Highly Condensed Potato Pericentromeric Heterochromatin Contains rDNA-Related Tandem Repeats

Robert M. Stupar,[1] Junqi Song,[1] Ahmet L. Tek, Zhukuan Cheng, Fenggao Dong and Jiming Jiang[2]

Department of Horticulture, University of Wisconsin, Madison, Wisconsin 53706

Manuscript received January 24, 2002
Accepted for publication August 16, 2002

ABSTRACT

The heterochromatin in eukaryotic genomes represents gene-poor regions and contains highly repetitive DNA sequences. The origin and evolution of DNA sequences in the heterochromatic regions are poorly understood. Here we report a unique class of pericentromeric heterochromatin consisting of DNA sequences highly homologous to the intergenic spacer (IGS) of the 18S•25S ribosomal RNA genes in potato. A 5.9-kb tandem repeat, named 2D8, was isolated from a diploid potato species *Solanum bulbocastanum*. Sequence analysis indicates that the 2D8 repeat is related to the IGS of potato rDNA. This repeat is associated with highly condensed pericentromeric heterochromatin at several hemizygous loci. The 2D8 repeat is highly variable in structure and copy number throughout the Solanum genus, suggesting that it is evolutionarily dynamic. Additional IGS-related repetitive DNA elements were also identified in the potato genome. The possible mechanism of the origin and evolution of the IGS-related repeats is discussed. We demonstrate that potato serves as an interesting model for studying repetitive DNA families because it is propagated vegetatively, thus minimizing the meiotic mechanisms that can remove novel DNA repeats.

THE term heterochromatin was first used by Emil Heitz (1928) to describe condensed regions of plant and animal karyotypes. Heterochromatin can be distinguished from euchromatin by a number of cytological staining techniques. Euchromatin is lightly stained and is generally associated with transcriptional activity. Heterochromatin is cytologically defined as darkly stained and deeply condensed chromatin and is often associated with transcriptional inactivity, late DNA replication, suppressed genetic recombination, and silencing of euchromatic genes (reviewed by Renauld and Gasser 1997; Wallrath 1998). Molecular characterization of cytologically defined heterochromatin features is often a daunting task because of the complex and repetitive nature of the DNA sequences associated with most heterochromatin. For example, the centromeric regions of all human and *Arabidopsis thaliana* chromosomes and the entire heterochromatin portion of the *Drosophila melanogaster* genome could not be sequenced by the most advanced sequencing technologies (Adams *et al.* 2000; Arabidopsis Genome Initiative 2000; International Human Genome Sequencing Consortium 2001).

The authors dedicate this work to the memory of friend and colleague Robert E. Hanneman, Jr., who devoted his professional life to potato genetics.

Sequence data from this article have been deposited with the EMBL/GenBank Data Libraries under accession nos. AF473838 and AF473839.

[1]These authors contributed equally to this work.

[2]*Corresponding author:* Department of Horticulture, University of Wisconsin, Madison, WI 53706. E-mail: jjiang1@facstaff.wisc.edu

The association of transposable elements and satellite repeats with heterochromatin has been extensively documented (reviewed by Csink and Henikoff 1998; Bennetzen 2000). However, the origin of heterochromatin-associated DNA, especially the nontransposon sequences, is poorly understood. Short, simple repetitive sequence motifs, independent of selective pressure, are believed to develop the higher-order sequence periodicity of typical heterochromatic DNA (Hancock 1996). Therefore, heterochromatin is thought to spawn from "seed" sequences, which spread heterochromatic behavior by expanding tandem duplications (reviewed by Redi *et al.* 2001). Replication slippage, unequal crossing over, and rolling-circle amplification are among the proposed mechanisms of tandem-repeat amplification (Rochaix *et al.* 1974; Dover 1986; Flavell 1986; Strand *et al.* 1993).

Here we report the discovery of a unique class of pericentromeric heterochromatin consisting of a tandemly repeated DNA element, named 2D8 repeat, that is highly homologous to the intergenic spacer (IGS) of the 18S•25S ribosomal RNA genes in potato. The IGS-related repeats are highly variable in structure and copy number throughout the Solanum genus, suggesting that these repeats are evolutionary dynamic. The 2D8 repeat is a rare heterochromatin-associated DNA element that shows homology to DNA sequences with known function.

MATERIALS AND METHODS

Plant materials: The bacterial artificial chromosome (BAC) clones used in this study were derived from two BAC libraries.

One library was constructed from clone PT29 of *Solanum bulbocastanum* (Song *et al.* 2000), a diploid ($2n = 2x = 24$) wild relative of cultivated potato. The second library was constructed from haploid clone USW1 ($2n = 2x = 24$), derived from potato (*S. tuberosum*, $2n = 4x = 48$) variety Katahdin. Other Solanum and related species used in the study were provided by Dr. David Spooner from the U.S. Department of Agriculture-Agricultural Research Service and Department of Horticulture, University of Wisconsin-Madison. These species included *S. morelliforme*, *S. cardiophyllum* (spp. *cardiophyllum*), *S. lesteri*, *S. capsicibaccatum*, *S. chacoense* (spp. *chacoense*), *S. boliviense*, *S. infundibuliforme*, *S. agrimonifolium*, *S. albornozii*, *S. verrucosum*, *S. multidissectum*, *S. oplocense*, *S. curtilobum*, *S. fendleri* (spp. *fendleri*), *S. iopetalum*, *S. etuberosum*, *S. suaveolens*, *S. muricatum*, *S. taeniotrichum*, *S. appendiculatum*, *S. dulcamara*, *S. ochranthum*, *S. pseudocapsicum*, *S. diploconos*, *Lycopersicon esculentum*, and *Capsicum pubescens*.

Isolation, sequencing, and sequence analysis of repetitive DNA elements: BAC DNAs were isolated using standard alkaline lysis methods (Song *et al.* 2000). BAC clone 2D8 was digested with *Hin*dIII and the DNA fragments were separated by agarose gel electrophoresis. The two resulting fragments, 5.9 and 8.8 kb, respectively, were subcloned into plasmid pUC18. Sequencing of the subclones was conducted using an ABI Prism automated DNA sequencer at the University of Wisconsin Biotechnology Center. BLAST similarity searches were performed using the BLASTN algorithm. All further sequence analyses were performed using the Lasergene '99 software package (DNAStar, Madison, WI).

Statistical analyses were performed to test if specific motifs were occurring at a disportionately high frequency in the sequenced subclones. The known frequencies of A, T, C, and G bases for each sequence of interest were used to calculate the expected frequency of autonomously replicating sequence (ARS) consensus sequence (ACS) motifs (5′-WTTTAYRT TTW-3′) for all bases arranged at random. The analysis was conducted at two stringencies, expected frequencies for 9/11 base matches and 11/11 base matches. The expected match frequency was then compared with the observed match frequency by chi-square analysis to identify significant deviations from the expected ACS-related frequencies. A similar search was performed for A-tract sequences ≥6 consecutive adenine bases. For each sequence of interest, the known frequency of adenine bases was used to generate the expected total number of adenine bases involved in A-tracts ≥6. The total number of expected A-tract bases for one 5′ → 3′ sequence direction was calculated according to the equation

$$\text{Expected Adenines in tracts } (\geq 6) = \text{Total Bases} \times \left[\%\text{A} - \left(\sum_{n=1}^{5} n(\%\text{A})^n(\%\text{Not A})^2\right)\right]$$

where n represents the total base length of the A-tract. The number of residues expected to be within A-tracts was then compared with the number of A-tract residues observed. Chi-square analysis was performed to identify significant deviations from the expected abundance of A-tract residues.

