the plant journal

SEB SOCIETY FOR EXPERIMENTAL BIOLOGY

The Plant Journal (2008) **56**, 627–637 doi: 10.1111/j.1365-313X.2008.03626.x

High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6

Dóra Szinay[1,†], Song-Bin Chang[1,†,‡], Ludmila Khrustaleva[1], Sander Peters[2], Elio Schijlen[2], Yuling Bai[3], Willem J. Stiekema[2,4,5], Roeland C.H.J. van Ham[2,4,5], Hans de Jong[1,5] and René M. Klein Lankhorst[2,5,*]

[1]*Laboratory of Genetics, Wageningen University, Arboretumlaan 4, 6703 BD Wageningen, The Netherlands,*
[2]*Wageningen University and Research Centre, Plant Research International, PO Box 16, 6700 AA Wageningen, The Netherlands,*
[3]*Wageningen UR Plant Breeding, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands,*
[4]*Laboratory of Bioinformatics, Wageningen University, Dreijenlaan 3, 6703 HA Wageningen, The Netherlands, and*
[5]*Centre for Biosystems Genomics, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands*

Received 22 February 2008; revised 5 June 2008; accepted 4 July 2008; published online 15 August 2008.
[*]For correspondence (fax +31 317 477266; e-mail rene.kleinlankhorsst@wur.nl).
[†]These authors have contributed equally to this work.
[‡]Present address: Department of Agronomy, National Taiwan University, Taipei 10617, Taiwan.

Summary

Within the framework of the International Solanaceae Genome Project, the genome of tomato (*Solanum lycopersicum*) is currently being sequenced. We follow a 'BAC-by-BAC' approach that aims to deliver high-quality sequences of the euchromatin part of the tomato genome. BACs are selected from various libraries of the tomato genome on the basis of markers from the F2.2000 linkage map. Prior to sequencing, we validated the precise physical location of the selected BACs on the chromosomes by five-colour high-resolution fluorescent *in situ* hybridization (FISH) mapping. This paper describes the strategies and results of cytogenetic mapping for chromosome 6 using 75 seed BACs for FISH on pachytene complements. The cytogenetic map obtained showed discrepancies between the actual chromosomal positions of these BACs and their markers on the linkage group. These discrepancies were most notable in the pericentromere heterochromatin, thus confirming previously described suppression of cross-over recombination in that region. In a so called pooled-BAC FISH, we hybridized all seed BACs simultaneously and found a few large gaps in the euchromatin parts of the long arm that are still devoid of seed BACs and are too large for coverage by expanding BAC contigs. Combining FISH with pooled BACs and newly recruited seed BACs will thus aid in efficient targeting of novel seed BACs into these areas. Finally, we established the occurrence of repetitive DNA in heterochromatin/euchromatin borders by combining BAC FISH with hybridization of a labelled repetitive DNA fraction (Cot-100). This strategy provides an excellent means to establish the borders between euchromatin and heterochromatin in this chromosome.

Keywords: fluorescent *in situ* hybridization, tomato, genome sequencing, multi-colour FISH, Cot-100, plant chromosomes.

Introduction

In 2004, the International Solanaceae Genome Project launched an initiative to sequence the euchromatin part of the genome of tomato (*Solanum lycopersicum*) as the focus of its systems approach to increase diversity and adaptation in crop plants (Mueller *et al.*, 2005). Tomato was chosen as the model for the Solanaceae because it has a relatively small genome size of approximately 950 Mb (Arumuganathan and Earle, 1991), numerous lines, mutants and chromosomal variants, a saturated genetic map and outstanding chromosome morphology with well-differentiated euchromatin and heterochromatin regions in all 12 chromosomes. While the heterochromatic regions constitute approximately 75% of

the genome and are believed to have low gene content (Khush *et al.*, 1964; Peterson *et al.*, 1996 Rick, 1971; Van der Hoeven *et al.*, 2002), the remaining 25% of the DNA is organized into long continuous stretches of gene-rich euchromatin blocks (Peterson *et al.*, 1996).

