The Plant Cell, Vol. 26: 1436–1447, April 2014, www.plantcell.org

RESEARCH ARTICLES

Boom-Bust Turnovers of Megabase-Sized Centromeric DNA in *Solanum* Species: Rapid Evolution of DNA Sequences Associated with Centromeres[C][W]

Haiqin Zhang,[a,b] Andrea Koblížková,[c] Kai Wang,[a] Zhiyun Gong,[a,d] Ludmila Oliveira,[a,e] Giovana A. Torres,[e] Yufeng Wu,[a] Wenli Zhang,[a] Petr Novák,[c] C. Robin Buell,[f] Jiří Macas,[c,1] and Jiming Jiang[a,1,2]

[a] Department of Horticulture, University of Wisconsin–Madison, Madison, Wisconsin 53706
[b] Triticeae Research Institute, Sichuan Agricultural University, Wenjiang, Sichuan 611130, People's Republic of China
[c] Institute of Plant Molecular Biology, Biology Centre ASCR, Ceske Budejovice CZ-37005, Czech Republic
[d] Key Laboratory of Crop Genetics and Physiology of Jiangsu Province/Key Laboratory of Plant Functional Genomics of Ministry of Education, Yangzhou University, Yangzhou 225009, People's Republic of China
[e] Departmento de Biologia, Universidade Federal de Lavras, Lavras MG 37200, Brazil
[f] Department of Plant Biology, Michigan State University, East Lansing, Michigan 48824

Centromeres are composed of long arrays of satellite repeats in most multicellular eukaryotes investigated to date. The satellite repeat–based centromeres are believed to have evolved from "neocentromeres" that originally contained only single- or low-copy sequences. However, the emergence and evolution of the satellite repeats in centromeres has been elusive. Potato (*Solanum tuberosum*) provides a model system for studying centromere evolution because each of its 12 centromeres contains distinct DNA sequences, allowing comparative analysis of homoeologous centromeres from related species. We conducted genome-wide analysis of the centromeric sequences in *Solanum verrucosum*, a wild species closely related to potato. Unambiguous homoeologous centromeric sequences were detected in only a single centromere (*Cen9*) between the two species. Four centromeres (*Cen2*, *Cen4*, *Cen7*, and *Cen10*) in *S. verrucosum* contained distinct satellite repeats that were amplified from retrotransposon-related sequences. Strikingly, the same four centromeres in potato contain either different satellite repeats (*Cen2* and *Cen7*) or exclusively single- and low-copy sequences (*Cen4* and *Cen10*). Our sequence comparison of five homoeologous centromeres in two *Solanum* species reveals rapid divergence of centromeric sequences among closely related species. We propose that centromeric satellite repeats undergo boom-bust cycles before a favorable repeat is fixed in the population.

INTRODUCTION

The centromere is a specialized chromosomal domain where the kinetochore is assembled and is defined by the presence of the centromere-specific histone H3 variant cenH3 (CENP-A in human) (Henikoff et al., 2001). The function of the centromere is conserved in all eukaryotes. By contrast, centromeric DNAs are among the most rapidly evolving sequences in eukaryotic genomes (Bensasson, 2011). Centromeres in many plant and animal species are composed of long arrays of satellite repeats and/or retrotransposon-related repetitive sequences (Henikoff et al., 2001; Jiang et al., 2003). Centromeric repeats appear to evolve rapidly (Melters et al., 2013). For example, rice (*Oryza sativa*) centromeres contain a 155-bp satellite repeat CentO (Dong et al., 1998; Cheng et al., 2002b) and CRR (Centromeric Retrotransposon of Rice) elements (Cheng et al., 2002b; Nagaki et al., 2005). The centromeres of *Oryza brachyantha*, a wild species diverged from rice for ~15 million years (Ammiraju et al., 2008), contain a 156-bp satellite repeat CentO-F (Lee et al., 2005; Yi et al., 2013), which shares no sequence similarity with CentO, and a centromere-enriched retrotransposon FRetro3 (Gao et al., 2009), which does not belong to the centromeric retrotransposon family and was detected in both centromeric and pericentromeric regions. Thus, the centromeres of these two species are composed of completely different repetitive DNA sequences.

Many existing centromeres are believed to have originated as neocentromeres that activated de novo from noncentromeric regions by acquiring cenH3/CENP-A (Kalitsis and Choo, 2012; Rocchi et al., 2012). Most newly formed neocentromeres are associated with gene "desert" regions and do not contain satellite repeats (Marshall et al., 2008; K. Wang et al., 2014). The evolutionarily new centromeres presumably accumulate satellite repeats and/or retrotransposons during evolution and eventually become repeat-based centromeres (Yan et al., 2006; Kalitsis and Choo, 2012). However, there is very limited information on this presumed transition from a neocentromere to a mature centromere. In particular, we have little knowledge regarding the emergence and survival of novel centromeric satellite repeats in

[1] These authors contributed equally to this work.
[2] Address correspondence to jjiang1@wisc.edu.
The authors responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantcell.org) are: Jiří Macas (macas@umbr.cas.cz) and Jiming Jiang (jjiang1@wisc.edu).
[C] Some figures in this article are displayed in color online but in black and white in the print edition.
[W] Online version contains Web-only data.
www.plantcell.org/cgi/doi/10.1105/tpc.114.123877

evolutionarily new centromeres. These evolutionary questions cannot be readily addressed in most model higher eukaryotes in which centromeres are occupied by a single family of satellite repeats because the DNA sequence of individual centromeres cannot be sequenced, tracked, and compared among related species.

We recently reported that the centromeres of potato (*Solanum tuberosum*) chromosomes are unique in DNA composition (Gong et al., 2012). Six of the 12 potato centromeres (*Cen1*, *Cen2*, *Cen3*, *Cen5*, *Cen7*, and *Cen8*) contain megabase-sized arrays of satellite repeats that are distinct to each of these six centromeres. By contrast, five potato centromeres (*Cen4*, *Cen6*, *Cen10*, *Cen11*, and *Cen12*) are composed of single- and low-copy DNA sequences, and no satellite repeats were detected in these centromeres. Thus, these five potato centromeres structurally resemble neocentromeres. The remaining centromere, *Cen9*, contains both single-copy sequences and satellite repeats (Gong et al., 2012). Most of the centromeric satellite repeats in potato were amplified recently from retrotransposon-related sequences and are not present in wild *Solanum* species closely related to potato (Gong et al., 2012). In addition, different centromeric haplotypes were found to be associated with three potato centromeres, including haplotypes containing megabase-sized satellite repeats and haplotypes that do not contain the same repeats (L.S. Wang et al., 2014). Thus, potato centromeric DNA appears to be under ongoing and rapid evolution. To further understand the evolution of centromeric DNA in *Solanum* species, we conducted genome-wide analysis of DNA sequences associated with the cenH3 nucleosomes in *Solanum verrucosum* (2n=2x=24), a wild species closely related to potato. Centromeric DNA sequences extracted from potato and *S. verrucosum* allowed for direct comparison of five pairs of homoeologous centromeres (*Cen2*, *Cen4*, *Cen7*, *Cen9*, and *Cen10*) in the two species. We demonstrate a rapid divergence of the centromeric sequences between these two closely related species. We hypothesize that centromeric satellite repeats may undergo boom-bust cycles before a favorable repeat is fixed in the population.

RESULTS

DNA Sequences Associated with cenH3 Nucleosomes in *S. verrucosum*

S. verrucosum is a wild species that is most closely related to potato and was proposed as the progenitor of cultivated potato (Hawkes, 1990). Sequence analysis of the mitochondrial genomes revealed that *S. verrucosum* is the maternal ancestor for all Mexican polyploid species (Sanetomo and Hosaka, 2013). Recent cytogenetic studies also confirmed that *S. verrucosum* is the possible "A" genome donor of polyploid *Solanum* species (Pendinen et al., 2012). Cultivated potato is an autotetraploid and contains four copies of the A genome (2n=4x=48, AAAA).

