Mol Genet Genomics (2014) 289:1307–1319
DOI 10.1007/s00438-014-0891-8

ORIGINAL PAPER

Chromosomal organizations of major repeat families on potato (*Solanum tuberosum*) and further exploring in its sequenced genome

Xiaomin Tang · Erwin Datema · Myriam Olortegui Guzman · Jan M. de Boer · Herman J. van Eck · Christian W. B. Bachem · Richard G. F. Visser · Hans de Jong

Received: 8 March 2014 / Accepted: 18 July 2014 / Published online: 9 August 2014

Abstract One of the most powerful technologies in unraveling the organization of a eukaryotic plant genome is high-resolution Fluorescent in situ hybridization of repeats and single copy DNA sequences on pachytene chromosomes. This technology allows the integration of physical mapping information with chromosomal positions, including centromeres, telomeres, nucleolar-organizing region, and euchromatin and heterochromatin. In this report, we established chromosomal positions of different repeat fractions of the potato genomic DNA (Cot100, Cot500 and Cot1000) on the chromosomes. We also analysed various repeat elements that are unique to potato including the moderately repetitive P5 and REP2 elements, where the REP2 is part of a larger *Gypsy*-type LTR retrotransposon and cover most chromosome regions, with some brighter fluorescing spots in the heterochromatin. The most abundant tandem repeat is the potato genomic repeat 1 that covers subtelomeric regions of most chromosome arms. Extensive multiple alignments of these repetitive sequences in the assembled RH89-039-16 potato BACs and the draft assembly of the DM1-3 516 R44 genome shed light on the conservation of these repeats within the potato genome. The consensus sequences thus obtained revealed the native complete transposable elements from which they were derived.

Keywords Potato (*Solanum tuberosum*) · Repetitive sequences · Fluorescent in situ hybridization · Pachytene chromosome

Abbreviations

BAC	Bacterial artificial chromosome
BESs	BAC end sequences

Communicated by S. Hohmann.

Electronic supplementary material The online version of this article (doi:10.1007/s00438-014-0891-8) contains supplementary material, which is available to authorized users.

X. Tang · J. M. de Boer · H. J. van Eck · C. W. B. Bachem · R. G. F. Visser
Wageningen UR Plant Breeding, Wageningen University and Research Centre, 6708 PB Wageningen, The Netherlands
e-mail: xiaomintang2013@hotmail.com

J. M. de Boer
e-mail: janmdeboer@gmail.com

H. J. van Eck
e-mail: herman.vaneck@wur.nl

C. W. B. Bachem
e-mail: christian.bachem@wur.nl

R. G. F. Visser
e-mail: richard.visser@wur.nl

X. Tang
Department of Biology, Colorado State University, Fort Collins, CO 80523, USA

E. Datema
Keygene N.V., 216, 6700 AE Wageningen, The Netherlands
e-mail: erwin.datema@keygene.com

M. O. Guzman · H. de Jong (✉)
Laboratory of Genetics, Wageningen University, 6700 AH Wageningen, The Netherlands
e-mail: hans.dejong@wur.nl

H. J. van Eck · R. G. F. Visser
Center for BioSystems Genomics (CBSG), Wageningen, The Netherlands

CCD	Charge-coupled device
Cot	(=Co × *t*) Co is the initial concentration of single-stranded DNA in mole per liter, and *t* is the reannealing time in seconds. Cot is a measure of DNA complexity
Cy 3, 3.5, 5	Cyanide dyes
DAPI	4′,6-Diamidino-2-phenylindole
DEAC	Diethylaminocoumarin
DM	DM1-3 516 R44 genotype
FISH	Fluorescent in situ hybridization
FITC	Fluorescein isothiocyanate
ISH	In situ hybridization
LINEs	Long interspersed nuclear elements
LTRs	Long terminal repeats
NOR	Nucleolar-organizing region
ORFs	Open reading frames
PFGE	Pulsed field gel electrophoresis
PGR1	Potato genomic repeat 1: a subtelomeric tandem repeat family in potato
PGSC	Potato genome sequencing consortium
RH	RH89-039-16 genotype
SINEs	Short interspersed nuclear elements
TEs	Transposable elements
TGRI	Tomato genomic repeat I: a subtelomeric tandem repeat family in tomato
TPE	Telomere position effect
TR	Telomeric repeat
TRF	Tandem repeats finder

Introduction

A substantial portion of the eukaryotic genome is composed of tandem (satellite) and dispersed repetitive sequences that are known to hinder DNA sequence assembly and thus annotation efforts (Treangen and Salzberg 2012). A great part of these repeats have a genome-wide distribution under the control of transposable elements (Lisch 2013; Wicker et al. 2007). The most common transposable elements in eukaryotic genome are RNA transposons, also called retroelements. They can be broadly classified into a class with long terminal repeats (LTRs) and those without. Long terminal repeats, which play a central role in the process by which the RNA copy of an LTR element is reverse transcribed into double-stranded DNA. *Ty*1/*Copia* family and *Ty*3/*Gypsy* family are two most common LTR retroelements making up substantial parts of many eukaryotic genomes and are particularly abundant in the larger plant genomes (Lisch 2013; Wicker et al. 2007).

One of the useful tools to characterize chromosomal organization of repetitive sequences is Fluorescent in situ hybridization (FISH). This technology is particularly powerful to establish microscopic positions of DNA sequences on chromosomal and nuclear targets. It has been established that tandem arrays of repeated sequences mainly locate close to the telomeres, around centromeres and a few at interstitial sites of chromosomes, while moderately repetitive DNA sequences are distributed on the chromosomes with only a few of them showing preferences in some regions, such as centromeric and pericentromeric heterochromatin regions (Anzai et al. 2001; Brown 2007; Hemleben et al. 2007). In addition, various studies on plant chromosomes have shown that *Ty*1/*Copia* elements appear more in euchromatin, while *Ty3/Gypsy* elements dominate in heterochromatic regions (Belyayev et al. 2001; Brandes et al. 1997; Heslop-Harrison et al. 1997; Lamb et al. 2007; Mroczek and Dawe 2003; Pearce et al. 1996, 1997; Pich and Schubert 1998; Presting et al. 1998). There is increasing evidence that tandem repeats and retroelements are structural and/or functional components of eukaryotic chromosomes, as they are closely associated with constitutive heterochromatin features (Bender 2004; Hall et al. 2002; Lippman et al. 2004; Volpe et al. 2002; Zhong et al. 2002). The centromeric satellites and retroelements are essential keys for centromere recognition by kinetochore proteins (Nagaki et al. 2003, 2011; Zhong et al. 2002). Repeats in the pericentromeric regions play important roles in initiating the recruitment of histone modification enzymes and promote the formation and maintenance of heterochromatin by RNAi machinery (Bender 2004; Hall et al. 2002; Lippman et al. 2004; Volpe et al. 2002; Zhong et al. 2002).

Many tandem repeats can diverge rapidly even between closely related species. They develop and amplify a certain repeat motif that spreads around as a "footprint" in the nuclear genome at high copy number, but are less common or even absent in related species. Such sequences are referred to as species-specific repeats (Lapitan 1992). The fact that pericentromeric heterochromatin is less in genes and is evolving rather rapidly with respect to repeat composition may explain why chromosome pairing and meiotic recombination are often repressed (up to 1,000-fold) in heterochromatin vs. euchromatin, especially in interspecific hybrids (Charlesworth et al. 1994). However, some satellite elements are more widespread across taxa. For example, the *Solanum brevidens* DNA sequences pSB1 and pSB7 are common in the whole Etuberosa group of genus *Solanum* (Malkamaki et al. 1996); the satellite repeat 2D8 that is derived from the intergenic spacer of the 18S-25S ribosomal RNA genes in potato is found in almost all *Solanum* species (Stupar et al. 2002).