Within the framework of the International Solanaceae Genome Project, a collective of several Dutch research groups is sequencing tomato chromosome 6. This chromosome was selected because it harbours several economically important genes such as the root-knot nematode resistance gene *Mi-1* (Ammiraju *et al.*, 2003; Deberdt *et al.*, 1999; Kaloshian *et al.*, 1998; Van Daelen *et al.*, 1993; Zhong *et al.*, 1999) and *Oidium lycopersicum* resistance genes (Huang *et al.*, 2000). In addition, substantial genetic and physical information on this chromosome had been generated previously as chromosome 6 has been used in various genetic map studies (Liharska *et al.*, 1997; Van Wordragen *et al.*, 1994, 1996; Weide *et al.*, 1993), studies on the molecular organization of paracentromere (pericentromere) sequences (Weide *et al.*, 1998), genetic analysis of alien chromosomal segments of introgression hybrids (Liharska *et al.*, 1996), and high-resolution FISH of the TGR1 tandem and telomere repeats (Zhong *et al.*, 1998). Also, chromosome 6 has been studied as a monosomic addition in a tetraploid potato background (De Jong *et al.*, 2000).

At pachytene, chromosome 6 is easily distinguishable by its centromere position and characteristic heterochromatin blocks in the long and short chromosome arms (Ramanna and Prakken, 1967; Zhong *et al.*, 1998). In addition, the chromosome has the lowest euchromatin percentage of the complement, previously estimated at approximately 20 Mb of euchromatin and 33.4 Mb of heterochromatin (Peterson *et al.*, 1996).

The tomato sequencing project follows the BAC-by-BAC approach, which has also been successfully applied to sequence the genomes of rice (International Rice Genome Sequencing Project, 2005) and *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000). Currently, the genome of *Medicago truncatula* is also being sequenced using this strategy (Young *et al.*, 2005). In the BAC-by-BAC approach or BAC-walking procedure (Peters *et al.*, 2006), the first step involves anchoring of a limited number of BAC clones to the genome that will then serve as starting points for further BAC contig building and sequencing. The anchoring of these so-called 'seed-BACs' is carried out by screening BAC libraries with genetic markers and subsequent linking of retrieved BACs to the genetic loci defined by the used markers. The accuracy and reliability of the anchoring process is highly dependent on the quality of the genetic map from which the genetic markers were derived. As the exact map locations of genetic markers and relative positions between markers cannot always be determined unequivocally, especially in genomic regions in which recombination is suppressed (Sherman and Stack, 1995), verification of the positions of anchored seed BACs is an absolute requirement in generating a genome sequence by BAC walking.

In tomato, fluorescent *in situ* hybridization (FISH) on pachytene complements has successfully been applied to chromosome identification, study of meiotic chromosome pairing, and positioning of heterochromatin and euchromatin, for instance (De Jong *et al.*, 1999). BAC clones have large genomic inserts of 50–150 kb that makes them most suitable for FISH studies on pachytene chromosomes. FISH can therefore be applied as a reliable technology to verify the position of anchored seed BACs on tomato chromosomes. However, large inserts often contain long stretches of tandem and dispersed repetitive sequences, especially when they originate from pericentromere and telomere heterochromatin regions. The use of such BACs for FISH produces abundant fluorescence signals over many loci due to excessive cross-hybridization of the repetitive sequences in the probe. This problem was circumvented by using the repeat fraction of genomic DNA, Cot-100, to suppress hybridization of the repetitive sequences in the BACs on the chromosomal target (Budiman *et al.*, 2004; Chang *et al.*, 2007).

In this paper, we describe new variants of multi-colour FISH that are very powerful for processing larger numbers of BACs on pachytene complements, without losing the detailed morphology of heterochromatin. We show how improved BAC FISH is indispensible in constructing a backbone of anchored seed BACs on the euchromatin part of tomato chromosome 6. The new method also allows the identification of chromosomal areas with low seed BAC coverage, and subsequent specific targeting of novel seed BACs towards these areas, and defines more accurately the borders of heterochromatin and euchromatin with respect to repeat content.