We developed an antibody against the cenH3 protein in potato (Gong et al., 2012). This antibody recognized *S. verrucosum* centromeres in high specificity (Figures 1A to 1C). We conducted chromatin immunoprecipitation (ChIP) in *S. verrucosum* using this antibody. The immunoprecipitated DNA was labeled as a probe in fluorescence in situ hybridization (FISH) on meiotic pachytene chromosomes. The centromeric regions, visible as the primary constrictions, can be unambiguously identified on *S. verrucosum* pachytene chromosomes because these regions were distinctly less stained by 4′,6-diamidino-2-phenylindole (DAPI) than

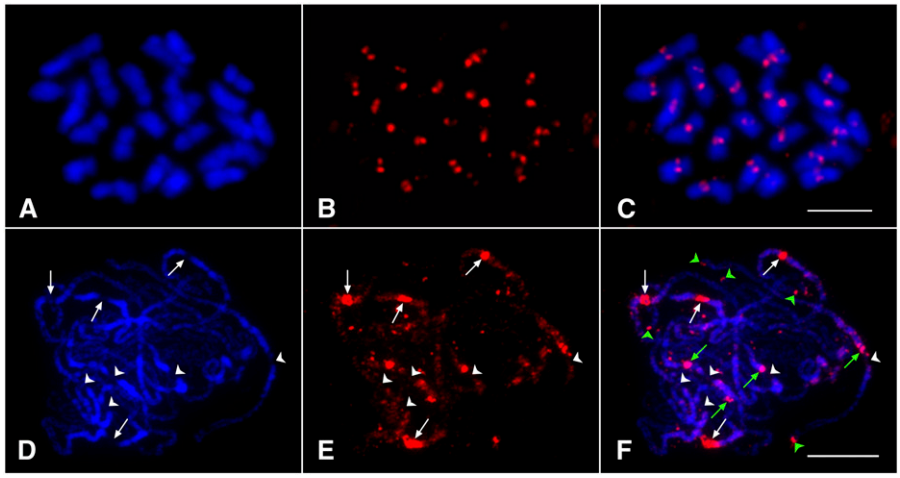

Figure 1. Immunofluorescence and ChIP-FISH Using Potato cenH3 Antibodies.

(A) Somatic metaphase chromosomes prepared from root tips of *S. verrucosum.*

(B) Immunofluorescence derived from anti-cenH3 antibodies.

(C) Image merged from **(A)** and **(B)**. Bar = 5 μm.

(D) Meiotic pachytene chromosomes from *S. verrucosum.* A total of four arrows and five arrowheads point to the primary constrictions that are distinctly less stained by DAPI.

(E) FISH signals derived from immunoprecipitated DNA. No major signals are associated with the five centromeres indicated by arrowheads. By contrast, major signals are associated with the four centromeres indicated by arrows.

(F) Image merged from **(D)** and **(E)**. Green arrows point to major FISH signals located in pericentromeric and interstitial regions. Green arrowheads point to FISH signals located at the end of the chromosomes. Bar = 10 μm.

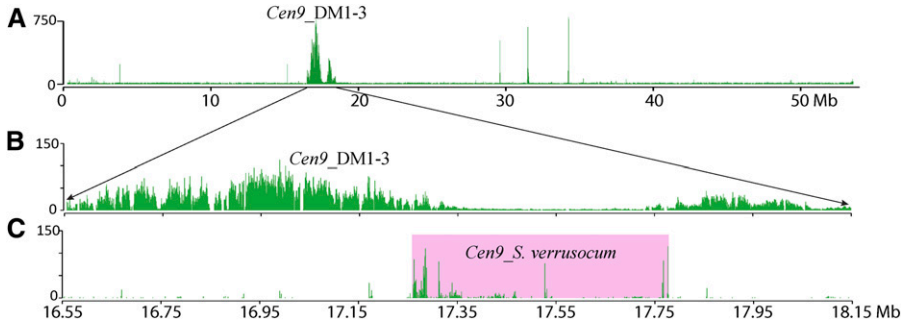

Figure 2. Identification of Sequences Associated with *Cen9* of *S. verrucosum*.

(A) Distribution of the density of cenH3 ChIP-seq reads from DM1-3 potato along chromosome 9 of the DM1-3 reference genome (read number per million reads in 10-kb windows). This ChIP-seq result was reported by Gong et al. (2012).
(B) The cenH3 binding region of *Cen9* of DM1-3 potato was mapped between 16.55 and 18.15 Mb.
(C) Distribution of the density of cenH3 ChIP-seq reads from *S. verrucosum* within *Cen9* of DM1-3. Significant enrichment of ChIP-seq reads was detected only at 17.28 to 17.78 Mb. The density was calculated as read number per million reads in 1-kb windows in **(B)** and **(C)**. Box in **(C)** marks the region with significant ChIP-seq read enrichment in *Cen9* of *S. verrucosum*.
[See online article for color version of this figure.]

the pericentromeric heterochromatin (Figure 1D). Strong ChIP-FISH signals were clearly visible on at least four centromeres (white arrows in Figures 1D to 1F). These strong FISH signals were likely derived from highly repetitive sequences associated with cenH3 nucleosomes because such centromeric repeats would be enriched in the immunoprecipitated DNA, thereby resulting in strong hybridizations. By contrast, unambiguous FISH signals were not detected on several other centromeres (white arrowheads in Figures 1D to 1F), suggesting that these centromeres are likely composed of single- or low-copy sequences. We observed similar ChIP-FISH signal patterns in cultivated potato (Gong et al., 2012). Thus, both potato and *S. verrucosum* appear to include centromeres containing highly repetitive sequences and centromeres composed of primarily low-copy sequences.

Punctuated ChIP-FISH signals were visible at the ends of most pachytene chromosomes (green arrowheads in Figure 1F), suggesting that *S. verrucosum* centromeres contain telomere-similar repeats, which were reported previously in several other *Solanum* species (Tek and Jiang, 2004; He et al., 2013). We also observed several major FISH signals in pericentromeric regions that immediately flank the primary constrictions or in interstitial regions (green arrows in Figure 1F). These signals are likely derived from repeats located both in and outside of the centromeres.

Homoeologous DNA Sequences Detected in *Cen9* of Potato and *S. verrucosum*

To characterize the cenH3-associated DNA sequences in *S. verrucosum*, we conducted Illumina sequencing of a DNA sample derived from ChIP (ChIP-seq). We generated 4.13 million paired-end sequence reads (100 nucleotides) and mapped 0.65 million reads to the potato reference genome, which was developed from the homozygous doubled monoploid clone DM1-3 (Xu et al., 2011). We did not detect any ChIP-seq sequence enrichment in 11 of the 12 potato chromosomes. However, enrichment was detected between 17.28 and 17.78 Mb in chromosome 9, which is located within the cenH3 binding domain of *Cen9* (16.55-18.15 Mb) of potato (Gong et al., 2012) (Figure 2).

Two factors may explain the lack of unique sequences mapped to the rest of the 11 potato centromeres. First, ChIP-FISH results showed that several *S. verrucosum* centromeres may contain exclusively repetitive DNA sequences (Figure 1). These centromeric repeats in the ChIP-seq data set will be excluded in the sequence mapping process. Similarly, six of the 12 centromeres

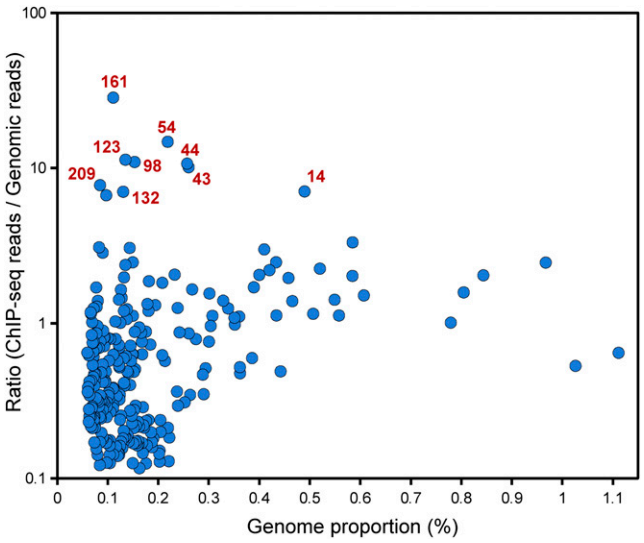

Figure 3. Identification of Repeat Clusters Associated with cenH3 Chromatin.