Potato is one of the most important vegetable crops worldwide and together with tomato is the first species of the Solanaceae family that have been sequenced (PGSC 2011; TGC 2012). Both species are reference genomes for all other Solanaceae crops and wild relatives and thus provide invaluable resources for various fundamental questions

on gene expression, genome organization and quality trait improvements (Mueller et al. 2005; http://solgenomics.net; Visser et al. 2009). With a genome size of 840 Mb, potato was estimated to contain about 68.6 % of repetitive DNA sequences (Tang et al. 2009). Cot fraction-painting is a simple and efficient way to visualize the extent and location of all the major repeats of potato on pachytene chromosomes with respect to the pericentromeric, centromeric regions, the (sub)telomeric blocks and the nucleolar organizer regions (NORs) (Peterson et al. 2002). In addition, unlabelled potato Cot DNA can be used as blocking agent in BAC-FISH painting on the chromosomes, suppressing signals from repetitive sequences in the BACs (Szinay et al. 2010; Tang et al. 2009).

Repetitive DNA sequences have been isolated from various species of the *Solanaceae* family. Some of their characteristics including satellite DNA, transposable elements, retrotransposons and ribosomal RNA genes are listed in Table 1. These sequences have been shown useful tools for phylogenetic studies, establishing parental origin of chromosomes in somatic hybrids, constructing karyotypes of plant species and for studying structural alterations, such as sequence amplification and interchanges in chromosomes. However, in potato very few of these repetitive DNA families with the exception of ribosomal DNA have been studied in detail. The 45S rDNA complex containing the 18S, 5.8S and 25S units has been shown at the end of the short arm of chromosome 2 in the NOR. The 5S rRNA gene cluster was mapped to a single locus in the short arm pericentromeric region of chromosome 1 close to the centromere (Dong et al. 2000). As part of a study on the organization of the potato genome and evaluation of specific DNA probes for tuber-bearing *Solanum* species diagnosis, a repetitive DNA sequence, termed P5, was cloned (Visser et al. 1988). In this study we have further characterized the chromosomal localization, repeat structure and nucleotide sequences of three repetitive DNAs in *Solanum tubersosum*, one tandem repeated DNA family PGR1 (potato genomic repeat 1) and two interspersed repeats P5 and REP2.

Materials and methods

Materials

For preparing cell spread preparations we used anthers of the diploid potato (*Solanum tuberosum* L.) genotypes G254 (Hermsen et al. 1978) and RH89-039-16 (Rouppe van der Voort et al. 1997). Ten potato BACs were used as cytogenetic markers to identify bivalents 3–12, respectively, in pachytene cell complement (Tang et al. 2009). Twenty-two plasmid clones derived from the *Solanum tuberosum* HH578 were analysed by FISH and sequenced. For detailed isolation procedures, please see Visser et al. (1988). In short, several hundred clones were obtained from the total EcoRI digests of nuclear DNA and used for cloning into plasmid pUC9. By clone hybridization with labelled nuclear HH578 DNA, only these 22 clones were selected based upon (1) the substantial strong signal these positive clones gave and (2) the ladder-pattern they gave in genomic Southern's, which is characteristic of hybridization with a repetitive probe. Further FISH and sequence results revealed the presence of three different classes of repeats. Here the representative repetitive DNA clones P5, REP2, as well as the subtelomeric repeats PGR1 (RH4-1 and REP3) were used.

Isolation of genomic DNA and production of Cot fractions

Cot fractions of potato genomic DNA were prepared according to Zwick et al. (1997) with some modifications. Total genomic DNA was extracted from leaves according to the CTAB method (Rogers and Bendich 1988) and sheared to fragment size of about 500 bp with a Vibra cell sonicator for 40 s. We denatured 1 μg/μL sonicated DNA in 0.3 M NaCl at 95 °C for 10 min and then transferred it to a water bath set at 62.4 °C (Chang et al. 2008) for single-strand DNA reassociation. Our reassociation time for Cot100 is 9 h 25 min, for Cot500 is 47 h 5 min and for Cot1000 is 94 h 10 min. The remaining single-strand DNA was digested with 1 U/μg S1-endonuclease (Fermentas) for 90 min at 37 °C and the double-strand DNA fraction was finally purified by ethanol precipitation.

DNA sequencing and analysis

The plasmid DNAs of RH4-1, REP3, REP2 and P5 were isolated using a standard alkaline extraction and sequenced using the BigDye Terminator Cycle Sequencing Kit version 2 (Applied Biosystems, Foster City, CA, USA) with M13 forward and M13 reverse primers and an MJ Research (Watertown, PA) PTC-100 thermo-cycler. Sequencing was performed using an ABI 3730 automatic DNA Analyzer. ABI sequencer trace data were evaluated using the program Phred. Poor quality and vector sequences were filtered with Cross-match. High-quality sequences (over Phred 30) were assembled and edited with Phrap and Consed (Ewing and Green 1998; Ewing et al. 1998). The nucleotide sequences of the cloned P5, REP2, REP3 and RH4-1 are now available in the GenBank databases under the accession numbers KC904079, KC904080, KC904081, KC904082, respectively.

Solanum tuberosum RH89-039-16 BAC sequences and the *S. phureja* DM1-3 516 R44 draft genome sequences (version 3) were kindly provided by the Potato Genome Sequencing Consortium (http://www.potatogenome.net). Potato RH89-039-16 BAC end sequences were downloaded

Table 1 Distribution and characteristics of various repetitive DNA elements in the *Solanaceae*

Repeat name	Size (bp)	Species	Characterization, chromosome position	Abundance (% genome)	References
pSB1	400	*Solanum brevidens*	Subtelomeric, centromeric, interstitial	0.04	Pehu et al. 1990; Rokka et al. 1998
pSB7	210	*Solanum brevidens*	Subtelomeric, centromeric, interstitial	0.01	Pehu et al. 1990; Rokka et al. 1998
pST3	1.5 k	*Solanum tuberosum*	Subtelomeric	0.028	Pehu et al. 1990; Rokka et al. 1998
pST10	200	*Solanum tuberosum*	ND	0.003	Pehu et al. 1990; Rokka et al. 1998
Sb4/2	1.728 k	*Solanum brevidens*	Subtelomeric	1	Preiszner et al. 1994
SPG repeat family	120-140	*Solanum spegazzinii*	Subtelomeric, twelve loci	0.26	Gebhardt et al. 1995
pSA287	183	*Solanum acaule*	ND	0.2–0.4	Schweizer et al. 1993, 1988
pRIT320	360	*Solanum tuberosum*	ND	0.5–2.6	Schweizer et al. 1993
pSDT382/pSDB6	370	*Solanum demissum*	ND	0.01–0.12	Schweizer et al. 1993; Zanke and Hemleben 1997
pSCH15	168	*Solanum circaeifolium*	ND	ND	Stadler et al. 1995
pSBH6 (95 % sequence homology to TGRI)	163	*Solanum bulbocastanum*	ND	ND	Stadler et al. 1995
2D8	5.9 k	*Solanum bulbocastanum*	Pericentromeric	0.875	Stupar et al. 2002
pSbTC1	2.8 k	*Solanum bulbocastanum*	Centromeric	0.76–1.44	Tek and Jiang 2004
Sobo	4.7 k	*Solanum bulbocastanum*	A single location in the pericentromeric region of chromosome 7	ND	Tek et al. 2005
P5	2.117 k	*Solanum tuberosum*	Pericentromeric, interspersed repeats	0.3	Visser et al. 1988; this study
REP2	2.085 k	*Solanum tuberosum*	Pericentromeric, interspersed repeats, Ty3/gypsy LTR retroelement	0.7	This study
PGR1/REP3	160	*Solanum tuberosum*	Subtelomeric, fourteen loci	0.2	This study
TGRI/Pleg15	162	*Solanum lycopersicum*	Subtelomeric, twenty loci	1.27	Ganal et al. 1988; Lapitan et al. 1989; Schweizer et al. 1988; Zhong et al. 1998
TGRII	780	*Solanum lycopersicum*	Pericentromeric, retrotransposon	0.33	Chang et al. 2008
TGRIII	509	*Solanum lycopersicum*	Pericentromeric, retrotransposon	0.11	Chang et al. 2008
TGRIV	7 k	*Solanum lycopersicum*	Centromeric, retrotransposon	ND	Chang et al. 2008
HRS60	182	*Nicotiana tabacum*	Subtelomeric, eleven loci	2	Koukalová et al. 1989
5S rDNA	400	*Solanum tuberosum*	Transcribed, single locus	ND	Dong et al. 2000
18S-5.8S-25S (45S) rDNA	9.1 k	*Solanum tuberosum*	NOR region, short arm chromosome 2	1.9–5.2	Dong et al. 2000; Schweizer et al. 1993
pAtT4	7	*Solanum tuberosum*	Telomere	ND	Visser et al. 2009
CL14	182	*Solanum tuberosum*	Subtelmeric, thirteen loci	0.4	Torres et al. 2011
CL34	339	*Solanum tuberosum*	Subtelmeric, sixteen loci	0.2	Torres et al. 2011
St18	1.18 k	*Solanum tuberosum*	Centromeric satellite repeat unique to chromosome 9	ND	Gong et al. 2012
St24	979	*Solanum tuberosum*	Centromeric satellite repeat unique to chromosome 1	ND	Gong et al. 2012
St49	2.754 k	*Solanum tuberosum*	Centromeric satellite repeat unique to chromosome 5	ND	Gong et al. 2012
St57	1.924 k	*Solanum tuberosum*	Centromeric satellite repeat unique to chromosome 7	ND	Gong et al. 2012