Results

Enhanced imaging of the pachytene chromosome morphology

Digital acquisition of DAPI-stained pachytene complements produced clear and sharp images of the chromosomes. It was observed that grey-scale images showed chromosome morphology far better than the dark-blue images produced by a colour camera or images obtained from scanned colour slides or negative films. We reduced the dynamic range of the DAPI images to dark/medium-grey (fewer than 200 grey levels) to avoid the bright grey tones of the chromosomes dominating the small pseudo-coloured signals of the fluorescent BAC probes in the final image overlay. A second important improvement was the use of a Hi-Gauss high-pass spatial filter and application of contrast correction of the DAPI image to compensate for slight blurring of the CCD

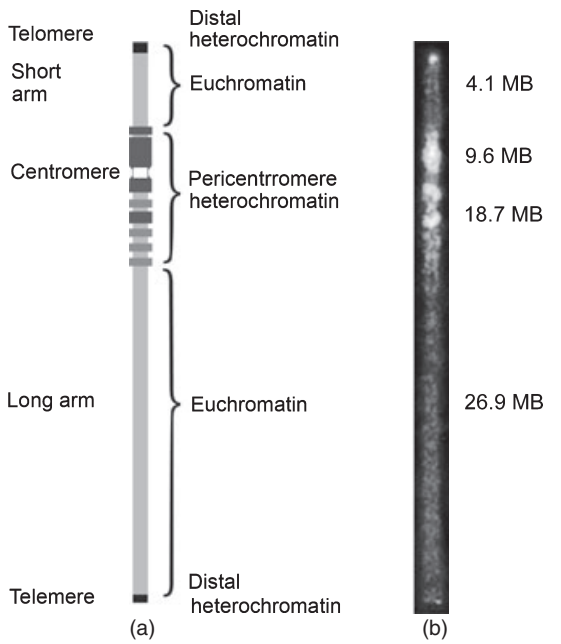

Figure 1. Genomic and cytogenetic characteristics of chromosome 6 of tomato (*Solanum lycopersicum*).
(a) Schematic representation of the pachytene chromosome, with estimations of DNA size in the euchromatin and heterochromatin regions.
(b) Straightened pachytene bivalent stained with DAPI.

image, thus producing better definition for accentuating minor details in chromomeres and heterochromatin banding of the chromosomes. We also straightened the chromosomes for better comparison of FISH patterns from different chromosomes.

Figure 1 shows an example of the improved DAPI staining image of chromosome 6. The chromosome clearly has an asymmetric centromere position, and its short arm is the smallest euchromatin region in the complement. Recent estimates of the euchromatin/heterochromatin proportions of the pachytene complement gave higher values of euchromatin in the short (4.1 Mb) and long arms (26.9 Mb) (Figure 1, and Chang *et al.*, 2008). DAPI staining reveals all diagnostic chromatin morphology, including the distal heterochromatin blocks of short and long arms, the large short-arm pericentromere block and the two long-arm pericentromere heterochromatin regions, the structural centromere region, and many tiny chromomeres in the euchromatin (Ramanna and Prakken, 1967). The polymorphic long-arm heterochromatin knob as described by Zhong *et al.* (1998) was not visible here, and may be even absent in the plant material that was used for our FISH experiments.

Multi-colour FISH

In order to improve the accuracy and efficiency of BAC detection, we chose a five-colour FISH protocol based on BAC probes directly labelled with fluorophores for blue, green, orange, red and far-red fluorescence. Pilot experiments with combinatorial and ratio labelling schemes as used for mammalian multicolour FISH studies gave biased interpretations of overlapping BAC signals, and so were no longer considered for multi-colour BAC detection. For verification purposes, we combined several sets of differentially labelled BACs in a single experiment, allowing mapping of multiple BACs in one hybridization experiment. Figure 2 shows a typical example in which the positions of three sets of five BACs labelled with five pseudo-colours were determined. These sets of BACs represented clones previously mapped by FISH to the short arm of chromosome 6, the pericentromere heterochromatin and the long arm, respectively. The high-resolution image allowed ordering of these BACs, as well as precise localization in relation to the centromere (Figure 2, arrow). In addition, the BACs could also be mapped in relation to the pericentromere heterochromatin domains on both the short arm and the long arm of the chromosome. However, a closer look at these boundaries, especially of the long-arm pericentromere, demonstrated a gradual transition of brightly fluorescing heterochromatin to the weaker euchromatin, such that the borders of the euchromatin region could not be defined unequivocally.