Fluorescence *in situ* hybridization: Haploid *S. tuberosum* clone USW1 and *S. bulbocastanum* clone PT29 were used for cytological studies. These two clones were also used for BAC library constructions. Chromosome and DNA fiber preparations were according to published protocols (Jackson *et al.* 1998; Dong *et al.* 2000). Plasmid pTa71, which contains the coding sequences for the 18S•26S ribosomal RNA genes of wheat (Gerlach and Bedbrook 1979), was used as a rDNA probe. The DNA was labeled with either digoxigenin-11-dUTP or biotin-16-dUTP. Probe preparations and signal detection for fluorescence *in situ* hybridization (FISH) and fiber-FISH were described previously (Jackson *et al.* 1998; Dong *et al.* 2000). Chromosomes were counterstained by propidium iodide (PI) or 4′,6-diamidino-2-phenylindole (DAPI) in Vectashield antifade solution (Vector Laboratories, Burlingame, CA). All images were captured digitally using a SenSys charge-coupled device (CCD) camera (Roper Scientific, Tucson, AZ) attached to an Olympus BX60 epifluorescence microscope. The CCD camera was controlled using IPLab Spectrum v3.1 software (Signal Analytics, Vienna, VA) on a Macintosh computer. Gray scale images were captured for each color channel and then merged. The final images were adjusted with Adobe Photoshop 5.1 (Adobe Systems, San Jose, CA).

Southern and dot-blot hybridizations: Genomic DNAs were isolated from leaf tissue of the plant species. DNA samples were digested with appropriate restriction enzymes. DNA fragments were separated by agarose electrophoresis and transferred to Hybond-N^+ membrane (Amersham Biosciences, Piscataway, NJ). Prehybridization and hybridization were performed at 65° in 5× SSC, 0.5% SDS, 0.02 M $NaPO_4$ (pH 6.5), 2 mM EDTA, 10 mM Tris (pH 7.4), and 0.02% denatured salmon sperm DNA. The membrane was hybridized with ^{32}P-labeled DNA probes for 24 hr. Following hybridization, the membrane was sequentially washed with 2× SSC + 0.1% SDS and 0.5× SSC + 0.1% SDS at 65° for 15 min each. The radioactive membrane was then exposed to X-ray film.

DNA samples used for dot-blot analysis were further purified through a cesium chloride gradient. Dot-blot hybridization was performed essentially as described by Miller *et al.* (1998).

RESULTS

The 2D8 repeat is associated with highly condensed pericentromeric heterochromatin: To identify potato BACs containing highly repetitive DNA sequences, filters prepared from a *S. bulbocastanum* BAC library (Song *et al.* 2000) were screened using a small amount of *S. bulbocastanum* genomic DNA as a probe. Clone 2D8 showed one of the strongest hybridization signals to the genomic DNA, suggesting that this BAC may contain one of the most repetitive DNA elements in the potato genome. Pulsed-field gel electrophoresis analysis showed that the insert size of 2D8 is ~230 kb. The *Hin*dIII restriction pattern of 2D8 gave only three bands: a strong band at 5.9 kb, a relatively weaker band at 8.8 kb, and the BAC vector band at 7.4 kb (Figure 1A). These results indicate that the insert of BAC 2D8 contains mainly a 5.9-kb tandem repeat. This 5.9-kb band was subcloned for further analysis and was named the "2D8 repeat." Southern hybridization of the 2D8 repeat to *Hin*dIII-digested genomic DNA of *S. bulbocastanum* verified the presence of a 5.9-kb repetitive unit, together with six additional bands of greater size (Figure 1B). The majority of the 2D8-related DNA remained undigested, presumably due to lack of *Hin*dIII restriction sites within many 2D8 sequences (Figure 1B).

FISH analysis of a 5.9-kb subclone on the somatic metaphase chromosomes of a haploid *S. tuberosum* clone USW1 revealed a single hemizygous hybridization site near the centromere of an unidentified chromosome (Figure 2, A–D). Weak FISH signals were observed at the nucleolus organizing regions (NOR). The cross-hybridization of the 2D8 repeat to NOR was confirmed by FISH analysis using a rDNA probe, pTa71 (Figure 2, B–D). FISH analysis on meiotic pachytene chromo-

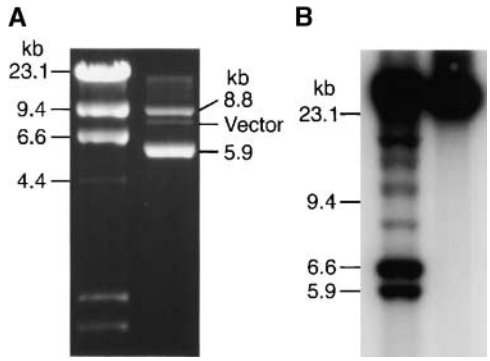

FIGURE 1.—(A) Agarose gel electrophoresis of BAC 2D8 following *Hin*dIII digestion (right lane). BAC 2D8 produces a strong 5.9-kb band and a much weaker 8.8-kb band. The marker lane (left) is lambda DNA digested with *Hin*dIII. (B) Southern blot hybridization using the 5.9-kb 2D8 subclone as a probe on *Hin*dIII-digested genomic DNA of *S. bulbocastanum* (left lane) and undigested *S. bulbocastanum* DNA (right lane). The *Hin*dIII-digested DNA sample shows the expected band at ~5.9 kb.

somes of USW1 revealed the expected single location of the 2D8 repeat. In early pachytene chromosomes the FISH signal is associated with a conspicuous heterochromatic feature that is brightly stained by DAPI (Figure 2, E–G). The 2D8-associated region is highly condensed and appears similar to the knobs of maize pachytene chromosomes described by McCLINTOCK (1929). Cross-hybridization of the 2D8 probe to the NOR was also observed in the pachytene phase in USW1 (Figure 2F).

In *S. bulbocastanum*, two strong, one intermediate, and one weak FISH signal were detected on four somatic metaphase chromosomes, with all signals located near the centromeric regions (Figure 2, H–J). Hybridization of 2D8 to NOR on *S. bulbocastanum* chromosomes was also observed (data not shown) but the cross-hybridization signals were not as strong as those observed in USW1. FISH on *S. bulbocastanum* pachytene chromosomes also yielded four distinct signals (Figure 2L), indicating that the four loci do not pair with one another and are therefore hemizygous. The high resolution pachytene FISH signals confirmed the pericentromeric locations of the 2D8 repeat (Figure 2M). The pachytene regions associated with the FISH signals are brightly stained by DAPI and highly condensed as compared to the distal euchromatic regions.

The 2D8 repeat is homologous to the IGS sequence of potato rDNA: One 5.9-kb 2D8 subclone was completely sequenced. This 5862-bp sequence consists of two diverged monomers of similar size and composition (GenBank accession no. AF473838). Each 2D8 monomer includes a cluster of AT-rich and a cluster of GC-rich subrepeats. Figure 3 shows the nested organization of repeats within BAC 2D8.