Table 1 continued

Repeat name	Size (bp)	Species	Characterization, chromosome position	Abundance (% genome)	References
St3-58	2.957 k	*Solanum tuberosum*	Centromeric satellite repeat unique to chromosome 2	ND	Gong et al. 2012
St3-238	3.814 k	*Solanum tuberosum*	Centromeric satellite repeat unique to chromosome 8	ND	Gong et al. 2012
SolS SINE family	106–244	*Solanum tuberosum*	Dispersed with preferred integration into short A-rich motifs	ND	Wenke et al. 2011

ND not determined

Table 2 An overview of the sequence database

Species	Type	Total length	A/T %
Solanum tuberosum RH89-039-16	BACs	178,073,876	64.6
Solanum tuberosum RH89-039-16	BAC ends	91,376,407	64.3
Solanum phureja DM1-3 516 R44	Draft genome	727,424,275	65.2
Solanum tuberosum	ESTs	138,211,100	57.8
Solanum lycopersicum	Draft genome	781,630,557	66.0

from NCBI GenBank as previously described in Zhu et al. (2008). *S. tuberosum* EST sequences were obtained from the SolEST database (D'Agostino et al. 2009). The *S. lycopersicum* draft genome sequence (version 2.10) was kindly provided by The International Tomato Genome Sequencing Consortium (http://solgenomics.net/genomes/Solanum_lycopersicum/). An overview of the sequence data is provided in Table 2.

The full REP2 sequence was identified from the RH BAC sequences using BLASTn 2.2.15 (Altschul et al. 1997) of the original HH578 REP2 sequence and subsequent post processing with a custom python script. The repeat sequences of P5, REP2 and REP3 were aligned to the potato and tomato sequences with BLASTn with low complexity filtering disabled and an *E*-value cutoff of 1e−10 for the potato sequence databases and 1e−03 for the tomato genome sequence. Alignments were processed using a custom python script to count repeat copies that covered at least 33 % of the query sequence. The alignments of the P5 and REP2 sequences against the BAC end and EST sequences were processed differently, such that all matches of at least 200 residues were counted, as the P5 and REP2 sequences were larger than the average length of the BAC end and EST sequences. Repeat copies identified with the method outlined above were extracted from the BAC sequences and aligned with ClustalW 2.0.12 (Larkin et al. 2007). Consensus sequences and coverage histograms of P5 and REP2 were determined using a custom python script. REP3 was further characterized by analysing the RH BAC sequences having this repeat with the Tandem Repeats Finder (TRF) program (Benson 1999). A consensus sequence for REP3 was determined from the sequence of BAC RH042N03 by combining 208 repeat copies from different clusters into a continuous larger sequence and re-analyzing this sequence with TRF. This alignment also provided data on the conservation of the REP3 (PGR1) consensus sequence. Protein domain annotation for the retrotransposons flanked by REP2 elements was performed with the InterProScan web interface from http://www.ebi.ac.uk/Tools/InterProScan/index.html using default parameters. The consensus sequences of the P5, REP2 and REP3 (PGR1) are listed in the supplementary files.

FISH and image capturing

Pachytene chromosome preparations were made as described by Zhong et al. (1996a) with few minor modifications. Each repeat DNA, Cot fraction DNA and BAC DNA (1 μg) was directly labelled with one of the five fluorescent nucleotides: Cy3-dUTP, Cy3.5-dCTP, Cy5-dUTP (GE Healthcare, Sweden), fluorescein-12-dUTP (fluorescein isothiocyanate, FITC), and diethylaminocoumarin (DEAC)-5-dUTP (PerkinElmer Inc.). The labelling methods followed the nick translation protocols of Amersham Bioscience (GE Healthcare, Sweden). FISH was performed according to the protocol of Zhong et al. (1996b). Slides were examined under a Zeiss Axioplan 2 Imaging Photomicroscope equipped with epifluorescence illumination, with filter sets for 4′,6-diamidino-2-phenylindole (DAPI), DEAC (blue), FITC (green), Cy3 (orange), Cy3.5 (red) and Cy5 (far-red) fluorescence. Selected images were captured by a Photometrics Sensys 1,305 × 1,024 pixel CCD camera. Image processing and FISH signal thresholding were performed with the Genus Image Analysis Workstation software (Applied Imaging Corporation). DAPI images were sharpened separately with a 7 × 7 Hi-Gauss high-pass spatial filter to accentuate minor details and heterochromatin differentiation of the chromosomes. The FISH signals of the different fluorescence images were combined with the DAPI image in a multichannel stack. DAPI images were displayed in grey, and the other fluorescence images except Cy 5 were

pseudocoloured corresponding to the original colours of the fluorescing dyes. For Cy5 we used the purple pseudocolour. Brightness and contrast of the composite images were further improved with Photoshop CS4 (Adobe).

Results

Cot analysis and mapping on potato pachytene cell complement

The Cot100, Cot500 and Cot1000 fractions were used as pools of highly repetitive, highly + moderately repetitive and highly + moderately + single/low-copy sequence components of the potato genome. We established the chromosomal locations of these repeat fractions by FISH on pachytene cell complements (Fig. 1a–e). Hybridization with Cot100 showed fluorescing foci on most chromosome ends and in the pericentromeric heterochromatin regions, except for some interstitial and distal ends (Fig. 1b, e). The Cot500 fraction, which contains the repeats of the Cot100 DNA pool as well, displayed similar FISH signals as Cot100 (Fig. 1c, e). FISH with Cot1000 DNA resulted in bright signals over heterochromatin regions of all chromosomes and to a lesser extent on some minor regions in the euchromatin areas (Fig. 1d, e). Besides, there were signals in the cytoplasm (Fig. 1b–e), which probably reflect hybridization of DNA sequences in the chloroplast and/(or) mitochondrial DNA. In view of the highly similar results obtained with Cot100 and Cot500, therefore, Cot100 of potato already can be sufficiently and efficiently used as blocking DNA in BAC-FISH painting.