Repeat clusters are represented by dots and their positions reflect genomic abundance of corresponding repeats (*x* axis) and their enrichment in ChIP-seq data (*y* axis). The *y* axis is the ratio of ChIP-seq sequence reads to genomic sequence reads for each repeat cluster and is in logarithmic scale. The *x* axis is the genome proportion of the genomic sequence reads for each repeat cluster. Only the largest clusters representing repeats with genome proportions >0.06% are shown. Clusters with ChIP-seq enrichment >7-fold selected for cytological analysis are marked with numbers.
[See online article for color version of this figure.]

in DM1-3 contain only satellite repeats and these repeats will not be well assembled in the reference genome; therefore, the sequences derived from the same centromeres of *S. verrucosum* will not be detected. Second, the single- and low-copy sequences in potato and *S. verrucosum* centromeres may have been evolving rapidly. The originally homologous *S. verrucosum* centromeric sequences have significantly diverged from those of DM1-3 in the same centromeres and can no longer be mapped to the DM1-3 genome using a standard mapping process. Therefore, these results reveal a general picture of very rapid divergence of the centromeric DNA sequences in these two closely related species.

Computational Identification of Repetitive DNA Sequences Enriched in Centromeres

ChIP-FISH revealed that several *S. verrucosum* centromeres may contain long arrays of satellite repeats (Figures 1D to 1F). To isolate these repeats, we first generated shotgun Illumina reads (100 nucleotides) from genomic DNA of *S. verrucosum*. We used a similarity-based sequence clustering approach (Macas et al., 2007; Novák et al., 2010) for de novo identification of repetitive DNA sequences from five million randomly selected sequence reads. The sequence proportion (%) of each computationally identified repeat family was estimated based on the number of sequence reads associated with individual clusters. We then mapped the cenH3 ChIP-seq reads to the repeat clusters and calculated the ratio of ChIP-seq reads to genomic reads associated with each cluster (Figure 3). This ratio is indicative of the level of enrichment of each repeat family in the centromeres. Potential satellite repeats were then identified based on their reconstructed consensus sequences and structure of the cluster graphs (Novák et al., 2010). We selected nine clusters for further analysis. Each of these nine clusters contained an abundant repeat (with a sequence proportion >0.1% of the genome) and the putative repeat had a minimum ratio of "ChIP-seq reads/genomic reads" of 7 (Figure 3, Table 1).

Cytological Confirmation of the Computationally Identified Centromeric Repeats

Primers were designed from each of the putative centromeric repeats (Supplemental Table 1). A single band with a size similar to the computational prediction was amplified from the genomic DNA of *S. verrucosum* for all selected repeats. The amplified DNA fragments were cloned and confirmed to contain desired repeats by sequencing (see Table 1 for GenBank accession numbers). The repeat clones were then labeled for FISH analysis.

Several repeats were found to be specific to one or two centromeres. Both Sv44 (Figure 4A) and Sv14 (data not shown) hybridized specifically to *Cen4* (Table 2). Sv161 represents a very complex cluster consisting of several subclusters. Primers were designed from two subclusters, Sv161.5 and Sv161.6. Both DNA fragments were mapped to *Cen7* (Figure 4B, Table 2). Sv54 hybridized to two different centromeres: *Cen2* and *Cen10*. The Sv54 array in *Cen10* appeared to be much longer than the array in *Cen2* based on the size and intensity of the FISH signals (Figure 4C, Table 2). Sv123 also hybridized to both *Cen2* and *Cen10*. FISH signals from Sv54 and Sv123 overlapped on the two centromeres (Figure 4D).

Three different repeats hybridized to all centromeres. Sv98 produced strong hybridization signals in most centromeres with a few centromeres showing relatively weak signals (Figures 5A and 5B). Signals generated from Sv43 were much weaker than those from Sv98; however, the intensity of the Sv43 signals was similar among different centromeres (Figures 5C and 5D). The patterns of FISH signals derived Sv98 and Sv43 were highly

Table 1. Characteristics of Sequence Clusters Containing Putative Centromeric Repeats

Cluster[a]	Genomic Sequence Reads[b]	ChIP-seq[b]	ChIP/WGS Ratio	Repeat	Repeat Type (Similarity)	Probe [GenBank, Cloned Fragment][c]	Chromosomal Locations
14	0.490	3.458	7.1	Sv14	Tandem (LTR retrotransposon chromodomain)[d]	c1531 [KF561463, 972 bp]	Centromere 4
44	0.258	2.744	10.7	Sv44	Tandem[d]	c1535 [KF561465, 1025 bp]	Centromere 4
161	0.111	3.156	28.4	Sv161.5	St24-like	c1546 [KF561472, 555 bp]	Centromere 7
				Sv161.6	St57-like	c1548 [KF561473, 388 bp]	Centromere 7
54	0.219	3.226	14.7	Sv54.1	Unknown (LTR-RE-like sequence)	c1537 [KF561466, 837 bp]	Centromeres 2 + 10
				Sv54.2	Unknown (LTR-RE gag)	c1538 [KF561467, 1593 bp]	Centromeres 2 + 10
123	0.135	1.529	11.3	Sv123	Unknown	c1542 [KF561470, 827 bp]	Centromeres 2 + 10
43	0.260	2.631	10.1	Sv43	Tandem, likely scrambled sequence derived from retrotransposons	c1533 [KF561464, 1109 bp]	All centromeres
98	0.154	1.676	10.9	Sv98.1	Ty3/gypsy, CRM clade, gag-pol region	c1539 [KF561468, 1570 bp]	All centromeres
				Sv98.2	Ty3/gypsy, CRM clade, LTR region	c1540 [KF561469, 942 bp]	All centromeres
132	0.131	0.919	7.0	Sv132	Ty3/gypsy, Tekay clade	c1544 [KF561471, 1045 bp]	Dispersed in pericentromeres
209	0.085	0.659	7.7	Sv209	Telomere-like repeat	St49 [JQ731639, 2754 bp]	Multiple centromeres

[a]Clusters generated by analysis of the genomic sequence data set.
[b]Proportion (%) of all sequence reads.
[c]Clones of genomic fragments obtained by PCR with primers based on predicted consensus sequences.
[d]Sv14 and Sv44 represent parts of a single tandem repeat with monomer of 2.7 kb.

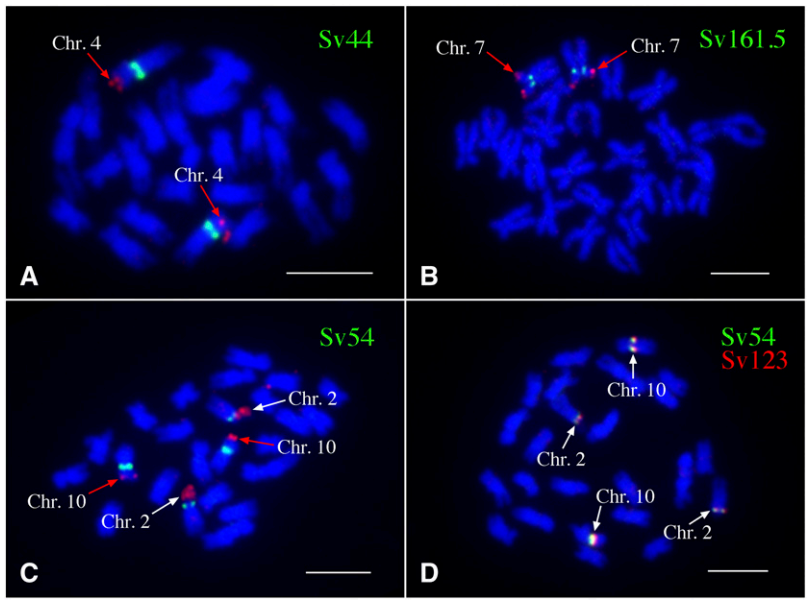

Figure 4. FISH Mapping of Centromeric Repeats in *S. verrucosum.*

(A) Repeat Sv44 was mapped to *Cen4* together with BAC clone 162B09 (arrows), which is specific to the short arm of chromosome 4.
(B) Repeat Sv161.5 was mapped to *Cen7* together with BAC clone 186I02, which is specific to the short arm of chromosome 7.
(C) Repeat Sv54 was mapped to *Cen2* and *Cen10* together with the 45S rDNA probe (white arrows), which is specific to chromosome 2, and BAC 178K07 (red arrows), which is specific to the short arm of chromosome 10.
(D) Repeats Sv123 (red) and Sv54 (green) were mapped to the same two centromeres. Note: The acrocentric chromosome 2 can be readily identified by the locations of Sv123/Sv54 toward the end of the chromosome.
Bars = 5 μm.

similar to those associated with the centromeric retrotransposon CRR in rice (Cheng et al., 2002b; Nagaki et al., 2005). Thus, these repeats may have spread to all centromeres via mechanisms similar to the CR retrotransposons reported in various plant species (Nagaki et al., 2005; Du et al., 2010; Neumann et al., 2011). Sv132 produced weak and dispersed signals in both centromeric and pericentromeric regions of all chromosomes (Figures 5E and 5F).