The dot-matrix display indicates a tandem-repetitive nature of sequences within the AT and GC-rich subrepeat clusters, with subrepeats displaying imperfect structural homology to one another (Figure 4A). The two monomeric units also display significant repeat divergence relative to one another. However, forward and reverse end-sequence analysis of five additional 5.9-kb subclones and two 8.8-kb subclones revealed a >98% sequence identity among all subclones. These results suggest that the 2D8 sequence is highly conserved among higher-order repeats of 5.9 and 8.8 kb, with the 8.8-kb fragments likely representing trimeric derivatives of the 5.9-kb dimer.

BLAST search revealed a high sequence homology (*E* value = 4*e*-26) of the 2D8 repeat to the IGS sequence of potato rDNA. The AT-rich and GC-rich subrepeat clusters each match the compatible domains of the potato rDNA IGS sequence (Figure 4B). Although the 2D8 sequence has diverged from the IGS sequence, the 2D8 repeat still maintains the ~74- and ~54-bp GC-rich repeats associated with the IGS, identified by BORISJUK and HEMLEBEN (1993).

Despite high sequence homology, the 2D8 sequence appears to be significantly diverged from the IGS. BORISJUK and HEMLEBEN (1993) reported the presence of a ATATATAAGGGGGG putative RNA polymerase I transcription initiation site (TIS) sequence within the potato IGS. The 2D8 dimer contains no such sites. The most similar sites within the 2D8 sequence match only 10 of the 14 bases specified by BORISJUK and HEMLEBEN (1993). There are no matches in the 2D8 sequence to any of the forms of the IGS TIS sites previously described in plants (PERRY and PALUKAITIS 1990).

Additionally, the AT-rich subrepeat clusters are far longer in the 2D8 repeat than in the IGS (Figure 4B). We also sequenced a 2.1-kb fragment from a BAC clone 105N20 (GenBank accession no. AF473839), which was isolated from a BAC library constructed from *S. tuberosum* haploid USW1. BAC 105N20 was identified due to its hybridization to the 2D8 repeat. A large, homologous AT-rich cluster was also identified in 105N20. We tested whether specific sequence motifs were prevalent in the AT-rich repeat clusters of the 2D8 and 105N20 sequences. We found that these regions contain an abundance of motifs related to the ARS ACS motif (5′-WTTTAYRTTTW-3′). The observed frequency of ACS-related elements in the 2D8 and 105N20 repeats is significantly greater than would be expected of a random arrangement of sequences with the known base compositions of the AT-rich domains (Table 1). Likewise, A-tract sequences of ≥6 consecutive adenine bases, which are known to associate with DNA bending and scaffold attachment regions (SARs; KOO *et al.* 1986; BOULIKAS 1995), were also found to be more abundant than would be expected for the known AT bases frequencies (Table 2).

Organization and distribution of the 2D8 repeat in the potato genome: FISH analysis on extended DNA fibers prepared from both USW1 and *S. bulbocastanum* generated long and contiguous beaded signals, confirming that the 2D8 repeat is organized as long tandem arrays. Applying the 2.87-kb/μm fiber-FISH calibration