Organization and dispersed distribution of two moderately repetitive DNA sequences in the potato genome

FISH with the P5 and REP2 probes was carried out to reveal the chromosomal distributions of the two repeats. They were detected randomly distributed throughout the genome with some more intense clusters predominantly associated with the pericentromeric heterochromatin regions. The FISH patterns of P5 and REP2 showed more or less co-localized characterizations (Fig. 1f). However, sequence analysis of these two clones revealed no sequence similarity between them (P5: 2,336 bp, GenBank accession number: KC904079; REP2: 649 bp, GenBank accession number: KC904080). P5 and REP2 have an A/T composition of 65 and 67 %, respectively. BLAST alignment of the two repetitive DNA elements to several potato sequence datasets showed that both sequences occur quite frequently in the potato genome (PGSC 2011), thus confirming the repetitive nature of these two sequences (Table 2). The size and abundance of the P5 and REP2 elements indicated that they are more closely related to the middle repetitive DNA sequences. In addition, neither sequence shows any good sequence homology to currently known potato or tomato genomic repeat sequences in the Solanaceae Repeat Database v3.2 (http://plantrepeats.plantbiology.msu.edu/), nor to other repeat sequences from RepBase 15.06. Further analysis of the REP2 sequence revealed that it is a part of the LTR sequence of a Gypsy-type retroelement. A representative of the full REP2 LTR sequence (Supplementary files, REP2 consensus sequence: 2,085 bp) was extracted from the RH89-03916 BAC sequences and used for subsequent analyses. Domain analysis of several intact copies of the retrotransposon flanked by REP2 elements revealed the presence of a retrotransposon GAG protein (IPR005162), a CCHC-type zinc finger (IPR001878), a retroviral aspartyl protease (IPR013242), a reverse transcriptase (IPR000477), a ribonuclease H (IPR012337) and an integrase (IPR001584) domain.

To estimate the copy number of these repetitive elements in the potato genome, we compared the P5 and REP2 sequences to the currently sequenced RH89-039-16 BACs (hereafter referred to as RH) and the DM1-3 516 R44 draft genome sequences (hereafter referred to as DM) (PGSC 2011). In total, 104 copies of the P5 element were found in 178 Mb of non-redundant BAC sequences from RH and 219 matches were found in 91 Mb of RH BAC end sequences. For the REP2 sequence, 241 copies were identified in the BAC sequences, 112 of which were paired LTRs of a Gypsy-type retrotransposon, and 491 matches were found to the BAC end sequences. However, both elements had a much lower frequency in the DM genome (Table 3). Multiple alignments of P5 and REP2 copies from the RH BAC sequences were exploited to reveal the sequence variation of these repeats. Figure S1 shows the consensus sequence conservation of the 104 P5 copies in the RH BACs (Supplementary files, P5 consensus sequence: 2,117 bp). In this figure there appear to be several distinct regions conserved at various levels of identity. The region between positions 569 and 1,000 is almost fully conserved in all repeat copies, whereas the surrounding regions are relatively heterogeneous. The P5 sequence from HH578 differs slightly from those P5 elements in the RH BACs: the first 1,376 residues aligned with 91 % identity to the RH P5 consensus, and the region between 1,651 and 1,955 aligned with 88 % identity. The 3′ end contains four GAGTTTTATGTGAAGCTATATGTGAAGGAATGATGTGAAT satellite repeat units. The P5 sequence contains several short open reading frames (ORFs), but has no significant similarity to either the UniProt plant protein database or to potato ESTs. Thus, this sequence does not appear to contain protein-coding elements and is not likely to represent part of a retrotransposon. In contrast, REP2 was well conserved (80 %) in RH over its full length (Figure S2), with the region between 581 and 1,022 of the

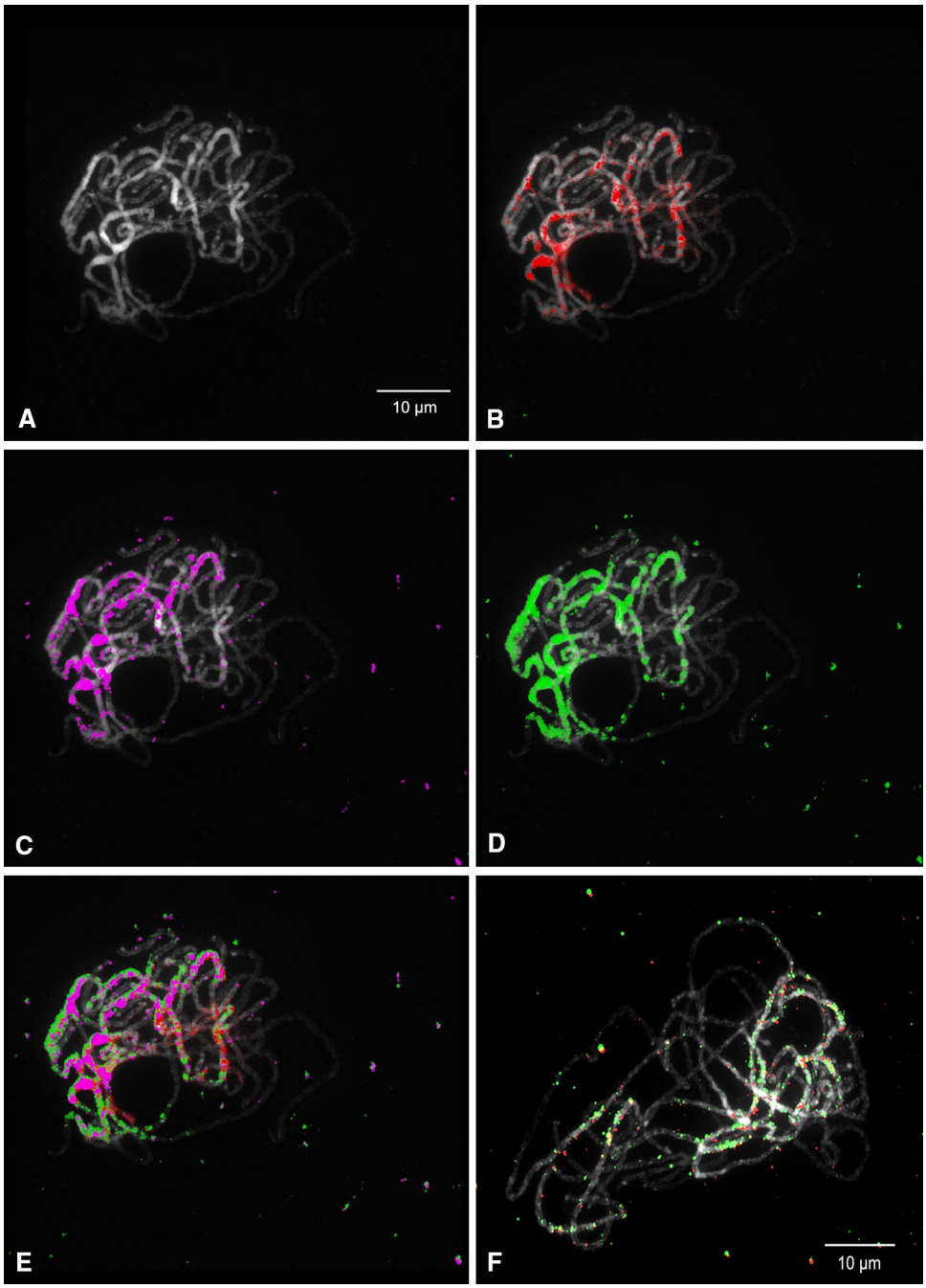

Fig. 1 **a**–**e** FISH analysis of different Cot DNA fractions on pachytene cell complement of potato. **a** DAPI, **b** Cot100, **c** Cot500, **d** Cot1000, **e** merged images of **a**–**d**. **f** FISH of repetitive sequences P5 (*green*) and REP2 (*red*) on pachytene cell complement of potato

consensus sequence corresponding to the region between 42 and 473 in the REP2 clone from HH578.

Comparison to the tomato draft genome sequences revealed that both repeats were specific to potato. While the P5 sequence had a small number of partial matches to the tomato genome, the matches are all located outside the strongly conserved region of the potato P5 sequence

Table 3 The potato repetitive DNA elements alignment to the sequence database

	BACs	BAC ends	DM	ESTs	Tomato
P5	104	219	51	0	0
REP2	241	491	346	0	0
REP3	1,307	1,175	5,906	0	153

and did not include the 3′ tandem repeats. There were no matches of the REP2 sequence to the tomato genome. This correlated with our FISH results, as we were not able to get any signals on tomato with these two probes (data not shown).