Previous studies on the composition of the pericentromere in tomato have revealed high amounts of various repetitive elements, including the TGR-II and TGR-III repeats (Ganal *et al.*, 1988; Lapitan *et al.*, 1989; Schweizer *et al.*, 1988), microsatellites (Broun and Tanksley, 1996) and retrotransposons of the Ty1-copia family and other families of retrotransposons (Chang *et al.*, 2008). The greater proportion of these repeats can be isolated as the so-called Cot-100 fraction using reassociation kinetics-based DNA-isolation techniques (Peterson *et al.*, 1998). Using Cot-100 as a probe in FISH, Chang *et al.* (2008) demonstrated that this repeat fraction of the tomato genome covers the heterochromatic areas of the chromosomes and can be used to assess more precisely the borders of heterochromatin and euchromatin. Here, we show that Cot-100 in combination with BACs in a multi-colour FISH provides a more robust indicator of repeats at the heterochromatin borders, even in cases where these borders are not clear in DAPI stained chromosomes. Figure 3 gives an example of a Cot-100 FISH in combination with the BACs 304P16 and 082G10 that are at the borders of the short- and long-arm pericentromeres, respectively (S.P., unpublished results). As shown in the figure, Cot-100 is more sensitive in demonstrating repeats in regions that were classified as less-condensed euchromatin on the basis of DAPI fluorescence intensity, and so is more informative about the repeat content of the chromosome region around the BAC.

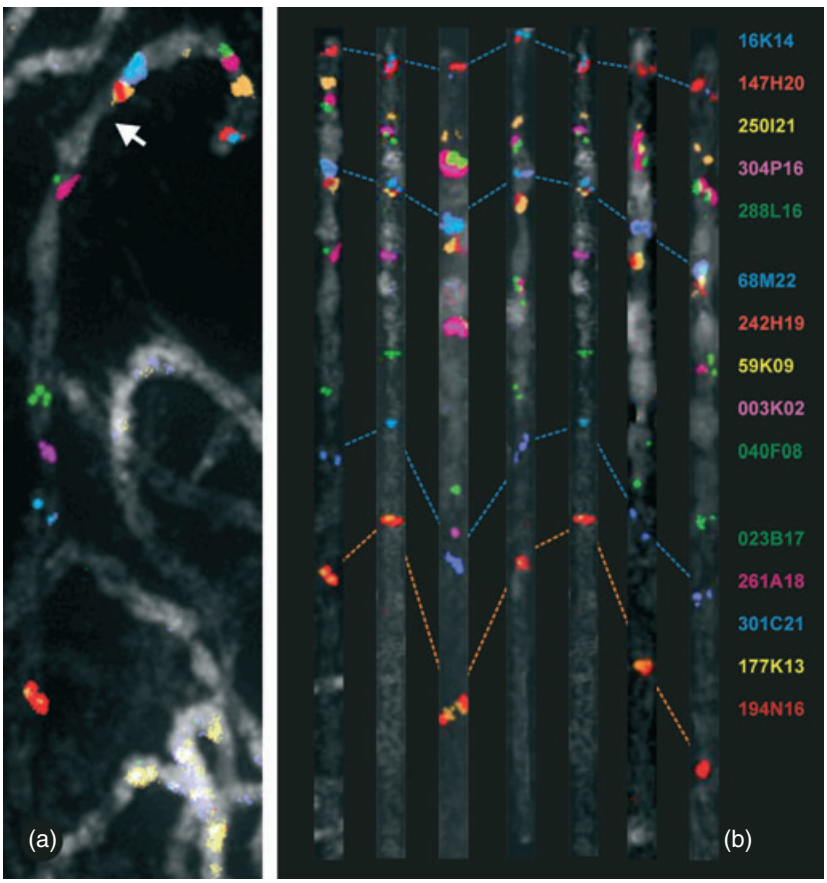

Figure 2. Multi-colour FISH using three sets of five seed BACs.
(a) Detail of chromosome pair 6 in a pachytene complement. The arrow indicates the position of the centromere.
(b) Seven examples of straightened bivalents, with the coloured names of the 15 different BACs.