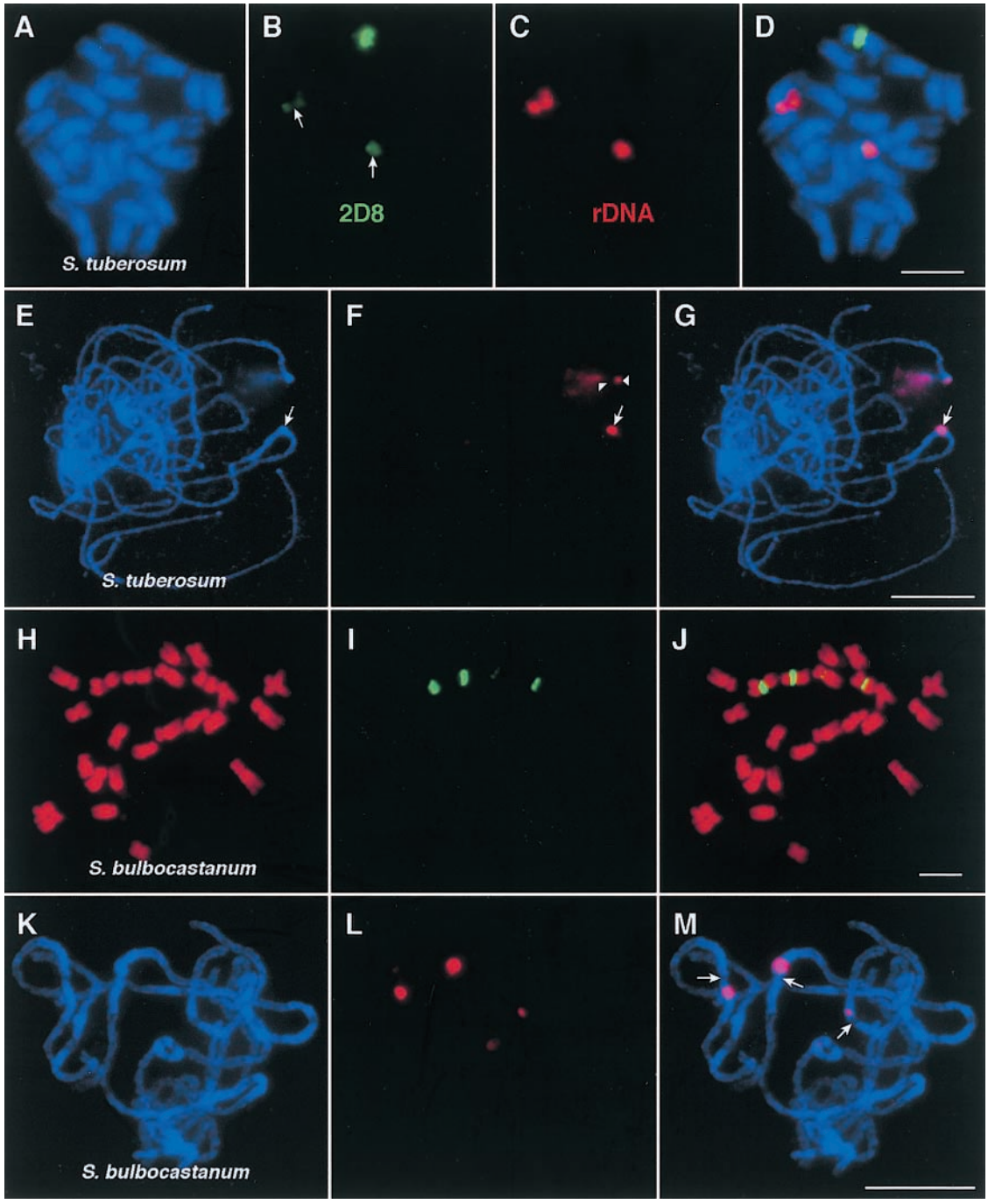

FIGURE 2.—FISH analyses of the 2D8 repeat in *S. tuberosum* and *S. bulbocastanum*. (A) Somatic metaphase chromosomes of haploid *S. tuberosum* clone USW1; (B) FISH signals derived from the 2D8 repeat; (C) FISH signals derived from a rDNA probe pTa71; (D) a merged image. Arrows in B point to the cross-hybridization of the 2D8 repeat to the NOR. (E) Meiotic pachytene chromosomes of haploid *S. tuberosum* clone USW1; (F) FISH signals derived from the 2D8 repeat; (G) a merged image. The FISH signal (arrows in F and G) is located on a conspicuous heterochromatic knob (arrow in E). The arrowheads (F) point to the cross-hybridization of the 2D8 repeat to the NOR. (H) Somatic metaphase chromosomes of *S. bulbocastanum*; (I) FISH signals derived from the 2D8 repeat; (J) a merged image. Two strong, one intermediate, and one weak hybridization were detected on four different chromosomes. All four hybridization sites are located close to the centromeres of the chromosomes (J). (K) Meiotic pachytene chromosomes of *S. bulbocastanum*; (L) FISH signals derived from the 2D8 repeat; (M) a merged image. Arrows in M point to the centromeres of three pachytene bivalents. Bars: D and J, 5 μm; G and M, 10 μm.

parameter (JACKSON *et al.* 1998), contiguous signals for individual fibers in USW1 were observed upward of 2 Mb, the limit of the fiber-FISH technique in plants (JACKSON *et al.* 1998). Figure 5A shows an ~1.8-Mb fiber-FISH signal observed in USW1.

Dot-blot hybridization was used to estimate the copy

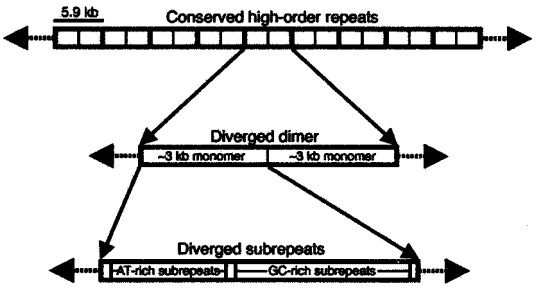

FIGURE 3.—Structure of the 2D8 repeat. Each 5.9-kb unit is partitioned into a dimeric substructure in which the monomers show some divergence from one another. The ~3-kb monomers consist of AT-rich and GC-rich subrepeat clusters.

number of the 2D8 repeat in potato. Using the DNA content standard of 1.8 pg/1C for the potato genome (ARUMUGANATHAN and EARLE 1991), we estimated that the 2D8-related sequence occupies ~7 and 3.5 Mb of the *S. bulbocastanum* (PT29, including four 2D8 loci) and USW1 (a single 2D8 locus) genomes, respectively (Figure 6).

Conservation of the 2D8 repeat was analyzed by Southern blot analysis across a wide range of Solanum species and related taxa. Figure 7 shows that 2D8-related sequences were detected in all tuber-bearing species from Solanum section *Petota*, including representative species from all four groups in series *Tuberosa* (HAWKES 1990; SPOONER and CASTILLO 1997), and non-tuber-bearing species *S. etuberosum* (section *Etuberosum*) and *S. ochranthum* (section *Juglandifolium*; SPOONER *et al.* 1993). Particularly intense signals were observed from *S. agrimonifolium*, *S. albornozii*, and *S. fendleri*, suggesting that these species contain an especially high copy number of 2D8-related sequences. The Southern hybridization results also indicate that copy number and restriction patterns of the 2D8-related sequences display no strong correlation with relatedness among taxa, indicating that the repeat family may be evolutionarily dynamic, capable of rapid expansion, contraction, and structural alterations.

The distribution of the 2D8-related sequences was also studied among distantly related Solanum sections (SPOONER *et al.* 1993). Hybridization to the 2D8 repeat was detected in the following outgroups: *S. diploconos*, *S. muricatum*, *S. taeniotrichum*, *S. appendiculatum*, and *Lycopersicum esculentum*. However, the 2D8 probe failed to hybridize to the species within the following outgroups: *S. pseudocapsicum*, *S. suaveolens*, *S. dulcamara*, and *C. pubescens*. The 2D8 probe also failed to hybridize to species of various distantly related taxa, such as tobacco, soybean, and Arabidopsis (data not shown).

Southern hybridization was further used to study the DNA methylation status of the 2D8 repeat in *S. bulbocastanum* and USW1 (Figure 8). Genomic DNAs were digested by methylation-sensitive and -insensitive isoschizomers, blotted, and probed with the 5.9-kb 2D8 clone.

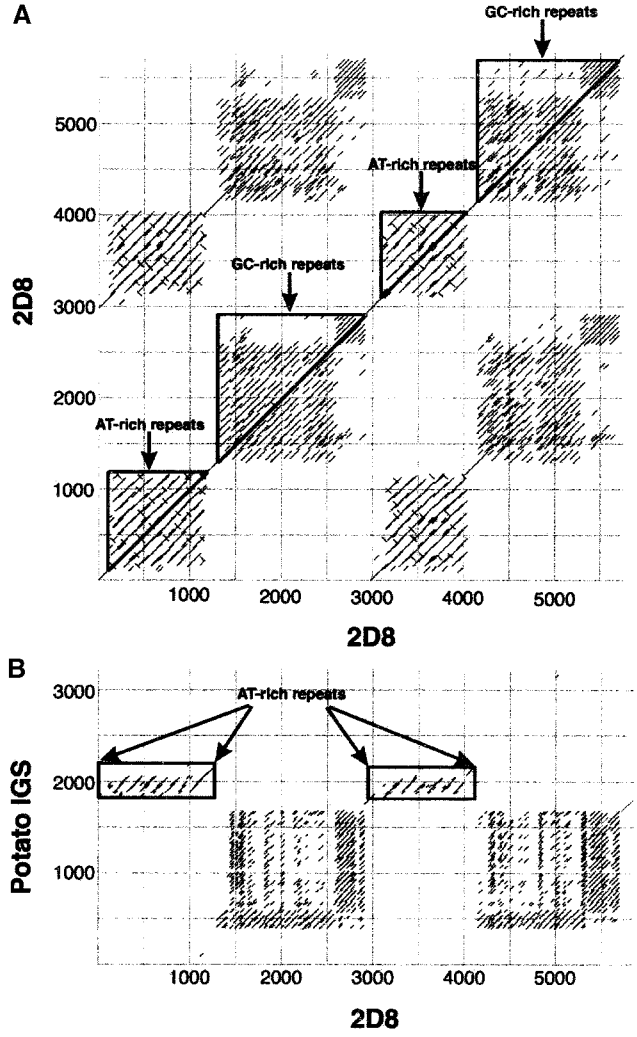

FIGURE 4.—Dot-matrix analyses of the 2D8 repeat. (A) A self comparison of the 2D8 subclone was generated using the MegAlign application from the Lazergene '99 software package. The matrix stringency was set to 32/50 identical nucleotides. The dimeric structure of the subclone is observed as two discrete monomers, with each consisting of AT-rich subrepeats adjacent to GC-rich subrepeats. (B) Dot-matrix comparison of the 5.9-kb 2D8 subclone and the potato rDNA IGS. The matrix stringency was set to 21/30 identical nucleotides. The 2D8 sequence displays exclusive homology to both the GC-rich and the AT-rich regions of the potato IGS. The AT-rich alignments appear elongated and rectangular in shape because the repeat cluster is much larger in the 2D8 repeat than in the potato IGS.