Physical mapping of the subtelomeric repeat family PGR1 in potato pachytene chromosomes

FISH analyses showed that the two RH4-1 and REP3 clones overlapped fully and are located on 14 of the 24 *Solanum tuberosum* chromosome ends. This FISH pattern is typical for subtelomeric repeats, suggesting that these two clones represent subtelomeric repeats. Alignment of the REP3 (162 bp, GenBank accession number: KC904081) and RH4-1 (357 bp, GenBank accession number: KC904082) sequences revealed that RH4-1 is a dimer of REP3. Thus, we renamed the two clones into the subtelomeric repeat family PGR1 and used REP3 as its representative probe. After reprobing the cell complement with the aid of ten cytogenetic BAC markers (Tang et al. 2009), we were able to study distribution patterns of these repetitive sequences for each individual chromosome arm (Fig. 2a, b). Brightly DAPI-stained and knob-like heterochromatin structures were observed at the distal ends of several chromosomes. We found that these distal knobs were associated with the

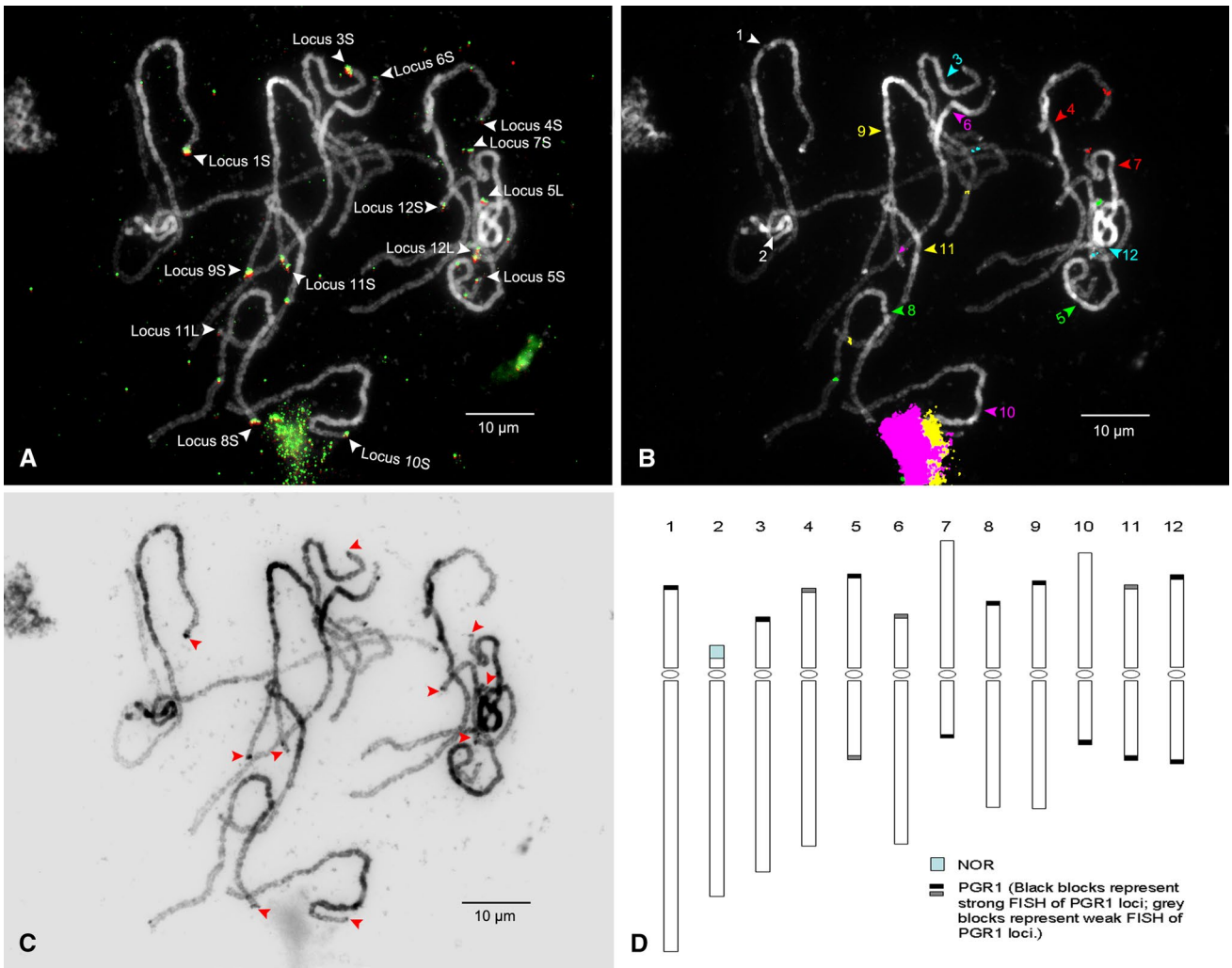

Fig. 2 Physical mapping the subtelomeric repeats REP3 and RH4-1 clones in pachytene chromosomes of *Solanum tuberosum* cultivar G254. **a** Fourteen loci (denoted by *white arrowheads*) (*red* for REP3, *green* for RH4-1) were detected and mostly located on the distal regions of the short arms of chromosomes. **b** Five multi-colour FISH with ten chromosome-specific BACs allowed the identification of all twelve bivalents in one FISH experiment except chromosomes 1 and 2, which can be identified unambiguously by their morphology. The centromeres are indicated by *arrowheads*, which are located within a lightly stained chromatin domain. **c** DAPI-stained chromosomes were converted to a *black-and-white* image to enhance the visualization of the conspicuous heterochromatin knobs (pointed by *red arrowheads*). **d** Ideogram of chromosomal locations of the PGR1 based on FISH results. The relative length of each chromosome and the centromere index of potato were drawn based on Table 2 data of Tang et al. (2009)

Table 4 PGR1 subtelomeric repeats in sequenced BACs from potato clone RH89-039-16

BAC name	Accession	Tiling path	Chromosome	Genetic map location	Chromosome region	PGR1 Copies
RH042N03	AC233382	4,846	5	bin 004	End of long arm	696
RH041F06	AC235622	5,611	9	bin 001	End of short arm	619
RH009K01	AC235609	1,379	9	bin 045	Long arm euchromatin	274

subtelomeric repeats PGR1 (Fig. 2a, c). The PGR1 loci were assigned to almost all short arms of chromosomes as well as 5L, 11L and 12L, with the exception of the locus of short arm of chromosome 2 (2S) which harbors the NOR. The sizes and fluorescence intensity of the PGR1-related knobs as observed with DAPI staining correlate with the copy number of the PGR1 repeat, which can be judged from the size and intensity of the FISH signals. FISH of PGR1 in 4S, 5S, 6S and 11L displays relatively faint signals suggesting that there are fewer copies of the PGR1 at these chromosomal locations. Moreover, these small loci are not associated with a knob-like structure (Fig. 2c) but showed only lightly stained euchromatin as other ends that do not contain PGR1 sequences. The results reflect that the distribution of the PGR1 repetitive sequence is specific and that the copy number of repeats within a cluster is not the same at different chromosomal locations. Our FISH results also suggest that the largest PGR1 loci are organized into large tandem arrays spanning as many as several hundred kilobases. Besides, PGR1 is a species-specific repetitive DNA sequence, which could not be detected in the closely related tomato genome (data not shown).

When PGR1 (represented by the REP3 sequence) was BLASTed to the RH BAC sequences, we found more than 1,300 copies (Table 3), almost all of which were present in three BAC tiling paths (Table 4). The BACs in these three tiling paths were anchored to the north arm telomeres of chromosome 5 and chromosome 9 and in the south arm euchromatin of chromosome 9 like an “invisible knob”, containing 523, 626, and 103 copies, respectively. The first two of these BAC locations corresponded to our FISH results. Re-analysis of these BAC sequences with tandem repeats finder software (Benson 1999) identified additional copies of PGR1, bringing the total PGR1 copy number in these clones to 1,589 (Table 4). Similarly, 5,906 copies of PGR1 were detected in the draft assembly of the DM genome, 80 % of which localize in 22 distinct super scaffolds containing more than 50 copies.

