–2528.
9. Hughes, M. K. & Hughes, A. L. (1993) *Mol. Biol. Evol.* **10,** 1360–1369.
10. Ferris, S. D. & Whitt, G. S. (1976) *Nature (London)* **265,** 258–260.
11. Bietz, J. A. (1987) in *Wheat and Wheat Improvement*, ed. Heyne, E. G. (Am. Soc. Agron., Madison, WI), pp. 215–242.
12. Payne, P. I., Holt, L. M., Jackson, E. A. & Law, N. C. (1984) *Philos. Trans. R. Soc. London B* **304,** 359–371.
13. Lynch, M. & Conery, J. S. (2000) *Science* **290,** 1151–1155.
14. McClintock, B. (1984) *Science* **226,** 792–801.
15. Feldman, M., Liu, B., Segal, G., Abbo, S., Levy, A. A. & Vega, J. M. (1997) *Genetics* **147,** 1381–1387.
16. Song, K., Lu, P., Tang, K. & Osborn, T. C. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 7719–7723.
17. O'Neill, R. J. W., O'Neill, M. J. & Graves, J. A. M. (1998) *Nature (London)* **393,** 68–72.
18. Comai, L., Tyagi, A. P., Winter, K., Holmes-Davis, R., Reynolds, S. H., Stevens, Y. & Byers, B. (2000) *Plant Cell* **12,** 1551–1568.
19. Vrana, P. B., Fossella, J. A., Matteson, P., del Rio, T., O'Neill, M. J. & Tilghman, S. M. (2000) *Nat. Genet.* **25,** 120–124.
20. Pikaard, C. S. & Chen, Z. J. (1998) in *RNA Polymerase I: Transcription of Eukaryotic Ribosomal RNA*, ed. Paule, M. R. (R. G. Landes, Austin, TX), pp. 277–294.
21. Reeder, R. H. (1985) *J. Cell Biol.* **101,** 2013–2016.
22. Chen, Z. J., Comai, L. & Pikaard, C. S. (1998) *Proc. Natl. Acad. Sci. USA* **95,** 14891–14896.
23. Chen, Z. J. & Pikaard, C. S. (1997) *Genes Dev.* **11,** 2124–2136.
24. Chen, Z. J. & Pikaard, C. S. (1997) *Proc. Natl. Acad. Sci. USA* **94,** 3442–3447.
25. O'Kane, S., Schaal, B. & Al-Shehbaz, I. (1995) *Syst. Bot.* **21,** 559–566.
26. Bachem, C. W., van der Hoeven, R. S., de Bruijn, S. M., Vreugdenhil, D., Zabeau, M. & Visser, R. G. (1996) *Plant J.* **9,** 745–753.
27. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
28. Konieczny, A. & Ausubel, F. M. (1993) *Plant J.* **4,** 403–410.
29. Cubas, P., Lauter, N., Doebley, J. & Coen, E. (1999) *Plant J.* **18,** 215–222.
30. Cubas, P., Vincent, C. & Coen, E. (1999) *Nature (London)* **401,** 157–161.
31. Doebley, J., Stec, A. & Hubbard, L. (1997) *Nature (London)* **386,** 485–488.
32. Theologis, A., Ecker, J. R., Palm, C. J., Federspiel, N. A., Kaul, S., White, O., Alonso, J., Altafi, H., Araujo, R. & Bowman, C. L. (2000) *Nature (London)* **408,** 816–820.
33. Osterlund, M. T., Hardtke, C. S., Wei, N. & Deng, X. W. (2000) *Nature (London)* **405,** 462–466.
34. Haufler, C. H. & Soltis, D. E. (1986) *Proc. Natl. Acad. Sci. USA* **83,** 4389–4393.
35. Gastony, G. J. (1991) *Proc. Natl. Acad. Sci. USA* **88,** 1602–1605.
36. Pichersky, E., Soltis, D. & Soltis, P. (1990) *Proc. Natl. Acad. Sci. USA* **87,** 195–199.
37. Gottlieb, L. D. & Ford, V. S. (1997) *Mol. Biol. Evol.* **14,** 125–132.
38. Comai, L. (2000) *Plant Mol. Biol.* **43,** 387–399.
39. Russo, V. E. A., Martienssen, R. A. & Riggs, A. D. (1996) *Epigenetic Mechanisms of Gene Regulation* (Cold Spring Harbor Lab. Press, Plainview, NY).
40. Phillips, J. P., Tainer, J. A., Getzoff, E. D., Boulianne, G. L., Kirby, K. & Hilliker, A. J. (1995) *Proc. Natl. Acad. Sci. USA* **92,** 8574–8578.
41. Prudhomme, M., Mejean, V., Martin, B. & Claverys, J. P. (1991) *J. Bacteriol.* **173,** 7196–7203.
42. Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. (1996) *Nature (London)* **379,** 131–137.
43. Lee, J. T. (2000) *Cell* **103,** 17–27.
44. Kelley, R. L. & Kuroda, M. I. (2000) *Cell* **103,** 9–12.