Confirmation and physical mapping of seed BACs

Aided by the available genetic markers for chromosome 6 on the F2.2000 genetic map (Fulton *et al.*, 2002), a total of 75 candidate seed BACs were retrieved from available BAC libraries. Each of these BACs was tested either individually or in sets using multi-colour FISH to verify the proposed location on chromosome 6 as predicted by the marker locations. Also, for each BAC it was determined whether the physical location occurred in euchromatin or in heterochromatin. Of the 75 BACs analysed, 51 were confirmed as *bona fide* seed BACs, as FISH clearly confirmed that these BACs are in the euchromatin of chromosome 6. Five BACs were discarded because they were found either in the pericentromere heterochromatin or in the centromere of chromosome 6. An additional 19 BACs were rejected either because the probe showed multiple FISH signals in the pericentromeres of most or all chromosomes, or because they gave single foci on one of the other chromosomes.

Few BACs gave FISH positions that differed substantially from their predicted position on the linkage map (Figure 4). Such discrepancies occurred mostly for BACs in the pericentromere, as shown in Figure 4, or near the distal regions of the short and long arms where cross-over recombination is known to be suppressed. In order to determine the suppression of recombination along the short arm of chromosome 6, we calculated Mb/cM ratios for a number of BACs. The distance of each BAC from the top of the short arm of chromosome 6 was measured, and distances obtained were converted into Mb based on mean ratio values of 6.3 Mb μm^{-1} for heterochromatin and 0.6 Mb μm^{-1} for euchromatin (Budiman *et al.*, 2004). From these measurements, it was estimated that recombination frequency is reduced approximately ninefold from the central euchromatic region of the short arm via the pericentromeric heterochromatin to the centromere (Figure 4). It was thus concluded that the observed discrepancy between genetic map positions and FISH positions for a number of BACs can most likely be attributed to the absence of cross-overs in the pericentromere.

Determination of chromosomal coverage by pooled-BAC FISH

Sufficient and even coverage of the target genome with seed BACs is of the utmost importance in BAC-by-BAC genomic sequencing projects. Uncovered parts of chromosomes represent major gaps, sometimes referred to as 'oceans' or 'seas', that are hard to close by extending seed BACs into

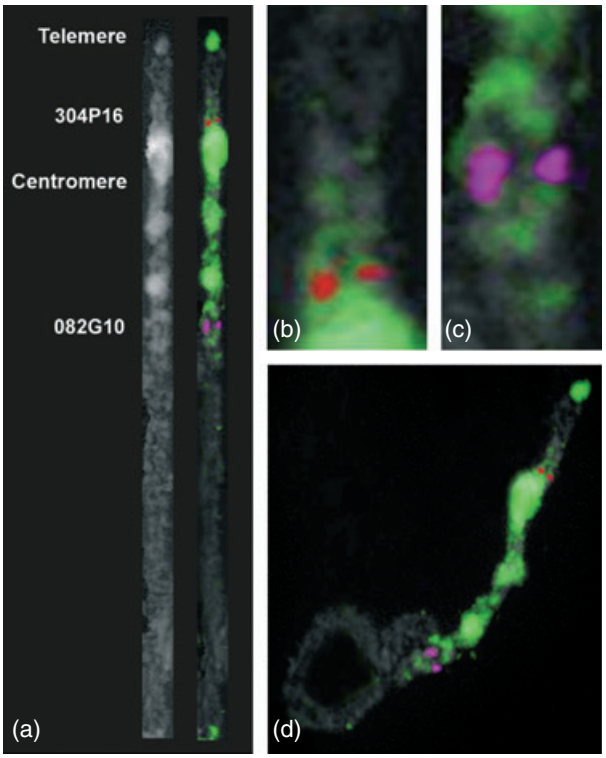

Figure 3. Labeling of the heterochromatic part of tomato chromosome 6 by FISH with the Cot-100 genomic DNA fraction (green signal).
The differently labelled BAC clones residing in the heterochromatin/euchromatin borders of the short arm and of the long arm are pseudo-coloured in red and magenta, respectively. The images (b) and (c) are detailed magnifications of (a) and (d).