Within the three BAC sequences, the PGR1 repeats are organized in clusters that each contain between 3 and 169 tandemly arranged copies of the repeat sequence. With a predominant unit size of 182 bp, the full PGR1 tandem repeat is 20 bp larger than the cloned REP3 fragment. A consensus sequence for PGR1 was calculated from the RH042N03 sequence (Supplementary files, PGR1 consensus sequence: 182 bp). The average sequence conservation is 90 % (Figure S3). There were no hits of the consensus sequence to RepBase 15.06 and the Solanaceae Repeat Database v3.2 (http://plantrepeats.plantbiology.msu.edu/). However, when the sequence was compared to the NCBI non-redundant nucleotide database (http://www.ncbi.nlm.nih.gov/) we found many hits, mostly to tomato BACs, a few to *S. brevidens* repetitive sequences and several *S. bulbocastanum* satellite repeats. While the PGR1 subtelomeric repeat sequence is unique to potato, 153 partial matches to the tomato genome were identified (Table 3). On average, these matches covered 46 % of the PGR1 sequence with 84 % identity. Seventy-four to seventy-eight per cent sequence similarities were found between PGR1 and a repetitive sequence of *S. brevidens* Sb4/2 depending on its subrepeat. This homology is restricted only to one of the *S. brevidens* subrepeats Sb4AX (Preiszner et al. 1994), which shows 74 % sequence similarity with TGRI (Zhong et al. 1998) in tomato. Similarly to PGR1, the *S. brevidens*-specific repeats were detected at the ends of several chromosomes but the TGRI localized as a subtelomeric repeat on almost all tomato chromosomes.

Discussion

In contrast to the rapid progress of genetic and molecular studies of the potato genome, only limited success has been achieved toward a molecular cytogenetic characterization of the potato genome. Very few heterochromatic or euchromatic features in the potato genome have been defined at the DNA-sequence level (Stupar et al. 2002; Tang et al. 2009; Tek and Jiang 2004). In this report, we presented a simplified technique for isolating different Cot-DNA fractions based on typical DNA reassociation of sheared denatured genomic DNA and showed their FISH signals on pachytene cell complements. Furthermore, the characterization of major repetitive DNA families including the subtelomeric tandem repeat family (PGR1) and two moderately repetitive DNA elements (P5 and REP2) provided valuable information about this important crop. Their consensus sequences that were extracted are excellent approximations of the active transposon elements (TEs) from which they were derived. Consensus sequences are also preferred reference sequences for screening and

annotation of repetitive elements, especially the most divergent ones (Jurka et al. 2011).We used a Vibra cell sonicator to fragment the potato genomic DNA, in which the DNA subjected to brief periods of sonication was sheared into smaller fragments about 500 bp. Compared to the old lysing instruments such as autoclave (Zwick et al. 1997) it proved to be more efficient and/or faster. And with smaller fragments of Cot DNA blocking in our probes, we could get better FISH images with little or no interference background. The FISH experiments of Cot fractions in potato have shown that most of the repeats are confined to clearly distinguishable knobs at the telomeres, the pericentromere heterochromatin regions and the large NOR. The FISH results of P5 on pachytene chromosomes showed a different distribution as was concluded in a previous in situ hybridization (ISH) on mitotic complements (Visser et al. 1988) in which the 1.4-kb probe of P5L was found on telomere and (peri)centromere regions of all chromosomes. These conflicting results may be due to the poor sensitivity and resolution of ISH on metaphase chromosomes with radioactively labelled probes. The size and abundance of the P5 and REP2 indicated that they are two middle repetitive DNA sequences. Apart from basic organizational similarity, P5 and REP2 show complete sequence non-identity, suggesting that although FISH signals were co-localized in the same physical position, they are probably not intermingled, but most likely form closely adjacent uninterrupted clusters or are interspersed with other repetitive elements, such as transposable elements, satellite sequences or single sequences, resulting in complex interactions. This speculation was supported by sequence analysis. In RH BAC tiling paths we looked at, 66 tiling paths contain P5, 110 tiling paths contain REP2, and only 24 tiling paths contain both. In addition, we found only a few (1–3) copies of P5 or REP2 in BAC tiling path. Therefore, co-localization in sequence resolution was not seen so much as visualized by FISH. This difference may be explained by the limited resolution of FISH. 104 and 219 copies of the P5 element were found in the RH BACs and BAC end sequences, respectively. Extrapolating this to a 2 × 840 Mb genome (i.e., 840 Mb for each genetic linkage phase) and assuming the BACs represent a random selection of the genome, there would be approximately 1,000–4,000 copies of P5 in the potato genome. The upper limits are likely an overestimation as we included partial matches and did not consider the unequal distribution of the enzyme-digested BAC libraries in the genome. We estimated that on average, the P5 would make up 0.3 % of the potato genome. For the REP2 sequence, 241 and 491 copies were identified in the BACs and BAC end sequences, respectively, corresponding to roughly 2,500–9,000 copies in the genome. On average, we estimated the REP2 to make up 0.7 % of the potato genome. Approximately half of the REP2 copies we identified in the BAC sequences are paired and flanking a Gypsy-type retrotransposon, suggesting that at the lower limit there would be around 600 copies of this retroelement in the genome and an additional 1,200 solo LTRs. Since the BAC sequence data are highly fragmented, it is likely that more intact copies of the retrotransposon are present in the genome. Both the P5 and the REP2 elements have a much lower frequency in the draft assembly of the DM genome, which is likely to be the result of the non-random exclusion of repetitive sequences from the whole-genome assembly. Thus, the DM genome assembly likely under represents many repeats. Datema et al. (2008) observed a high percentage of LINE, SINE elements and class II DNA transposons in the large-scale potato BAC End Sequences (BESs) analysis. Comparisons between P5 and sequences of representative SolS SINE family in potato (GenBank accession numbers HE583424 and HE583588 in Wenke et al. 2011) showed no significant sequence similarity. Although we cannot give an accurate annotation for P5 at this moment, its sequence does not contain protein-coding elements, and thus it is more closely related to LINEs/DNA transposon.

Tandem repeats are generally associated with heterochromatic domains, such as the nucleolar organizers, knobs and pericentromeric regions. This structural conservation of repeat clusters suggests that sequence organization rather than sequence content is important in chromosome condensation (Tabata et al. 2000). In this paper, we showed that several heterochromatic knobs in potato pachytene chromosomes were associated with the subtelomeric repeat PGR1. Most of the PGR1 sequences were physically located very close to the telomeres of short arms of potato chromosomes. There is a correlation between knob size and the copy number of the PGR1 repeat. These results suggested that the tandem repeats might play a role in the condensation of the heterochromatic knobs. By use of PFGE (pulsed field gel electrophoresis) (Ganal et al. 1991) and fiber-FISH (Zhong et al. 1998), all chromosome ends in tomato were found to have their own unique molecular organization with respect to the telomeric repeat (TR) clusters and the subtelomeric repeat (TGRI) arrays. Taken together, it is possible that the structure of the potato telomeres are similar, in the sense that it exists in three forms: (1) only TR; (2) TR and PGR1 separated by a spacer which is not more than several hundred kilobases; (3) TR and PGR1 are directly associated together. In yeast and human, subtelomeric repeats can buffer the gene silencing from the telomere position effect (TPE) (Baur et al. 2001; Gottschling et al. 1990). The genes on subtelomeric repeat-containing chromosomal ends show high transcriptional activity due to the blockage of TPE by subtelomeric repeats. As the strength of the silencing effect is dependent on the distance between gene and telomere, it is speculated that the telomere-associated repeats, such as PGR1 domains and the

spacer sequence may also work as buffering blocks separating chromosome ends from unique sequences and potentially preserve the high expression of distal genes.