The Sv209 repeat is a degenerate telomeric sequence and shows sequence similarity to the St49 repeat in potato. PCR primers were difficult to design from such short and degenerate sequences. FISH using the St49 probe from potato (Gong et al., 2012) showed that about half of the *S. verrucosum* centromeres contain a considerable amount of telomere-similar sequences (Figures 5G and 5H).

Table 2. Centromeric DNA Sequences in Potato (*S. tuberosum*) and *S. verrucosum*

Centromere	Potato	*S. verrucosum*
Cen1	St24 repeat[a]	Sequence data not available
Cen2	St3.58 repeat	Sv54, Sv123 repeat
Cen3	St3.294 repeat	Sequence data not available
Cen4	Single- and low-copy sequences	Sv44, Sv14 repeats
Cen5	St49	Sequence data not available
Cen6	Single- and low-copy sequences	Sequence data not available
Cen7	St57	St57 (Gong et al., 2012), Sv161.5, Sv161.6 repeats[b]
Cen8	St3.238	Sequence data not available
Cen9	St18, St3.294 repeats, single- and low-copy sequences[c]	Single- and low-copy sequences
Cen10	Single- and low-copy sequences	Sv54, Sv123 repeats[d]
Cen11	Single- and low-copy sequences	Sequence data not available
Cen12	Single- and low-copy sequences	Sequence data not available

[a]The St24 repeat is homologous to the Sv161.5 repeat in *S. verrucosum*.
[b]Sv161.6 is homologous to the St57 repeat in potato.
[c]The St18 repeat was detected in the centromere of an unidentified *S. verrucosum* chromosome, which is not *Cen9* (Gong et al., 2012).
[d]The Sv54 repeat hybridized to potato chromosome 9 (Figure 7I).

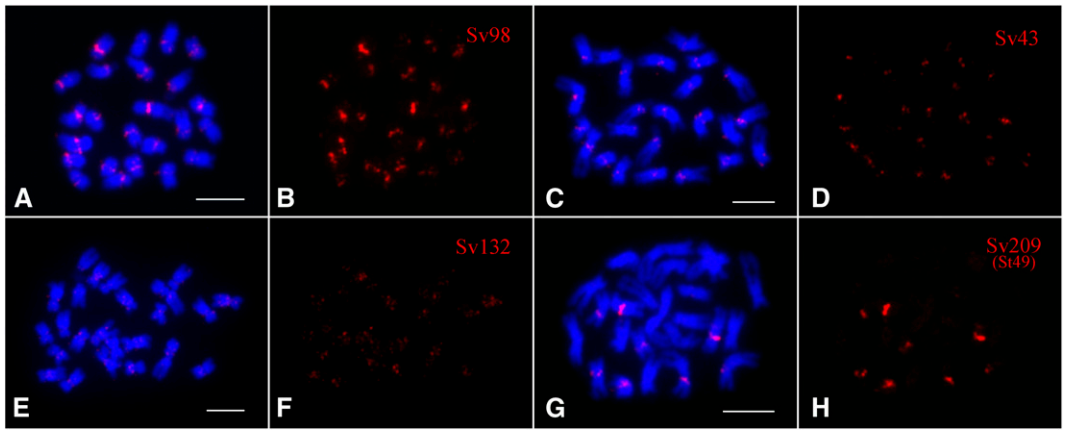

Figure 5. FISH Mapping of Repeats Associated with Multiple Centromeres in *S. verrucosum*.

(A) Sv98 hybridized to all centromeres with different signal sizes and intensities.
(B) Digitally separated FISH signals from **(A)**.
(C) Sv43 hybridized all centromeres with a similar signal size and intensity.
(D) Digitally separated FISH signals from **(C)**.
(E) Sv132 generated weak and dispersed FISH signals in the centromeric and pericentromeric regions of all chromosomes.
(F) Digitally separated FISH signals from **(E)**.
(G) FISH of Sv209 on metaphase chromosomes of *S. verrucosum*. Probe St49, which is homologous to Sv209, was used as the FISH probe.
(H) Digitally separated FISH signals from **(G)**.
Bars = 5 μm.

The Structure of the Centromeric Satellite Repeats Revealed by Fiber-FISH

The overlapping FISH signals derived Sv44-Sv14, Sv54-Sv123, and Sv161.5-Sv161.6 probe pairs indicated that each pair of repeats is likely derived from different parts of the same repeat monomer but were separated into different clusters or subclusters during the computational analysis. We conducted dual-color fiber-FISH experiments to reveal the relationship of these three pairs of DNA probes. The fiber-FISH signals from Sv44 and Sv14 were always associated with the same DNA fibers (Figure 6A). DNA fibers associated exclusively with Sv44 or Sv14 were not observed. These results confirmed that Sv44 and Sv14 represent different parts of the same satellite repeat monomer. Similar fiber-FISH results were obtained for the Sv54-Sv123 and Sv161.5-Sv161.6 probe pairs (Figures 6B and 6C).

The fiber-FISH signals derived from the Sv14 and Sv44 probe pairs were primarily overlapping, frequently resulting in yellow fluorescence dots on DNA fibers (Figure 6A). This result suggested that these two DNA fragments are closely adjacent to

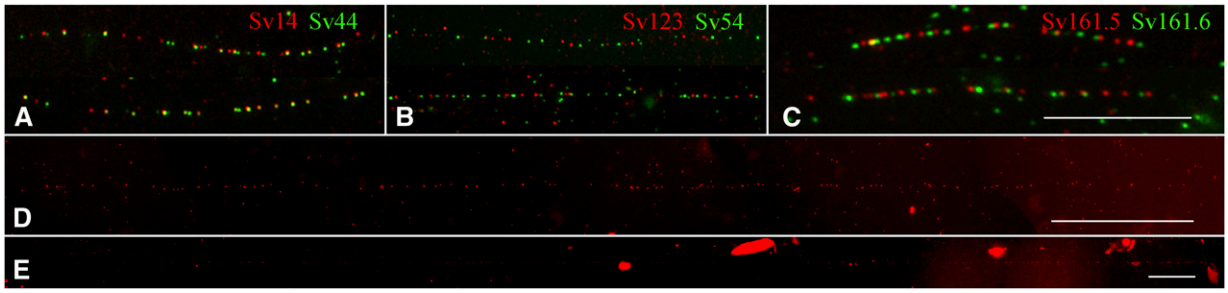

Figure 6. Fiber-FISH Mapping of the Centromeric Satellite Repeats in *S. verrucosum*.

(A) Two partial fiber-FISH signals derived probes Sv14 (red) and Sv44 (green). Note: Most fluorescence signal spots derived from the two probes overlapped.
(B) Two partial fiber-FISH signals derived probes Sv123 (red) and Sv54 (green). Note: Most fluorescence signal spots derived from the two probes did not overlap.
(C) Two complete fiber-FISH signals derived probes Sv161.5 (red) and Sv161.6 (green). Bar = 10 μm.
(D) A complete fiber-FISH signal derived probes Sv54/Sv123. Bar = 50 μm.
(E) A complete fiber-FISH signal derived probes Sv14/Sv44. Bar = 50 μm.
The original images in **(D)** and **(E)** were derived from dual-color fiber-FISH using two probes. The signals from the two probes were merged, converted into gray images, and digitally straightened. The straightened images were then pseudocolored into red for better contrast.

each other within the same repeat monomer. This prediction was confirmed by PCR amplification of an ~2-kb fragment by a combination of Sv14 reverse primer Svr14-R1 and Sv44 reverse primer Svr44-R1 (Supplemental Table 1). In comparison, the fiber-FISH signals derived from the Sv123 and Sv54 probe pairs rarely overlapped (Figure 6B). The densities of the fluorescence spots of the fiber-FISH signals derived from these two pairs of probes were lower than those derived from contiguous DNA sequences, suggesting that the true monomer associated with these two repeats contain additional sequences that yet to be cloned.