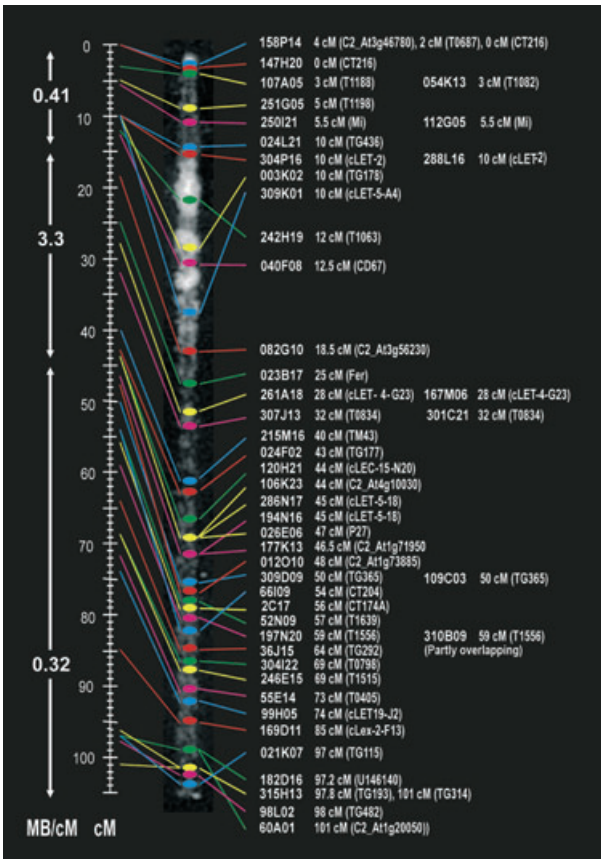

Figure 4. Overview of the BAC FISH map of chromosome 6, based on the position of seed BACs on the chromosome.
The figure clearly shows some reversal of order between genetic map and chromosome map positions at the distal ends and pericentromeric region of tomato chromosome 6. The double-headed arrows on the left indicate the short-arm euchromatin, pericentromere and long-arm euchromatin, with estimates of Mb cM^{-1}.

longer contigs. This BAC-walking approach allows closure of only relatively small gaps between adjacent seed BACs, but 'oceans' of Mb size cannot generally be crossed. Instead, in these cases, additional seed BACs have to be placed within the large gaps.

Coverage of tomato chromosome 6 with the validated set of 51 seed BACs was assessed using a novel pooled-BAC FISH approach in which all available seed BACs per chromosome arm were labelled with a single fluorochrome and subsequently hybridized in a single experiment. Figures 5 and 6 show the results of these experiments for the short and long chromosome arms, respectively.

For the short-arm pooled-BAC FISH, we used 18 BACs that cover the majority of the short-arm euchromatin and distal heterochromatin (approximately 4.1 Mb), and only three significant gaps were evident (Figure 5g), which span a euchromatin area between 600 and 1000 kb. For the long arm, the pooled-BAC FISH experiment was performed using a set of 33 BACs derived from the large euchromatic region (approximately 26.9 Mb). As shown in Figure 6, these BACs were distributed over the entire euchromatin, but at least five major gaps were evident, covering euchromatin regions of up to several Mb. These gaps could either be the consequence of a bias in the genetic maps and/or BAC libraries used, or, alternatively, result from random distribution of BACs along the chromosome. To test this hypothesis, we simulated the distribution of 33 BACs randomly plotted over a length of 26.9 Mb using a standard spreadsheet function for generating random real numbers (data not shown). The results of such a simulation confirmed that such gaps can indeed occur in a small sample of BACs that are randomly distributed over the chromosome. Hence, there are no indications of an experimental error underlying the lack of BAC coverage for some parts of chromosome 6.

Targeting the gaps on the long arm

As the observed physical gaps on the long arm are too large to be closed effectively by BAC walking, a novel screening of the available BAC libraries was necessary to retrieve additional seed BACs located within these gaps. To target novel seed BACs that may reside in one of the gaps, we tested the