Both *Eco*RII and *Bst*NI cleave the CC(A/T)GG site, but only *Eco*RII is sensitive to methylation of the interior cytosine. The differential restriction patterns between *Eco*RII and *Bst*NI digestions (Figure 8) indicate extensive methylation at the CNG cytosine residues within the 2D8 sequences in both *S. bulbocastanum* and *S. tuberosum*. *Hpa*II and *Msp*I recognize the CCGG sequence, but neither can cut when the 5′ C is methylated. Only *Msp*I, not *Hpa*II, can cleave if the internal C is methylated. Figure 8 shows that the restriction patterns of the *Hpa*II

TABLE 1

Sequence analysis of 11-bp reading frames that exhibit ACS-related sequences within the AT-rich regions of the 2D8 and 105N20 repeats and the potato rDNA IGS

Sequence	Total bases	% AT	Exp. ACS (9/11)	Obs. ACS (9/11)	Chi-square	*P* value
2D8 AT cluster 1	1217	82.82	80.3	105	7.57	0.0059*
2D8 AT cluster 2	1179	81.34	72.6	97	8.17	0.0043*
105N20 AT cluster	2065	82.96	133.0	185	20.3	<0.0001*
Potato IGS	367	79.56	18.2	28	5.26	0.0219
			(11/11)	(11/11)		
2D8 AT cluster 1	1217	82.82	1.20	7	28.1	<0.0001*
2D8 AT cluster 2	1179	81.34	1.06	3	3.52	0.0607
105N20 AT cluster	2065	82.96	1.91	5	5.02	0.025
Potato IGS	367	79.56	0.24	1	2.47	0.1157

*Significant difference ($P \leq 0.01$) between observed and expected frequencies of ACS-related elements. Exp., expected; obs., observed.

and *Msp*I digestions are similar. The fact that *Hpa*II cut the 2D8 loci of both genomes and shares some prominent bands with *Msp*I indicates that some CG cytosines are unmethylated within the repetitive locus. This is in contrast to previous reports in which 5′-methylation is complete at CG cytosines in repetitive plant DNA (Bennetzen *et al.* 1994; Miller *et al.* 1998; Cheng *et al.* 2001).

Additional IGS-related DNA sequences in the potato genome: We identified additional clones in the *S. bulbocastanum* BAC library that contain sequences related to the rDNA IGS. One such BAC, 26J19, is structurally similar to BAC 2D8. *Hin*dIII digestion of BAC 26J19 produced only six fragments ranging from 6 to 10 kb. End sequencing of plasmid clones containing these fragments revealed high sequence similarity with potato rDNA IGS (data not shown). FISH analysis showed that BAC 26J19 hybridizes to two *S. bulbocastanum* chromosomes that also bear a 2D8 locus (Figure 5E). On one chromosome the 26J19 locus is centromere-proximal to the 2D8 locus, while on the other chromosome the 26J19 locus is centromere-distal to the 2D8 locus. Cross-hybridization of 26J19 with NOR was also observed in FISH experiments. Interestingly, no significant cross-hybridization between the 2D8 and 26J19 probes was observed (Figure 5, B–E).

Another BAC 4A4 was also identified by its strong hybridization to the potato genomic DNA. *Hin*dIII digestion of 4A4 produced numerous bands, suggesting a different structure from that of BAC clones 2D8 and 26J19. Several *Hin*dIII fragments of BAC 4A4 were subcloned and end sequenced. Sequences related to both rDNA spacer and rDNA coding regions were found in these subclones. Only one major locus was detected in *S. bulbocastanum* when BAC 4A4 was used as a FISH probe (Figure 5G). The 4A4 locus is located on a *S. bulbocastanum* chromosome that also carries a major 2D8 locus. Faint cross-hybridization of BAC 4A4 to the 2D8 loci in *S. bulbocastanum* was observed in some cells (Figure 5G). Similarly, cross-hybridization of the 2D8 probe to the 4A4 locus was also observed on some pachytene chromosomes (Figure 5H).

DISCUSSION

We have discovered a tandem repeat, 2D8, which is associated with the pericentromeric heterochromatin in potato. The most interesting characteristic of the 2D8 repeat is its high homology to the rDNA IGS sequence. Three different scenarios can explain this relationship: (1) The 2D8 repeat elements are derived from the IGS

TABLE 2

Number of nucleotides involved in A-tracts ≥6 bp within the AT-rich regions of the 2D8 and 105N20 repeats and the potato rDNA IGS

Sequence	Total bases	% AT	Exp. A-tract bases	Obs. A-tract bases	Chi-square	*P* value
2D8 AT cluster 1	1217	82.82	76.0	160	92.9	<0.0001*
2D8 AT cluster 2	1179	81.34	76.1	154	79.8	<0.0001*
105N20 AT cluster	2065	82.96	115.5	305	310.9	<0.0001*
Potato IGS	367	79.56	17.3	53	73.6	<0.0001*

*Significant difference ($P \leq 0.01$) between observed and expected frequencies of ACS-related elements.