Fiber-FISH also allowed direct measurements of the lengths of the three centromeric satellite repeat arrays. The Sv161.5/Sv161.6 repeat array spanned only 50.3 ± 9.8 kb (n = 20) (Figure 5C). The Sv54/Sv123 repeats generated a large FISH signal on *Cen10* and a smaller signal on *Cen2* (Figure 4C). Fiber-FISH signals derived from the *Cen2* array cannot be unambiguously identified because each short signal could be broken from a long signal derived from the *Cen10* array. Measurements on several long fiber-FISH signals revealed that the long Sv54/Sv123 array spanned 1442.6 ± 248.8 kb (n = 9) (Figure 6D). The Sv14/Sv44 array was measured to be several megabases, which reached the upper limit of fiber-FISH detection. The average length of the four longest fiber-FISH signals from the Sv14/Sv44 probes was 4017.1 kb, ranging from 3334.9 to 4591.7 kb (Figure 6E).

The cenH3 binding domains of potato centromeres range from 1063 to 2404 kb (Gong et al., 2012). Thus, we predict that the Sv54/Sv123 array in *Cen10* and Sv14/Sv44 in *Cen4* likely span the entire cenH3 binding domains of these two centromeres.

Origin of the Centromeric Repeats

Several of the identified centromeric repeats showed sequence similarities to Ty3/gypsy retrotransposons. However, only Sv98 was found to have a typical LTR-retrotransposon structure. This element represented a member of the centromere-targeted CRM clade of retrotransposons (Neumann et al., 2011). Other repeats appeared to be amplified from scrambled retrotransposon-related sequences (Table 1; Supplemental Figure 1). This was, for example, evident for the Sv14/Sv44 sequences that proved to be parts of a single tandem repeat with monomer length of ~2.7 kb and showed partial similarity to a LTR retrotransposon chromodomain sequence (Supplemental Figure 1). The Sv161 repeat shared sequence similarity with the St24 and St57 satellites from potato, both amplified from retroelements (Gong et al., 2012). Partial retrotransposon similarities and similar graph structures were also observed for repeats Sv54 and Sv123. These results suggest that most of the centromeric repeats in *S. verrucosum* were amplified directly from retrotransposons or from scrambled sequences originally derived from retrotransposons.

Evolution of the Centromeric Repeats in *Solanum* Species

To reveal the evolution of the centromeric repeats identified in *S. verrucosum*, we selected a set of four diploid *Solanum* species (Table 3) for FISH mapping. DM1-3, representing cultivated potato, is a homozygous doubled monoploid clone that has been fully sequenced (Xu et al., 2011). *Solanum jamesii* (B genome species), *Solanum chomatophilum* (P genome species), and *Solanum palustre* (E genome species) are wild diploid potato species with different evolutionary distances from cultivated potato (Lou et al., 2010).

Sv14/Sv44 did not generate FISH signals in any of the species examined (Table 3). Sv161.6 hybridized to *Cen7* in both *S. verrucosum* and DM1-3 but did not hybridize to any other species (Table 3). By contrast, Sv161.5 hybridized to *Cen7* in *S. verrucosum* but to *Cen1* in DM1-3 (Figure 7C). Sv161.5 did not hybridize to

Table 3. Summary of FISH Analysis of the Centromeric Repeats in *Solanum* Species

Repeat	*S. verrucosum* (A Genome)	DM1-3 Potato (A Genome)	*S. chomatophilum* (P Genome)	*S. jamesii* (B Genome)	*S. palustre* (E Genome)
Sv14	Specific to *Cen4*	No hybridization	No hybridization	No hybridization	No hybridization
Sv44	Specific to *Cen4*	No hybridization	No hybridization	No hybridization	No hybridization
Sv161.5 (St24)	Specific to *Cen7*	Specific to *Cen1*	No hybridization	No hybridization	Weak signals in multiple centromeres
Sv161.6 (St57)	Specific to *Cen7*	Specific to *Cen7*	No hybridization	No hybridization	No hybridization
Sv54	Specific to *Cen2* and *Cen10*	Specific to *Cen9*	Single centromeric signal on *Cen2*	No hybridization	No hybridization
Sv123	Specific to *Cen2* and *Cen10*	No hybridization	Three centromeric signals, on *Cen1*, *Cen2*, and another unknown	No hybridization	No hybridization
Sv98	Centromeric signals in all chromosomes but varied in size and intensity in different centromeres	Centromeric signals in all chromosomes but varied in size and intensity in different centromeres	No hybridization	No hybridization	No hybridization
Sv43	Centromeric signals (weak) in almost all chromosomes	Centromeric signals (weak) in almost all chromosomes	Centromeric signals (weak) in multiple (~7) chromosomes	No hybridization	No hybridization
Sv132	Very weak signals in most centromeres and pericentromeres	Very weak signals in most centromeres and pericentromeres	No hybridization	Very weak signals in most centromeres and pericentromeres	No hybridization

S. jamesii or *S. chomatophilum* chromosomes but generated very weak signals in several unidentified *S. palustre* chromosomes (Figure 7F). Sv54, specific to both *Cen2* and *Cen10* in *S. verrucosum*, hybridized to *Cen9* in DM1-3 (Figure 7I) and to one copy of *Cen2* in *S. chomatophilum* (Figure 7L). Sv123, by contrast, did not hybridize to DM1-3, but hybridized to one copy of *Cen2* and one copy of *Cen1* in *S. chomatophilum* (Figure 7O). Very weak FISH signals were also observed in several other centromeres in *S. chomatophilum.*

Repeats Sv98, Sv43, and Sv132, which hybridize to all centromeres in *S. verrucosum*, showed similar FISH signal patterns in DM1-3. However, the hybridization of these repeats in other species was either extremely weak or not detectable (Table 3).

In summary, most centromeric repeats in *S. verrucosum* emerged recently, since these repeats were not detected in other wild species with different genomes. The massive Sv14/Sv44 array is specific to *S. verrucosum*. The Sv161 and Sv54 repeats appeared to have emerged in potato and *S. verrucosum* independently,