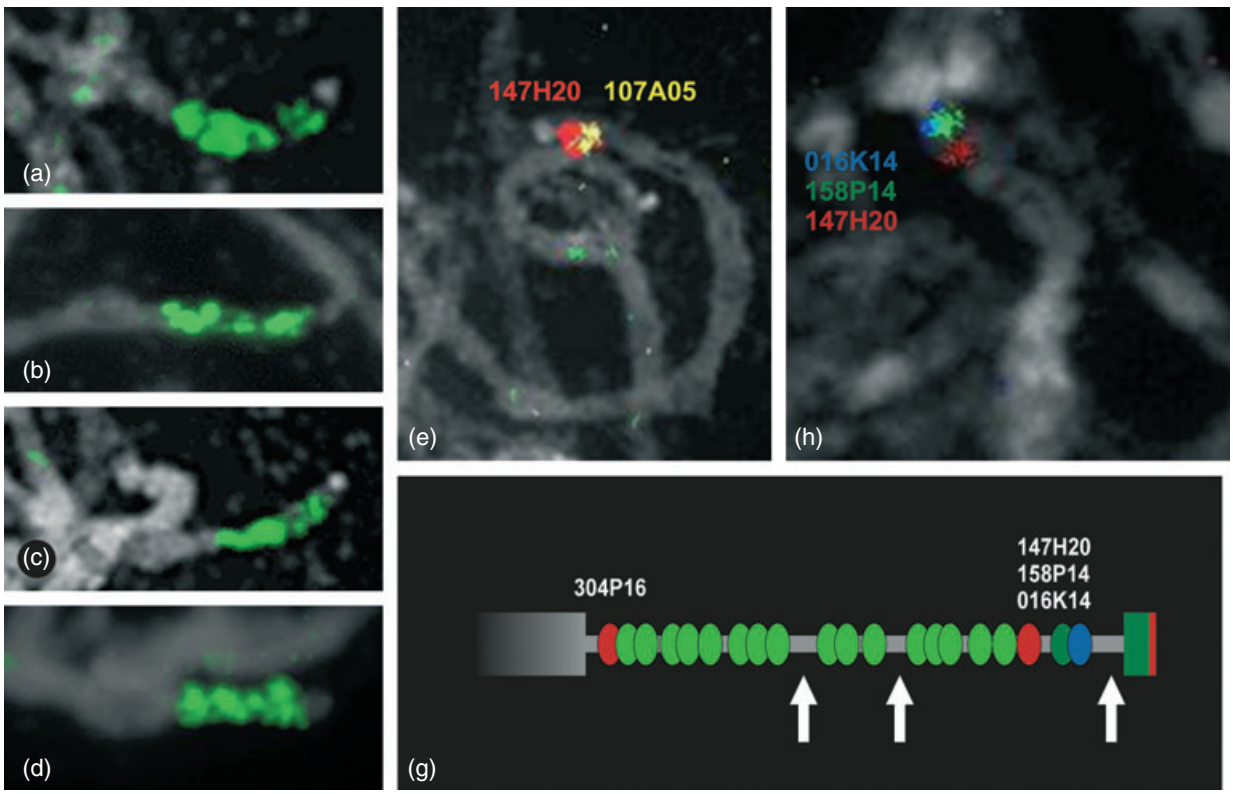

Figure 5. Pooled BAC FISH for the short arm of chromosome 6.
(a)–(d) Compilation of four different pachytene spreads after hybridization with 18 seed BACs.
(e, f) Multi-colour FISH showing proximal and terminal BACs for the euchromatin part of the short arm.
(g) Schematic representation showing the gaps (arrows) between the seed BACs and the positions of proximal and terminal BACs.

assumption that suppression of recombination is limited on the euchromatin part of the long arm of tomato chromosome 6, and hence that a reasonably good correlation exists between the genetic map and the physical map for this part of the tomato genome. Any large physical gap thus would correspond to a gap in the genetic map, showing the positions of the available seed BACs for the long arm.

As illustrated in Figure 6(a), the genetic map of anchored seed BACs on chromosome 6 contains five major gaps, ranging in size between 3 and 12 cM. For each of these gaps, markers residing within them or bordering them were used to screen the available BAC libraries for novel candidate seed BACs, which yielded three novel BACs. The physical location of these new BACs was assessed by a combination of pooled BAC FISH and multi-colour FISH. All 33 confirmed seed BACs for the long arm were labelled with the red-fluorescing Cy3.5, and the additional new candidate seed BACs were labelled with FITC (green), Cy3 (orange) and DEAC (blue), and were used simultaneously in a single FISH experiment (Figure 6b).