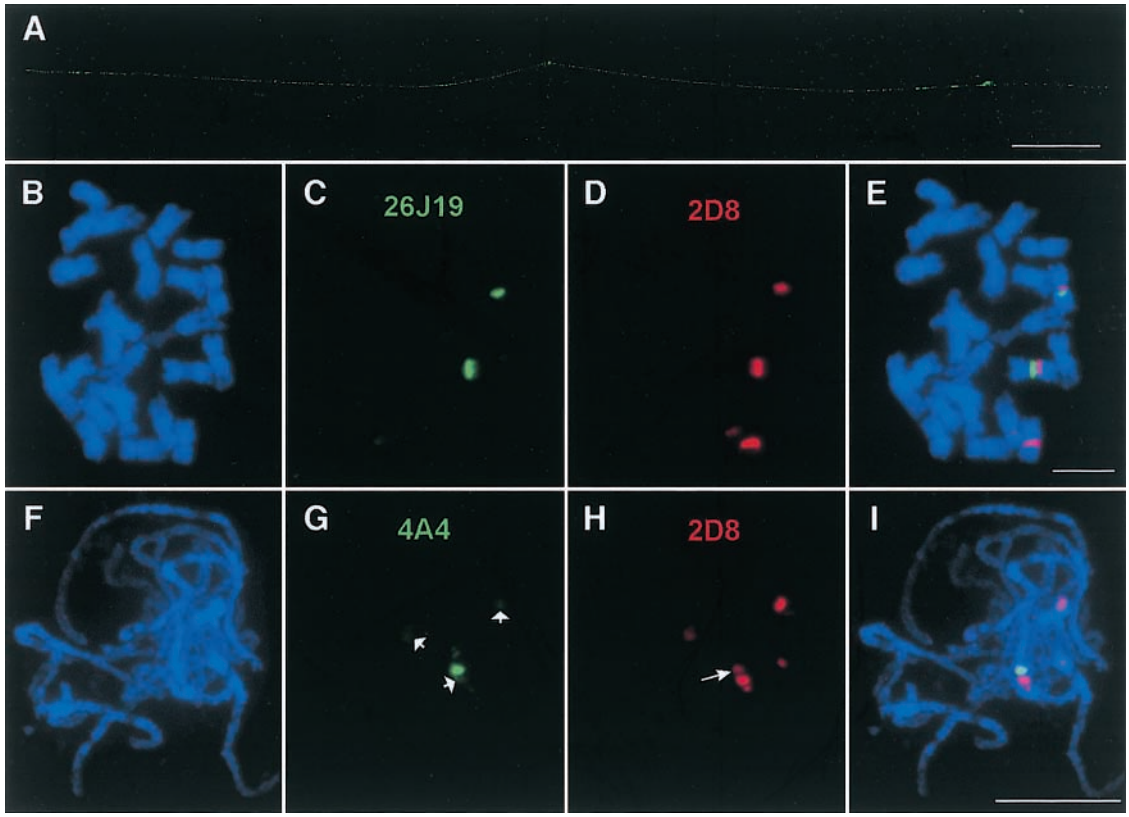

Figure 5.—The high-order structure of the 2D8 repeat and its relationship with other repeats related to the rDNA IGS. (A) A >600-μm FISH signal on DNA fibers prepared from *S. tuberosum* haploid clone USW1 using the 2D8 repeat as a probe. This signal represents ~1.8 Mb of DNA (2.87 kb/μm; see Jackson *et al.* 1998). Bar, 50 μm. (B) Somatic metaphase chromosomes of *S. bulbocastanum*; (C) FISH signals derived from BAC 26J19; (D) FISH signals derived from the 2D8 repeat; (E) a merged image. Bar, 5 μm. Two hybridization sites of 26J19 (C) are located on two chromosomes that also carry a 2D8 locus (D and E). (F) Meiotic pachytene chromosomes of *S. bulbocastanum*. (G) FISH signals derived from BAC 4A4; (H) FISH signals derived from the 2D8 repeat; (I) a merged image, Bar, 10 μm. Arrows in G point to the faint cross-hybridization of the 4A4 repeat to the 2D8 hybridization sites. The arrow in H points to the cross-hybridization of the 2D8 repeat to the major 4A4 hybridization site.

repeats; (2) the IGS repeat elements are derived from the 2D8 repeats; and (3) the 2D8 and IGS repeat elements are derived from an independent common ancestor.

There is a strong possibility that the 2D8 repeat elements were originally derived from the IGS repeats, as several previous studies have found evidence of rDNA-related sequence elements dispersed throughout eukaryotic genomes. These elements are termed rDNA orphons, and they have mainly been characterized as noncoding, small-unit tandem repeats of variable copy numbers. Such elements have been identified in various eukaryotic species, including yeast (Childs *et al.* 1981), animals (Arnheim *et al.* 1980; Kominami and Muramatsu 1987; De Lucchini *et al.* 1988; Lohe and Roberts 1990; Guimond and Moss 1999), and plants (Unfried *et al.* 1991; Falquet *et al.* 1997). It has been suggested that IGS invasion into satellite DNA may occur frequently during evolution, with most events remaining undetected due to the divergence of the sequences (Falquet *et al.* 1997). The 2D8 repeat sequences in this study may have been directly transposed or retrotransposed from the rDNA IGS origin and then diverged through an accumulation of mutations.

Within the first scenario, the 2D8 family may be derived from an intermediate satellite element previously derived from the rDNA IGS. As our results suggest, the 2D8 family is not the only IGS-related element found within the *S. bulbocastanum* genome. The 2D8, 26J19, and 4A4 repeat families may have arisen independently or, alternatively, the families may have evolved from one another. It is interesting to note that the 2D8, 26J19, and 4A4 repeats are located adjacently on the same chromosomal regions, suggesting that they may evolve from one another by duplication-related chromosome rearrangements. Large interchromosomal duplications have been reported recently in the human genome and are associated with the pericentromeric regions of most human chromosomes (Horvath *et al.* 2001).

Reciprocally, it has also been proposed that satellite DNA may be transposed into the rDNA IGS (Schmidt *et al.* 1982; Schmidt 1984; Maggini *et al.* 1991). It is

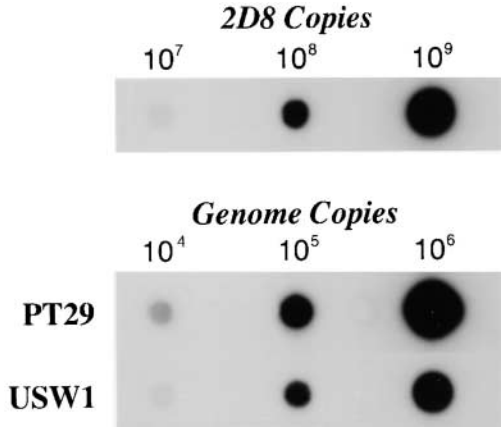

FIGURE 6.—Dot-blot analysis estimates the copy number of the 2D8 repeat in *S. bulbocastanum* (PT29) and potato (haploid clone USW1) genomes. Serial dilutions of known quantities from the 2D8 repeat and the genomic DNA of the two potato species were probed with the 5.9-kb 2D8 subclone. Genomic repeat copy number is calculated on the signal intensities generated from the genomic DNAs compared to the intensities of the 2D8 standards.

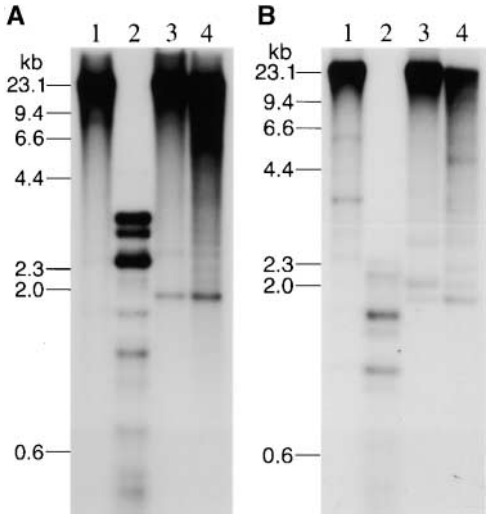

FIGURE 8.—Methylation analyses of the 2D8 repeat in (A) *S. bulbocastanum* (PT29) and (B) *S. tuberosum* (USW1). DNA was digested by *Eco*RII (lane 1), *Bst*NI (lane 2), *Hpa*II (lane 3), and *Msp*I (lane 4), respectively, and probed with the 5.9-kb 2D8 subclone.

therefore possible that the rDNA IGS repeat elements were originally derived from 2D8-related repeats. However, BORISJUK *et al.* (1994) used interspecific comparisons to demonstrate that Solanum species exhibit only slight variation in the length and sequences of the IGS, indicating that these regions are structurally conserved throughout the genus. Therefore, transposition of 2D8-derived sequences into the rDNA IGS would have had to occur prior to speciation within the Solanum genus.

The third possibility is that the 2D8 and rDNA IGS elements are both derived from an independent common ancestor. Within this scenario, the 2D8 and the IGS repeat elements would be directly derived from either the same ancestral element or related ancestral elements that are homologous to each other.