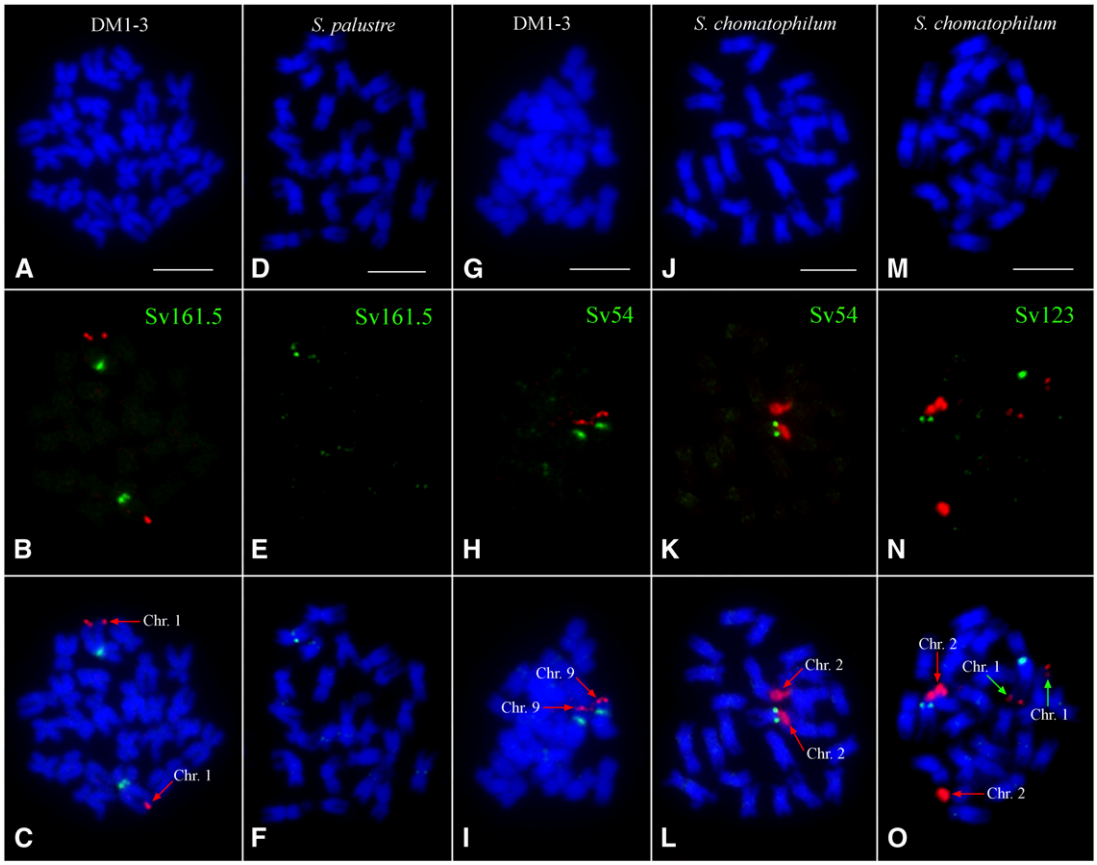

Figure 7. FISH Mapping of *S. verrucosum* Centromeric Repeats in Different *Solanum* Species.

(A) A somatic metaphase cell from DM1-3 potato.
(B) FISH signals derived from Sv161.5 (green) and chromosome 1–specific BAC 96H03 (red).
(C) Image merged from **(A)** and **(B)**. Arrows point to signals from BAC 96H03.
(D) A somatic metaphase cell from *S. palustre*.
(E) FISH signals derived from Sv161.5 (green).
(F) Image merged from **(D)** and **(E)**.
(G) A somatic metaphase cell from DM1-3 potato.
(H) FISH signals derived from Sv54 (green) and chromosome 9–specific BAC 135I22 (red).
(I) Image merged from **(G)** and **(H)**. Arrows point to signals from BAC 135I22.
(J) A somatic metaphase cell from *S. chomatophilum*.
(K) FISH signals derived from Sv54 (green) and 45S rDNA (red) that is specific to chromosome 2.
(L) Image merged from **(J)** and **(K)**. Arrows point to signals from 45S rDNA. Note: Sv54 is detected only on one of the two copies of chromosome 2.
(M) A somatic metaphase cell from *S. chomatophilum*.
(N) FISH signals derived from Sv123 (green), 45S rDNA (red), and chromosome 1–specific BAC 96H03 (red).
(O) Image merged from **(M)** and **(N)**. Red arrows point signals from 45S rDNA. Green arrows point to signals from BAC 96H03.
Bars = 5 μm.

since these repeats (Sv161.5 and Sv54) hybridized to different centromeres in the two species.

DISCUSSION

We previously demonstrated that *Cen4* and *Cen10* of potato contain exclusively single- and low-copy sequences. No satellite repeats were discovered in these two centromeres (Gong et al., 2012). In comparison, both *Cen4* and *Cen10* in *S. verrucosum* contain megabase-sized satellite repeat arrays (Figure 4). These results suggest two possible evolutionary paths for these two centromeres. The ancestral *Cen4* and *Cen10*, which predate the two species, were likely composed of single and low copy sequences. The Sv44/Sv14 and Sv54/Sv123 repeats invaded *Cen4* and *Cen10*, respectively, in *S. verrucosum* after the divergence of the two species, whereas *Cen4* and *Cen10* of potato have maintained the single/low copy DNA composition (Figure 8A). Alternatively, the ancestral *Cen4* and *Cen10* contained the Sv44/Sv14 and Sv54/Sv123 repeats, respectively. These repeats were deleted in potato *Cen4* and *Cen10* after the divergence of the two species. Although centromeric satellite repeat arrays can expand and contract, resulting in different sizes of repeat arrays in different genotypes (Albert et al., 2010), deletion of the entire repeat array in a centromere would be lethal to centromere function or result in neocentromere activation (Ishii et al., 2008; Ketel et al., 2009; Shang et al., 2013). Thus, complete deletion of both repeats (Sv44/Sv14 and Sv54/Sv123) in both *Cen4* and *Cen10* in potato would be highly unlikely.

Potato *Cen2* contains the St3.58 satellite repeat array that spans ~1500 kb (Gong et al., 2012), whereas *Cen2* of *S. verrucosum* contains the Sv54/Sv123 repeat. Thus, the ancestral *Cen2* has been invaded by different satellite repeats since the divergence of the two species (Figure 8B). Interestingly, St3.58 is specific to *Cen2* in potato, whereas Sv54/Sv123 hybridized to both *Cen2* and *Cen10*. It is likely that the Sv54/Sv123 repeat originally emerged in one centromere and has since spread to a second chromosome.

A single satellite repeat is found in all centromeres in many animal and plant species. For example, all human centromeres contain the 171-bp α-satellite repeat, ranging from several hundreds of kilobases to a few megabases (Willard and Waye, 1987; Vafa and Sullivan, 1997). However, several species contain multiple satellite repeats associated with different centromeres, including chicken (*Gallus gallus*) (Shang et al., 2010), pea (*Pisum sativum*) (Neumann et al., 2012), potato (Gong et al., 2012), and common bean (*Phaseolus vulgaris*) (Iwata et al., 2013). Our results from the two *Solanum* species are in favor of the concept that a centromeric satellite repeat may emerge initially in a single centromere and then spread to all centromeres. However, it is unknown how a satellite repeat spread to different centromeres, which cannot be explained by the classical unequal crossing over mechanisms since centromeric regions lack chromosomal crossovers (Yan et al., 2008). It is interesting to note that at least one of the satellite repeats, Sv54/Sv123, has spread to two different centromeres in *S. verrucosum*. Similarly, most of the macrochromosomes in chickens contain a distinct satellite repeat. However, a 42-bp satellite repeat has spread to the centromeres of two macrochromosomes and several microchromosomes (Shang et al., 2010).

It has been discovered recently that the CENP-A/cenH3 nucleosomes are highly phased with the 171-bp monomers of the α-satellite repeats in humans (Hasson et al., 2013) and with the 155-bp CentO satellite repeats in rice (Zhang et al., 2013). In addition, the CentO repeats have an ~10-bp periodicity in A/T dinucleotide pattern and in nuclease cleavage, suggesting that CentO has evolved to minimize its bending energy on cenH3 nucleosomes (Zhang et al., 2013). Thus, some satellite repeats, such as the CentO repeat in rice, have evolved into a structure that can confer both translational and rotational phasing on cenH3 nucleosomes, which would be favorable for stabilization of the centromeric nucleosomes. Such repeats would be selectively maintained in the centromeres. The human α-satellite repeat has been found in both old and new world monkey species (Willard, 1991); thus, this repeat has occupied the centromeres of primate species for nearly 40 million years (Horvath and Willard, 2007), while no other satellite repeats have conserved among all of these primate species. Similarly, the centromeric satellite repeats from several distantly related grass species, including CentO from rice and the 156-bp CentC repeat from maize (*Zea mays*; Ananiev et al., 1998), share sequence similarity within an 80-bp region (Lee et al., 2005).

It is intriguing that some plant and animal species contain multiple centromeric repeats. Interestingly, most of the centromeric satellite repeats identified in chicken, and both potato and *S. verrucosum* do not have the structure ideal for cenH3 nucleosome organization. In

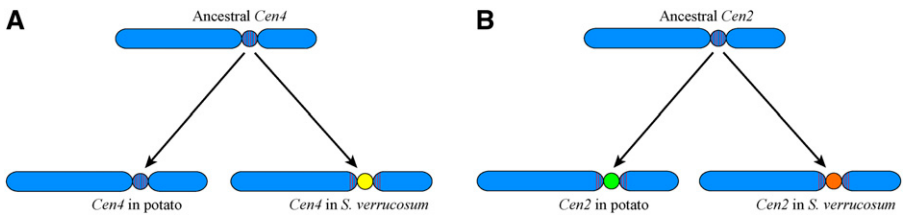

Figure 8. Models of Centromere Evolution in *Solanum* Species.