FISH signals derived from TGRI are located at almost all subtelomeric regions of tomato pachytene chromosomes as well as in the interstitial regions of many chromosomes (Ganal et al. 1991; Zhong et al. 1998). However, PGR1 is located only on about half of all the subtelomeric regions of potato pachytene chromosomes, and no obvious interstitial region was found. Although we cannot rule out the potential existence of minute clusters of PGR1 that might have been missed by FISH analyses (for example, an "invisible knob" in the south arm euchromatin of chromosome 9) because of the faint intensity of the fluorescence signal, the apparent lack of the PGR1 subtelomeric repeat suggested that it might not be necessary, or that other sequences in the subtelomeric region could also play a role in the maintenance of chromosome ends. This also supported by the fact that potato chromosome ends have heterogeneity in both sequence and structure. We compared the sequences of PGR1 and other two subtelomeric repeats CL14 and CL34, which were also found in *Solanum tuberosum* (Torres et al. 2011). The PGR1 consensus sequence has a 99 % sequence identity to the CL14 consensus sequence (Figure S4). Thirteen loci were detected from FISH mapping of CL14 on somatic metaphase chromosomes of potato (Torres et al. 2011). Although they did not assign the loci to the definite chromosome ends, they also found the CL14 are more frequently dominated at the ends of the short arms which are correlated with our results. Interestingly, the CL14 repeat showed a similar hybridization pattern among all *Solanum* species, while REP3 is a species-specific repetitive DNA clone, which could not be detected on FISH of tomato. This difference in specificity may come from the use of PCR-generated probe for the CL14 hybridizations of Torres et al. (2011), which may have broadened the hybridization range. There is no similarity between the PGR1 and CL34 repeats, indicating that these elements are two unrelated repeat families.

A high-quality sequence assembly of the potato genome has been published (PGSC 2011), but leaves questions about the exact organization of highly repetitive regions unanswered. More detailed characterization of euchromatin and heterochromatin at both the cytological and DNA sequence level will be a critical component for full understanding of the genome structure. Pachytene is the best meiotic stage to distinguish euchromatin and heterochromatin in chromosomes that can be identified individually. Our results have shown that pachytene chromosome-based FISH mapping is the most effective approach to integrate DNA sequence information with different types of chromatin structures and essential to fully characterize the potato genome.

Acknowledgments We thank Irma Straatman very much for technical assistance of propagation of the clones. This work was supported by grants from PGSC-NL funded by the Netherlands Technology Foundation (FES; Grant no. WGC.7795) and the Fund for Economic Structural Support (Netherlands Ministries of Economic Affairs and Agriculture).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Anzai T, Takahashi H, Fujiwara H (2001) Elimination of active *Tad* elements during the sexual phase of the *Neurospora crassa* life cycle. Fungal Genet Biol 33:49–57

Baur JA, Zou Y, Shay JW, Wright WE (2001) Telomere position effect in human cells. Science 292:2075–2077

Belyayev A, Raskina O, Nevo E (2001) Chromosomal distribution of reverse transcriptase containing retroelements in two Triticeae species. Chromosome Res 9:129–136

Bender J (2004) Chromatin-based silencing mechanisms. Curr Opin Plant Biol 7:521–526

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res 27:573–580

Brandes A, Heslop-Harrison JS, Kamm A, Kubis S, Doudrick RL, Schmidt T (1997) Comparative analy-sis of the chromosomal and genomic organization of Ty1-copia-like retrotransposons in pteridophytes, gymnosperms and angiosperms. Plant Mol Biol 33:11–21

Brown TA (2007) Genomes 3, Garland Science. Taylor & Francis Group, New York and London

Chang SB, Yang TJ, Datema E, van Vugt J, Vosman B, Kuipers A, Meznikova M, Szinay D, Lankhorst RK, Jacobsen E, de Jong H (2008) FISH mapping and molecular organization of the major repetitive sequences of tomato. Chromosome Res 16:919–933

Charlesworth B, Sniegowski P, Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220

D'Agostino N, Traini A, Frusciante L, Chiusano ML (2009) SolEST database: a "one-stop shop" approach to the study of Solanaceae transcriptomes. BMC Plant Biol 9:142

Datema E, Mueller LA, Buels R, Giovannoni JJ, Visser RGF, Stiekema WJ, van Ham RCHJ (2008) Comparative BAC end sequence analysis of tomato and potato reveals overrepresentation of specific gene families in potato. BMC Plant Biol 8:34

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

Ewing B, Green P (1998) Base-calling of automated sequencer traces using Phred. II. Error probabilities. Genome Res 8:186–194

Ewing B, Hillier L, Wendl MC, Green P (1998) Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res 8:175–185

Ganal MW, Lapitan NLV, Tanksley SD (1988) A molecular and cytogenetic survey of major repeated DNA sequences in tomato (*Lycopersicon esculentum*). Mol Gen Genet 213:262–268

Ganal MW, Lapitan NLV, Tanksley SD (1991) Macrostructure of the tomato telomeres. Plant Cell 3:87–94

Gebhardt C, Eberle B, Leonards-Schippers C, Walkemeier B, Salamini F (1995) Isolation, characterization and RFLP linkage mapping of a DNA repeat family of *Solanum spegazzinii* by which chromosome ends can be localized on the genetic map of potato. Genet Res Camb 65:1–10

Gong Z, Wu Y, Koblizkova A, Torres GA, Wang K, Iovene M, Neumann P, Zhang W, Novak P, Buell CR, Macas J, Jiang J (2012) Repeatless and repeat-based centromeresn in potato: implications for centromere evolution. Plant Cell 24:3559–3574

Gottschling DE, Aparicio OM, Billington BL, Zakian VA (1990) Position effect at *S cerevisiae* telomeres: reversible repression of Pol II transcription. Cell 63:751–762

Hall IM, Shankaranarayana GD, Noma KI, Ayoub N, Cohen A, Grewal SI (2002) Establishment and maintenance of a heterochromatin domain. Science 297:2232–2237

Hemleben V, Kovarik A, Torres-Ruiz RA, Volkov RA, Beridze T (2007) Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. Syst Biodivers 5(3):277–289

Hermsen JGT, Taylor LM, van Breukelen EWM, Lipski A (1978) Inheritance of genetic markers from two potato dihaploids and their respective parent cultivars. Euphytica 27:681–688

Heslop-Harrison JS, Brandes A, Taketa S, Schmidt T, Vershinin AV, Alkhimova EG, Kamm A, Doudrick RL, Schwarzacher T, Katsiotis A, Kubis S, Kumar A, Pearce SR, Flavell AJ, Harrison GE (1997) The chromosomal distributions of Ty1-copia group rerotransposable elements in higher plants and their implications for genome evolution. Genetica 100:197–204

Jurka J, Bao W, Kojima K, Kapitonov VV (2011) Repetitive elements: bioinformatic identification, classification and analysis. In: Encyclopedia of life sciences (ELS). Wiley, Chichester. doi:10.1002/9780470015902.a0005270.pub2

Koukalová B, Reich J, Matyásek R, Kuhrová V, Bezdek M (1989) A BamHI family of highly repeated DNA sequences of *Nicotiana tabacum*. Theor Appl Genet 78:77–80

Lamb JC, Meyer JM, Corcoran B, Kato A, Han F, Birchler JA (2007) Distinct chromosomal distributions of highly repetitive sequences in maize. Chromosome Res 15:33–49

Lapitan NLV (1992) Organization and evolution of higher plant nuclear genomes. Genome 35:171–181

Lapitan NLV, Ganal MW, Tanksley SD (1989) Somatic chromosome karyotype of tomato based on in situ hybridization of the TGRI satellite repeat. Genome 32:992–998

Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG (2007) Clustal W and Clustal X version 2.0. Bioinformatics 23(21):2947–2948

Lippman Z, Gendrel AV, Black M, Vaughn MW, Dedhia N, McCombie WR, Lavine K, Mittal V, May B, Kasschau KD, Carrington JC, Doerge RW, Colot V, Martienssen R (2004) Role of transposable elements in heterochromatin and epigenetic control. Nature 430:471–476

Lisch D (2013) How important are transposons for plant evolution? Nat Rev Genet 14(1):49–61

Malkamaki U, Clark MS, Rita H, Valkonen JPT, Pehu E (1996) Analyses of solanaceous species using repetitive genomic DNA sequences isolated from *Solanum brevidens*. Plant Sci 117:121–129

Mroczek RJ, Dawe RK (2003) Distribution of retroelements in centromeres and neocentromeres of maize. Genetics 165:809–819