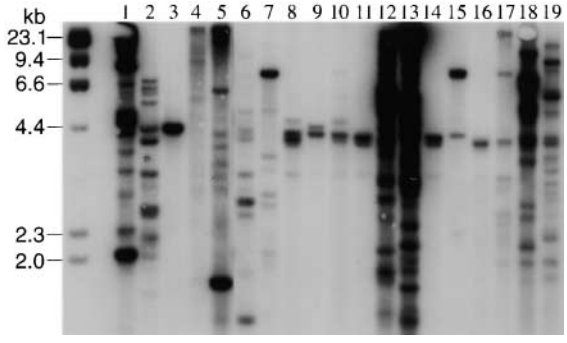

FIGURE 7.—Distribution of the 2D8-related sequences in Solanum and related species. *Dra*I-digested genomic DNAs were blotted and probed with the 5.9-kb 2D8 repeat. Lane 1, *S. bulbocastanum* accession PT29; 2, *S. tuberosum* haploid USW1; 3, *S. etuberosum*; 4, *S. ochranthum*; 5, *S. morelliforme*; 6, *S. cardiophyllum* (spp. *cardiophyllum*); 7, *S. lesteri*; 8, *S. capsicibaccatum*; 9, *S. chacoense* (spp. *chacoense*); 10, *S. boliviense*; 11, *S. infundibuliforme*; 12, *S. agrimonifolium*; 13, *S. albornozii*; 14, *S. verrucosum*; 15, *S. multidissectum*; 16, *S. oplocense*; 17, *S. curtilobum*; 18, *S. fendleri* (spp. *fendleri*); 19, *S. iopetalum*.

The 2D8 repeat family is evolutionarily dynamic throughout the Solanum species, in copy number, locus number, and sequence structure. Despite the dynamic nature of the 2D8 repeat, almost no divergence was identified within the 5.9- and 8.8-kb 2D8 repeat units. The near-perfect sequence identity among these units indicates that this locus has recently undergone concerted evolution. Numerous mechanisms, including unequal crossing over, gene conversion, and DNA transposition accompanied by amplification, are believed to drive the concerted evolution of repetitive DNA sequences (DOVER 1982). The concerted evolution of the IGS-related sequences may have occurred by recombinogenic mechanisms similar to those believed to operate in NOR evolution (KRYSTAL *et al.* 1981; DOVER and FLAVELL 1984; SCHLOTTERER and TAUTZ 1994; COPENHAVER and PIKAARD 1996). This would require a high rate of interlocus, intralocus, and/or sister chromatid recombination at the 2D8 loci, as the near-perfect sequence identity among the 2D8 repeat units indicates that amplification of this locus has occurred rapidly and recently. However, recombination is rare in pericentromeric regions in Solanaceae species (SHERMAN and STACK 1995). The hemizygosity of the 2D8 loci in *S. bulbocastanum* further suggests that recombinogenic mechanisms did not likely drive the concerted evolution of the 2D8 repeat.

An alternative hypothesis on the concerted evolution of the 2D8 repeat involves an extrachromosomal excision-amplification-reintegration mechanism (reviewed by STARK *et al.* 1989), whereby a specific DNA sequence is excised from the chromosome, amplified as an extra-

chromosomal element, and then reintegrated back into the genome at a different chromosomal location. Some of the previously reported rDNA orphons have been proposed to be derived from such extrachromosomal mechanisms (Lohe and Roberts 1990; Guimond and Moss 1999). It has been suggested that such a system may serve as a general means for genome-wide dispersions of repetitive sequences (Dvorak *et al.* 1987) and may function in the concerted evolution of paralogous NOR loci (Dvorak 1989). The DNA loci generated by such a system would originally be hemizygous. These loci may be maintained as hemizygous if the reintegration occurs in a region of suppressed recombination or/and if accomplished in a clonally propagating species such as the tuberizing Solanum species.

We found abundant sequence motifs in the 2D8 repeat that are similar to ACS, DNA bending, and SAR motifs (Tables 1 and 2). Similar motifs derived from the rDNA IGS of tobacco can increase the copy number and transcription level of heterologous DNA in transgenic constructs and have thus been termed amplification-promoting sequences (*aps*; Borisjuk *et al.* 2000). It is an intriguing possibility that *aps* sequences may have played a role in the higher-order amplification of the IGS-related tandem repeats in potato.

We are grateful to Dr. David Spooner for providing the DNA samples of the Solanum species and his phylogenetic expertise for our study. We thank Dr. Brian Yandell and Dr. Kiyotaka Nagaki for their advice on statistical analysis and Dr. Jason Lilly for technical assistance. We also thank Dr. Jan Dvorak for his valuable comments on the manuscript. This work was supported by Hatch Funds and partially by United States Department of Agriculture/National Research Institute grant 96-35300-3720 to J.J.

LITERATURE CITED

Adams, M. D., S. E. Celniker, R. A. Holt, C. A. Evans, J. D. Gocayne *et al.*, 2000 The genome sequence of *Drosophila melanogaster*. Science **287:** 2185–2195.

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

Arnheim, N., P. Seperack, J. Banerji, R. B. Lang, R. Miesfield *et al.*, 1980 Mouse rDNA nontranscribed spacer sequences are found flanking immunoglobulin CH genes and elsewhere throughout the genome. Cell **22:** 179–185.

Arumuganathan, K., and E. D. Earle, 1991 Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. **9:** 208–218.

Bennetzen, J. L., 2000 The many hues of plant heterochromatin. Genome Biol. **1:** 107.1–107.4.

Bennetzen, J. L., K. Schrick, P. S. Springer, W. E. Brown and P. SanMiguel, 1994 Active maize genes are unmodified and flanked by diverse classes of modified, highly repetitive DNA. Genome **37:** 565–576.

Borisjuk, N., and V. Hemleben, 1993 Nucleotide sequence of the potato intergenic spacer. Plant Mol. Biol. **21:** 381–384.

Borisjuk, N., L. Borisjuk, G. Petjuch and V. Hemleben, 1994 Comparison of nuclear ribosomal RNA genes among *Solanum* species and other Solanaceae. Genome **37:** 271–279.

Borisjuk, N., L. Borisjuk, S. Komarnytsky, S. Timeva, V. Hemleben *et al.*, 2000 Tobacco ribosomal DNA spacer element stimulates amplification and expression of heterologous genes. Nat. Biotechnol. **18:** 1303–1306.

Boulikas, T., 1995 Chromatin domains and prediction of MAR sequences. Int. Rev. Cytol. **162A:** 279–388.

Cheng, Z. K., R. M. Stupar, M. H. Gu and J. Jiang, 2001 A tandemly repeated DNA sequence is associated with both knob-like heterochromatin and a highly decondensed structure in the meiotic pachytene chromosomes of rice. Chromosoma **110:** 24–31.

Childs, G., R. Maxson, R. H. Cohn and L. Kedes, 1981 Orphons: dispersed genetic elements from tandem repetitive genes of eukaryotes. Cell **23:** 651–663.

Copenhaver, G. P., and C. S. Pikaard, 1996 Two-dimensional RFLP analyses reveal megabase-sized clusters of rRNA gene variants in *Arabidopsis thaliana*, suggesting local spreading of variants as the mode for gene homogenization during concerted evolution. Plant J. **9:** 273–282.

Csink, A. K., and S. Henikoff, 1998 Something from nothing: the evolution and utility of satellite repeats. Trends Genet. **14:** 200–204.