(A) The ancestral centromere of chromosome 4 of *Solanum* species contained only single- and low-copy sequences, including low density of genes (red vertical bars). Potato *Cen4* has mainly maintained this DNA composition. However, a satellite repeat (yellow) has invaded the *Cen4* of *S. verrucosum*, which pushed the original centromeric sequences into the pericentromeric regions.

(B) The ancestral centromere of chromosome 2 of *Solanum* species contained only single- and low-copy sequences. This centromere has been invaded by different satellite repeats (green and brown, respectively) in both species.

fact, none of these repeats has a monomer size typical for a single nucleosome (150 to 180 bp). The monomers of the potato centromeric repeats range from 979 bp to >5 kb (Gong et al., 2012). The monomeric sizes of the *S. verrucosum* centromeric repeats have not been determined, but they are >390 bp. We hypothesize that centromeric satellite repeats may undergo boom-bust cycles of evolution. A structurally favorable repeat will eventually emerge and will spread to other centromeres via poorly understood mechanisms. Such a repeat, such as the α-satellite repeat in humans, would then be fixed in the population and may survive for an extremely long period of time in evolution.

METHODS

Plant Materials

Solanum verrucosum (2n=2x=24, PI275260 and PI570643), a diploid wild potato species, was used for ChIP and cytogenetic studies. A homozygous doubled monoploid (2n=2x=24) clone, DM1-3 516R44 (DM1-3), was used to represent cultivated potato (*Solanum tuberosum*) for evolutionary studies. Three wild diploid *Solanum* species, including *Solanum jamesii* (B genome, PI620869), *Solanum chomatophilum* (P genome, PI243340), and *Solanum palustre* (E genome, PI558245), were also used in evolutionary studies. Seeds of all wild *Solanum* species were obtained from the USDA/ARS Potato Introduction Station, Sturgeon Bay, WI.

ChIP, ChIP-seq, and Mapping of ChIP-seq Reads to the Potato Genome

ChIP and ChIP-seq were performed as previously described (Gong et al., 2012). Sequence reads generated from ChIP-seq were aligned to the recently released potato genome sequence map derived from the DM1-3 clone (PGSC_DM_v3_2.1.10_pseudomolecule downloaded from http://potatogenomics.plantbiology.msu.edu/) using Bowtie (Langmead et al., 2009).

Repeat Identification and Characterization

Similarity-based clustering, repeat identification, and classification in a set of randomly selected five million paired-end whole-genome shotgun (WGS) Illumina reads was performed using RepeatExplorer pipeline (Novák et al., 2013). Investigation of cluster graphs was performed using the SeqGrapheR program (Novák et al., 2010). To identify repeats associated with cenH3 nucleosomes, the ChIP-seq reads were mapped to contigs representing assembled cluster reads using BLASTn (Altschul et al., 1997) with the parameters -m 8 -b 1 -e 1e-8 -W 9 -r 2 -q -3 -G 5 -E 2 -F F. Each ChIP-seq read was mapped to a maximum of one cluster, based on its best similarity detected among WGS contigs. Cloning of selected centromeric repeats was done using PCR primers designed according to their reconstructed consensus sequences derived from contigs assembled for each repeat cluster (Supplemental Table 1).

PCR Amplification of Computationally Identified Repeats

PCR amplification was performed in a 25-μL reaction mixture containing 50 to 100 ng template DNA, 1 μM each primer, 0.5 mM deoxynucleotide triphosphate mix, 2.5× PCR buffer, and 0.625 unit *Taq* DNA polymerase. The thermal cycling for PCR comprised 35 cycles, each with 0.30-min denaturation at 95°C, 0.45 -min annealing at 54°C, and an extension of 1 min at 72°C. An initial denaturation (3 min at 95°C) was preceded and followed by a final extension step (7 min at 72°C). The PCR products were electrophoresed on 1.0% agarose gels and then purified using a MinElute PCR purification kit (Qiagen).

FISH, Fiber-FISH, and Immunofluorescence

Preparation of mitotic and meiotic chromosomes, FISH, and fiber-FISH were performed following published protocols (Jackson et al., 1998; Dong et al., 2000; Lou et al., 2010). DNA probes for each centromere-specific satellite repeat were amplified by PCR from the *S. verrucosum* genomic DNA. Primers were designed from bioinformatically extracted repeat cluster (Supplemental Table 1). The amplified DNAs were labeled with either biotin-16-UTP or digoxigenin-11-dUTP (Roche Diagnostics) using a standard nick translation reaction. Chromosomes were counterstained with DAPI in Vectashield antifade solution (Vector Laboratories). The FISH images were processed with Meta Imaging Series 7.5 software. The final contrast of the images was processed using Adobe Photoshop CS3 software. The cytological measurements of the fiber-FISH signals were converted into kilobases using a 3.21 kb/μm conversion rate (Cheng et al., 2002a).

Root tips harvested from plants were fixed in 4% (w/v) paraformaldehyde for 15 min at room temperature. The root tips were washed by 1× PBS three times, each for 5 min, and were squashed on glass slides. After removal of the cover slip, the slides were dehydrated using ethanol (70, 90, and 100%) and then incubated overnight in a humidity chamber at 37°C with the rabbit primary sera antibody against potato cenH3 diluted 1:500 in TNB buffer (0.1 M Tris-HCl, pH 7.5, 0.15 M NaCl, and 0.5% blocking reagent). After three rounds of washing in 1× PBS, the slides were incubated with Cy3-conjugated goat anti-rabbit antibody (1:1000) at 37°C for 1 h. After three rounds of washing in 1× PBS, the slides dried at room temperature and the chromosomes were counterstained with DAPI.

Accession Numbers

Sequence data for the centromeric satellite repeat sequences can be found in the GenBank data library under accession numbers KF561463 to KF561473.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure 1. Graphical Representations of Repeat Clusters Enriched in ChIP-seq Data.

Supplemental Table 1. PCR Primers Used for Cloning Satellite Repeats in *S. verrucosum.*

ACKNOWLEDGMENTS

This work was supported by grants IOS-1237969 from the National Science Foundation to C.R.B. and J.J., Grant DBI-0922703 to J.J., and Grants LH11058 from the Ministry of Education, Youth, and Sport and RVO:60077344 from the Academy of Sciences of the Czech Republic to J.M.

AUTHOR CONTRIBUTIONS

J.J. and J.M. designed the research. H.Z., A.K., K.W., Z.G., L.O., G.A.T., and W.Z. performed the research. Y.W., P.N., J.M., and J.J. analyzed the data. J.J., J.M., and C.R.B. wrote the article. J.M. and J.J. are joint senior authors who contributed to this project equally.

Received February 2, 2014; revised March 14, 2014; accepted March 26, 2014; published April 11, 2014.

REFERENCES

Albert, P.S., Gao, Z., Danilova, T.V., and Birchler, J.A. (2010). Diversity of chromosomal karyotypes in maize and its relatives. Cytogenet. Genome Res. **129:** 6–16.

Altschul, S.F., Madden, T.L., Schäffer, A.A., Zhang, J.H., Zhang, Z., Miller, W., and Lipman, D.J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. **25:** 3389–3402.

Ammiraju, J.S.S., et al. (2008). Dynamic evolution of oryza genomes is revealed by comparative genomic analysis of a genus-wide vertical data set. Plant Cell **20:** 3191–3209.

Ananiev, E.V., Phillips, R.L., and Rines, H.W. (1998). Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions. Proc. Natl. Acad. Sci. USA **95:** 13073–13078.

Bensasson, D. (2011). Evidence for a high mutation rate at rapidly evolving yeast centromeres. BMC Evol. Biol. **11:** 211.

Cheng, Z.K., Buell, C.R., Wing, R.A., and Jiang, J. (2002a). Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. Chromosome Res. **10:** 379–387.

Cheng, Z.K., Dong, F.G., Langdon, T., Ouyang, S., Buell, C.R., Gu, M.H., Blattner, F.R., and Jiang, J.M. (2002b). Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell **14:** 1691–1704.