Mueller LA, Solow TH, Taylor N, Skwarecki B, Buels R, Binns J, Lin C, Wright MH, Ahrens R, Wang Y, Herbst EV, Keyder ER, Menda N, Zamir D, Tanksley SD (2005) The SOL genomics network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138(3):1310–1317

Nagaki K, Talbert PB, Zhong CX, Dawe RK, Henikoff S, Jiang J (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. Genetics 163(3):1221–1225

Nagaki K, Shibata F, Suzuki G, Kanatani A, Ozaki S, Hironaka A, Kashihara K, Murata M (2011) Coexistence of NtCENH3 and two retrotransposons in tobacco centromeres. Chromosome Res 19(5):591–605

Pearce SR, Harrison G, Li D, Heslop-Harrison JS, Kumar A, Flavell AJ (1996) The Ty1-copia group retrotransposons in Vicia species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305–315

Pearce SR, Harrison G, Heslop-Harrison JS, Flavell AJ, Kumar A (1997) Characterization and genomic organization of Ty1-copia group retrotransposons in rye (*Secale cereale*). Genome 40:617–625

Pehu E, Thomas M, Poutala T, Karp A, Jones MGK (1990) Species-specific sequences in the genus *Solanum*: identification, characterization, and application to study somatic hybrids of *S. brevidens* and *S. tuberosum*. Theor Appl Genet 80:693–698

Peterson DG, Schulze SR, Sciara EB, Lee SA, Bowers JE, Nagel A, Jiang N, Tibbitts DC, Wessler SR, Paterson AH (2002) Integration of Cot analysis, DNA cloning, and high-throughput sequencing facilitates genome characterization and gene discovery. Genome Res 12:795–807

Pich U, Schubert I (1998) Terminal heterochromatin and alternative telomeric sequences in Allium cepa. Chromosome Res 6:315–321

Potato Genome Sequencing Consortium (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189–197

Preiszner J, Takacs I, Bilgin M, Gyorgyey J, Dudits D, Feher A (1994) Organization of a *Solanum brevidens* repetitive sequence related to the TGRI subtelomeric repeats of *Lycopersicon esculentum*. Theor Appl Genet 89:1–8

Presting GG, Malysheva L, Fuchs J, Schubert I (1998) A *TY3/GYPSY* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J 16:721–728

Rogers SO, Bendich AJ (1988) Extraction of DNA from plant tissues. Plant Mol Biol Man A6:1–10

Rokka VM, Clark MS, Knudson DL, Pehu E, Lapitan NLV (1998) Cytological and molecular characterization of repetitive DNA sequences of *Solanum brevidens* and *Solanum tuberosum*. Genome 41:487–494

Rouppe van der Voort JN, van Zandvoort P, van Eck HJ, Folkertsma RT, Hutten RC, Draaistra J, Gommers FJ, Jacobsen E, Helder J, Bakker J (1997) Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Mol Gen Genet 255:438–447

Schweizer G, Ganal M, Ninnemann H, Hemleben V (1988) Species-specific DNA sequences for identification of somatic hybrids between *Lycopersicon esculentum* and *Solanum acaule*. Theor Appl Genet 75:679–684

Schweizer G, Borisjuk N, Borisjuk L, Stadler M, Stelzer T, Schilde L, Hemleben V (1993) Molecular analysis of highly repeated genome fractions in *Solanum* and their use as markers for the characterization of species and cultivars. Theor Appl Genet 85:801–808

Stadler M, Stelzer T, Borisjuk N, Zanke C, Schilde-Rentschler L, Hemleben V (1995) Distribution of novel and known repeated elements of *Solanum* and application for the identification of somatic hybrids among *Solanum* species. Theor Appl Genet 91:1271–1278

Stupar RM, Song J, Tek AL, Cheng Z, Dong F, Jiang J (2002) Highly condensed potato pericentromeric heterochromatin contains rDNA-related tandem repeats. Genetics 162:1435–1444

Szinay D, Bai Y, Visser R, de Jong H (2010) FISH applications for genomics and plant breeding strategies in tomato and other solanaceous crops. Cytogenet Genome Res 129:199–210

Tabata S, Kaneko T, Nakamura Y, Kotani H, Kato T, Asamizu E, Miyajima N, Sasamoto S, Kimura T, Hosouchi T et al (2000) Sequence and analysis of chromosome 5 of the plant *Arabidopsis thaliana*. Nature 408(6814):823–826

Tang X, de Boer JM, van Eck HJ, Bachem C, Visser RGF, de Jong H (2009) Assignment of genetic linkage maps to diploid *Solanum*

tuberosum pachytene chromosomes by BAC-FISH technology. Chromosome Res 17:899–915

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

Tek AL, Song J, Macas J, Jiang J (2005) Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics 170:1231–1238

Tomato Genome Consortium (2012) The tomato genome sequence provides insights into fleshy fruit evolution. Nature 485:635–641

Torres GA, Gong Z, Iovene M, Hirsch CD, Buell CR, Bryan GJ, Novak P, Macas J, Jiang J (2011) Organization and evolution of subtelomeric satellite repeats in the potato genome. G3 Genes|Genomes|Genetics 1:85–92

Treangen TJ, Salzberg SL (2012) Repetitive DNA and next-generation sequencing: computational challenges and solutions. Nat Rev Genet 13:36–46

Visser RGF, Hoekstra R, van der Leij FR, Pijnacker LP, Witholt B, Feenstra WJ (1988) In situ hybridization to somatic metaphase chromosomes of potato. Theor Appl Genet 76:420–424

Visser RGF, Bachem CWB, de Boer JM, Bryan GJ, Chakrabati SK, Feingold S, Gromadka R, van Ham RCHJ, Huang S, Jacobs JME, Kuznetsov B, de Melo PE, Milbourne D, Orjeda G, Sagredo B, Tang X (2009) Sequencing the potato genome: outline and first results to come from the elucidation of the sequence of the world's third most important food crop. Am J Potato Res 86:417–429

Volpe TA, Kidner C, Hall IM, Teng G, Grewal SIS, Martienssen RA (2002) Regulation of heterochromatic silencing and histone H3 lysine-9 methylation by RNAi. Science 297:1833–1837

Wenke T, Dobel T, Sorensen TR, Junghans H, Weisshaar B, Schmidt T (2011) Targeted identification of short interspersed nuclear element families shows their widespread existence and extreme heterogeneity in plant genomes. Plant Cell 23:3117–3128

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973–982

Zanke C, Hemleben V (1997) A new *Solanum* satellite DNA containing species-specific sequences which can be used for identification of genome parts in somatic hybrids of potato. Plant Sci 126:185–191

Zhong XB, de Jong JH, Zabel P (1996a) Preparation of tomato meiotic pachytene and mitotic metaphase chromosomes suitable for fluorescence in situ hybridization (FISH). Chromosome Res 4:24–28

Zhong XB, Fransz PF, van Eden JW, Zabel P, van Kammen A, de Jong JH (1996b) High resolution mapping by fluorescence in situ hybridisation to pachytene chromosomes and extended DNA fibers. Plant Mol Biol Rep 14:232–242

Zhong XB, Fransz PF, van Eden JW, Ramanna MS, van Kammen A, Zabel P, de Jong H (1998) FISH studies reveal the molecular and chromosomal organization of individual telomere domains in tomato. Plant J 13:507–517

Zhong CX, Marshall JB, Topp C, Mroczek R, Kato A, Nagaki K, Birchler JA, Jiang J, Dawe RK (2002) Centromeric retroelements and satellites interact with maize kinetochore protein CENH3. Plant Cell 14:2825–2836

Zhu W, Ouyang S, Iovene M, O'Brien K, Vuong H, Jiang J, Buell CR (2008) Analysis of 90 Mb of the potato genome reveals conservation of gene structures and order with tomato but divergence in repetitive sequence composition. BMC Genom 9:286

Zwick MS, Hanson RE, McKnight TD, Islam-Faridi MH, Stelly DM, Wing RA, Price HJ (1997) A rapid procedure for the isolation of Cot-1 DNA from plants. Genome 40:138–142