De Lucchini, S., F. Andronico, M. Andreazzoli, M. Guiliani, R. Savino *et al.*, 1988 Extra-chromosomal spacer sequences in Triturus. J. Mol. Biol. **204:** 805–813.

Dong, F., J. Song, S. K. Naess, J. P. Helgeson, C. Gebhardt *et al.*, 2000 Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato. Theor. Appl. Genet. **101:** 1001–1007.

Dover, G. A., 1982 Molecular drive, a cohesive model of species evolution. Nature **299:** 111–117.

Dover, G. A., 1986 Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. Trends Genet. **2:** 159–165.

Dover, G. A., and R. B. Flavell, 1984 Molecular co-evolution: rDNA divergence and the maintenance of function. Cell **38:** 622–623.

Dvorak, J., 1989 Evolution of multigene families: the ribosomal RNA loci of wheat and related species, pp. 83–97 in *Plant Population Genetics, Breeding, and Genetic Resources*, edited by A. H. D. Brown, M. T. Clark, A. L. Kahler and B. S. Weir. Sinauer Associates, Sunderland, MA.

Dvorak, J., D. Jue and M. Lassner, 1987 Homogenization of tandemly repeated nucleotide sequences by distance-dependent nucleotide sequence conversion. Genetics **116:** 487–498.

Falquet, J., R. Creusot and M. Dron, 1997 Molecular analysis of Phaseolus vulgaris rDNA unit and characterization of a satellite DNA homologous to IGS subrepeats. Plant Physiol. Biochem. **35:** 611–622.

Flavell, R. B., 1986 Repetitive DNA and chromosome evolution in plants. Philos. Trans. R. Soc. Lond. B Biol. Sci. **312:** 227–242.

Gerlach, W. L., and J. R. Bedbrook, 1979 Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res. **7:** 1869–1885.

Guimond, A., and T. Moss, 1999 A ribosomal orphon sequence from *Xenopus laevis* flanked by novel low copy number repetitive elements. Biol. Chem. **380:** 167–174.

Hancock, J. M., 1996 Simple sequences and the expanding genome. BioEssays **18:** 421–425.

Hawkes, J. G., 1990 *The Potato: Evolution, Biodiversity and Genetic Resources*. Belhaven Press, London.

Heitz, E., 1928 Das Heterochromatin der Moose I. Jahrb. Wiss Bot. **69:** 762–818.

Horvath, J. E., J. A. Bailey, D. P. Locke and E. E. Eichler, 2001 Lessons from the human genome: transitions between euchromatin and heterochromatin. Hum. Mol. Genet. **10:** 2215–2223.

International Human Genome Sequencing Consortium, 2001 Initial sequencing and analysis of the human genome. Nature **409:** 860–921.

Jackson, S. A., M. L. Wang, H. M. Goodman and J. Jiang, 1998 Application of fiber-FISH in physical mapping of *Arabidopsis thaliana*. Genome **41:** 566–572.

Kominami, R., and M. Muramatsu, 1987 Amplified ribosomal spacer sequence: structure and evolutionary origin. J. Mol. Biol. **193:** 217–222.

Koo, H. S., H. M. Wu and D. M. Crothers, 1986 DNA bending at adenine•thymine tracts. Nature **320:** 501–506.

Krystal, M., P. D'Eustachio, F. H. Ruddle and N. Arnheim, 1981 Human nucleolus organizers on nonhomologous chromosomes can share the same ribosomal gene variants. Proc. Natl. Acad. Sci. USA **78:** 5744–5748.

Lohe, A. R., and P. A. Roberts, 1990 An unusual Y chromosome

of *Drosophila simulans* carrying amplified rDNA spacer without RNA genes. Genetics **125:** 399–406.

Maggini, F., R. Cremonini, C. Zolfino, G. F. Tucci, R. D'Ovidio *et al.*, 1991 Structure and chromosomal localization of DNA sequences related to ribosomal subrepeats in *Vicia faba*. Chromosoma **100:** 229–234.

McClintock, B., 1929 Chromosome morphology in *Zea mays*. Science **69:** 629.

Miller, J. T., S. A. Jackson, S. Nasuda, B. S. Gill, R. A. Wing *et al.*, 1998 Cloning and characterization of a centromere-specific repetitive DNA element from *Sorghum bicolor*. Theor. Appl. Genet. **96:** 832–839.

Perry, K. L., and P. Palukaitis, 1990 Transcription of tomato ribosomal DNA and the organization of the intergenic spacer. Mol. Gen. Genet. **221:** 102–112.

Redi, C. A., S. Garagna, H. Zacharias, M. Zuccotti and E. Capanna, 2001 The other chromatin. Chromosoma **110:** 136–147.

Renauld, H., and S. M. Gasser, 1997 Heterochromatin: A meiotic matchmaker? Trends Cell Biol. **7:** 201–205.

Rochaix, J. D., A. Bird and A. Bakken, 1974 Ribosomal RNA gene amplification by rolling circles. J. Mol. Biol. **87:** 473–487.

Schlotterer, C., and D. Tautz, 1994 Chromosomal homogeneity of Drosophila ribosomal DNA arrays suggests intrachromosomal exchanges drive concerted evolution. Curr. Biol. **4:** 777–783.

Schmidt, E. R., 1984 Clustered and interspersed repetitive DNA sequences family of Chironomous. J. Mol. Biol. **178:** 1–15.

Schmidt, E. R., E. A. Godwin, H.-G. Keyl and N. Israelewski, 1982 Cloning and analysis of ribosomal DNA of Chironomus thummi piger and Chironomus thummi thummi: the nontranscribed spacer of Ch. th. thummi contains a highly repetitive DNA sequence. Chromosoma **87:** 389–407.

Sherman, J. D., and S. M. Stack, 1995 Two-dimensional spreads of synaptonemal complexes from solanaceous plants. VI. High-resolution recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics **141:** 683–708.

Song, J., F. Dong and J. Jiang, 2000 Construction of a bacterial artificial chromosome (BAC) library for potato molecular cytogenetics research. Genome **43:** 199–204.

Spooner, D. M., and T. Castillo, 1997 Reexamination of series relationships of South American wild potato (Solanaceae: *Solanum* sect. *Petota*): evidence from chloroplast DNA restriction site variation. Am. J. Bot. **84:** 671–685.

Spooner, D. M., G. J. Anderson and R. K. Jansen, 1993 Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). Am. J. Bot. **80:** 676–688.

Stark, G. R., M. Debatisse, E. Giulotto and G. M. Wahl, 1989 Recent progress in understanding mechanisms of mammalian DNA amplification. Cell **57:** 901–908.

Strand, M., T. A. Prolla, R. M. Liskay and T. D. Petes, 1993 Destabilization of tracts of simple repetitive DNA in yeast by mutations affecting DNA mismatch repair. Nature **365:** 274–276.

Unfried, K., K. Schiebel and V. Hemleben, 1991 Subrepeats of rDNA intergenic spacer present as prominent independent satellite DNA in *Vigna radiata* but not *Vigna angularis*. Gene **99:** 63–68.

Wallrath, L., 1998 Unfolding the mysteries of heterochromatin. Curr. Opin. Genet. Dev. **8:** 147–153.

Communicating editor: J. A. Birchler