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

Dong, F.G., Miller, J.T., Jackson, S.A., Wang, G.L., Ronald, P.C., and Jiang, J.M. (1998). Rice (*Oryza sativa*) centromeric regions consist of complex DNA. Proc. Natl. Acad. Sci. USA **95:** 8135–8140.

Du, J.C., Tian, Z.X., Hans, C.S., Laten, H.M., Cannon, S.B., Jackson, S.A., Shoemaker, R.C., and Ma, J.X. (2010). Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. Plant J. **63:** 584–598.

Gao, D.Y., et al. (2009). A lineage-specific centromere retrotransposon in *Oryza brachyantha.* Plant J. **60:** 820–831.

Gong, Z.Y., Wu, Y.F., Koblízková, A., Torres, G.A., Wang, K., Iovene, M., Neumann, P., Zhang, W.L., Novák, P., Buell, C.R., Macas, J., and Jiang, J.M. (2012). Repeatless and repeat-based centromeres in potato: implications for centromere evolution. Plant Cell **24:** 3559–3574.

Hasson, D., Panchenko, T., Salimian, K.J., Salman, M.U., Sekulic, N., Alonso, A., Warburton, P.E., and Black, B.E. (2013). The octamer is the major form of CENP-A nucleosomes at human centromeres. Nat. Struct. Mol. Biol. **20:** 687–695.

Hawkes, J.G. (1990). The Potato: Evolution, Biodiversity and Genetic Resources. (London: Smithsonian Institution Press).

He, L., Liu, J., Torres, G.A., Zhang, H.Q., Jiang, J.M., and Xie, C.H. (2013). Interstitial telomeric repeats are enriched in the centromeres of chromosomes in *Solanum* species. Chromosome Res. **21:** 5–13.

Henikoff, S., Ahmad, K., and Malik, H.S. (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. Science **293:** 1098–1102.

Horvath, J.E., and Willard, H.F. (2007). Primate comparative genomics: lemur biology and evolution. Trends Genet. **23:** 173–182.

Ishii, K., Ogiyama, Y., Chikashige, Y., Soejima, S., Masuda, F., Kakuma, T., Hiraoka, Y., and Takahashi, K. (2008). Heterochromatin integrity affects chromosome reorganization after centromere dysfunction. Science **321:** 1088–1091.

Iwata, A., et al. (2013). Identification and characterization of functional centromeres of the common bean. Plant J. **76:** 47–60.

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

Jiang, J.M., Birchler, J.A., Parrott, W.A., and Dawe, R.K. (2003). A molecular view of plant centromeres. Trends Plant Sci. **8:** 570–575.

Kalitsis, P., and Choo, K.H.A. (2012). The evolutionary life cycle of the resilient centromere. Chromosoma **121:** 327–340.

Ketel, C., Wang, H.S.W., McClellan, M., Bouchonville, K., Selmecki, A., Lahav, T., Gerami-Nejad, M., and Berman, J. (2009). Neocentromeres form efficiently at multiple possible loci in *Candida albicans.* PLoS Genet. **5:** e1000400.

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

Lee, H.R., Zhang, W.L., Langdon, T., Jin, W.W., Yan, H.H., Cheng, Z.K., and Jiang, J.M. (2005). Chromatin immunoprecipitation cloning reveals rapid evolutionary patterns of centromeric DNA in *Oryza* species. Proc. Natl. Acad. Sci. USA **102:** 11793–11798.

Lou, Q.F., Iovene, M., Spooner, D.M., Buell, C.R., and Jiang, J.M. (2010). Evolution of chromosome 6 of *Solanum* species revealed by comparative fluorescence in situ hybridization mapping. Chromosoma **119:** 435–442.

Macas, J., Neumann, P., and Navrátilová, A. (2007). Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula.* BMC Genomics **8:** 427.

Marshall, O.J., Chueh, A.C., Wong, L.H., and Choo, K.H.A. (2008). Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution. Am. J. Hum. Genet. **82:** 261–282.

Melters, D.P., et al. (2013). Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. **14:** R10.

Nagaki, K., Neumann, P., Zhang, D., Ouyang, S., Buell, C.R., Cheng, Z.K., and Jiang, J.M. (2005). Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol. Biol. Evol. **22:** 845–855.

Neumann, P., Navrátilová, A., Koblížková, A., Kejnovský, E., Hřibová, E., Hobza, R., Widmer, A., Doležel, J., and Macas, J. (2011). Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob. DNA **2:** 4.

Neumann, P., Navrátilová, A., Schroeder-Reiter, E., Koblížková, A., Steinbauerová, V., Chocholová, E., Novák, P., Wanner, G., and Macas, J. (2012). Stretching the rules: monocentric chromosomes with multiple centromere domains. PLoS Genet. **8:** e1002777.

Novák, P., Neumann, P., and Macas, J. (2010). Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics **11:** 378.

Novák, P., Neumann, P., Pech, J., Steinhaisl, J., and Macas, J. (2013). RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics **29:** 792–793.

Pendinen, G., Spooner, D.M., Jiang, J.M., and Gavrilenko, T. (2012). Genomic in situ hybridization reveals both auto- and allopolyploid origins of different North and Central American hexaploid potato (*Solanum* sect. *Petota*) species. Genome **55:** 407–415.

Rocchi, M., Archidiacono, N., Schempp, W., Capozzi, O., and Stanyon, R. (2012). Centromere repositioning in mammals. Heredity (Edinb) **108:** 59–67.

Sanetomo, R., and Hosaka, K. (2013). A recombination-derived mitochondrial genome retained stoichiometrically only among *Solanum verrucosum* Schltdl. and Mexican polyploid wild potato species. Genet. Resour. Crop Evol. **60:** 2391–2404.

Shang, W.-H., et al. (2013). Chromosome engineering allows the efficient isolation of vertebrate neocentromeres. Dev. Cell **24:** 635–648.

Shang, W.H., Hori, T., Toyoda, A., Kato, J., Popendorf, K., Sakakibara, Y., Fujiyama, A., and Fukagawa, T. (2010). Chickens possess centromeres with both extended tandem repeats and short non-tandem-repetitive sequences. Genome Res. **20:** 1219–1228.

Tek, A.L., and Jiang, J.M. (2004). The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. Chromosoma **113:** 77–83.

Xu, X., et al; Potato Genome Sequencing Consortium (2011). Genome sequence and analysis of the tuber crop potato. Nature **475:** 189–195.

Vafa, O., and Sullivan, K.F. (1997). Chromatin containing CENP-A and alpha-satellite DNA is a major component of the inner kinetochore plate. Curr. Biol. **7:** 897–900.

Wang, K., Wu, Y.F., Zhang, W.L., Dawe, R.K., and Jiang, J.M. (2014). Maize centromeres expand and adopt a uniform size in the genetic background of oat. Genome Res. **24:** 107–116.

Wang, L.S., Zeng, Z.X., Zhang, W.L., and Jiang, J.M. (2014). Three potato centromeres are associated with distinct haplotypes with or without megabase-sized satellite repeat arrays. Genetics **196:** 397–401.

Willard, H.F. (1991). Evolution of alpha satellite. Curr. Opin. Genet. Dev. **1:** 509–514.

Willard, H.F., and Waye, J.S. (1987). Hierarchical order in chromosome-specific human alpha satellite DNA. Trends Genet. **3:** 192–198.

Yan, H.H., Talbert, P.B., Lee, H.R., Jett, J., Henikoff, S., Chen, F., and Jiang, J.M. (2008). Intergenic locations of rice centromeric chromatin. PLoS Biol. **6:** e286.

Yan, H.H., et al. (2006). Genomic and genetic characterization of rice Cen3 reveals extensive transcription and evolutionary implications of a complex centromere. Plant Cell **18:** 2123–2133.

Yi, C.D., Zhang, W.L., Dai, X.B., Li, X., Gong, Z.Y., Zhou, Y., Liang, G.H., and Gu, M.H. (2013). Identification and diversity of functional centromere satellites in the wild rice species *Oryza brachyantha.* Chromosome Res. **21:** 725–737.

Zhang, T., Talbert, P.B., Zhang, W.L., Wu, Y.F., Yang, Z.J., Henikoff, J.G., Henikoff, S., and Jiang, J.M. (2013). The *CentO* satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres. Proc. Natl. Acad. Sci. USA **110:** E4875–E4883.