The BACs M082G10 and H023B17 were obtained from library screening with the genetic markers C2_At3g56230 and *Fer*, respectively. These markers are specific for gap I (Figure 6a), and it can thus be concluded that this gap on the genetic map of the seed BACs for chromosome 6 corresponds to the large physical gap proximal to the centromere (Figure 6b). BAC M012J12 corresponds to genetic marker C2_At1g16870, showing that the physical gap proximal to the telomere of chromosome 6 corresponds to genetic gap V [between cLEX-2-F13 (85 cM) and TG115 (97 cM)]. For the remaining three gaps, no novel seed BACs have yet been retrieved from the BAC libraries. These results show that gaps in the BAC assembly for the long-arm euchromatin of chromosome 6 coincidence well with gaps in the genetic map.

Discussion

For sequencing of complex eukaryotic genomes, two approaches are in use: BAC-by-BAC sequencing and whole-genome shotgun sequencing. In general, the BAC-by-BAC sequencing approach starts with construction of a sequence-ready minimal tiling path of overlapping BAC clones that are anchored to the genome using molecular markers present in the BACs. These BACs are then separately sequenced to a high accuracy of normally less than one error per 10 000 bases. The subsequent assembly of all obtained BAC sequences thus results in a high-quality

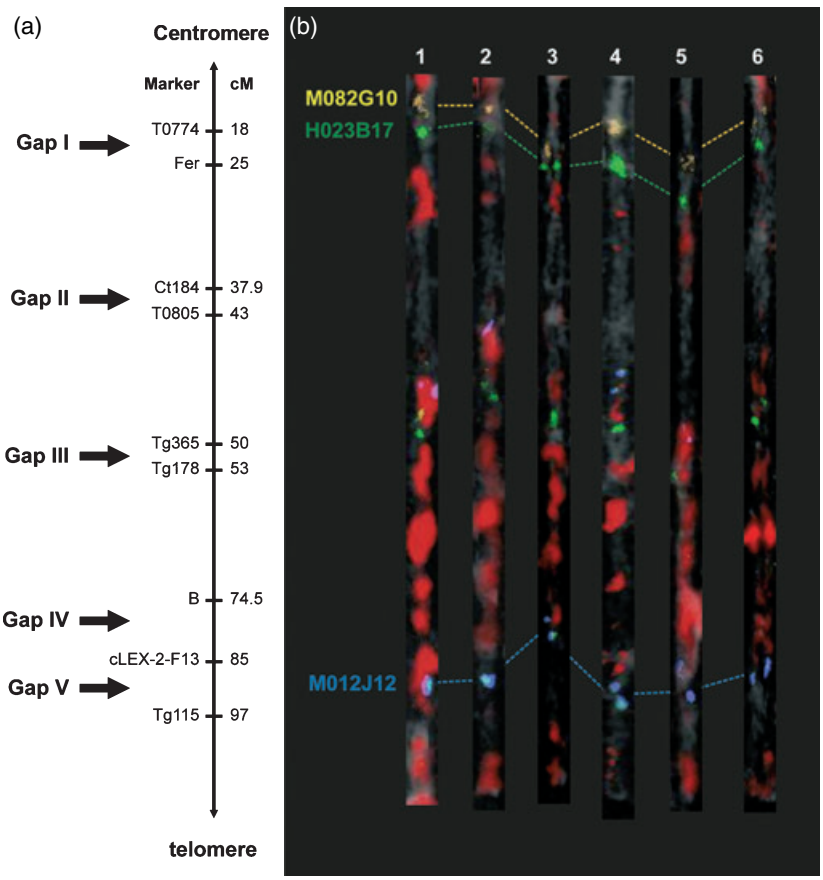

Figure 6. Pooled-BAC FISH of the long arm.
(a) Genetic map of tomato chromosome 6 showing gaps lacking genetically anchored seed BACs.
(b) Multi-colour pooled-BAC FISH. The red fluorescence signals are the pooled seed BACs of the long arm of chromosome 6. BACs targeted specifically towards gaps in the genetic map are shown in yellow (M082G10), green (H023B17) and blue (M012J12).

('golden standard') genomic sequence. However, a drawback is the rather lengthy and thus expensive process needed for constructing the minimal tiling path and the subsequent sequencing of individual BACs. The BAC-by-BAC approach has been successfully applied for the sequencing of *Caenorhabditis elegans* (Ainscough *et al.*, 1998), *Arabidopsis thaliana* (Arabidopsis Genome Initiative, 2000) and the japonica rice variety *Oryza sativa* ssp. *Nipponbare* (International Rice Genome Sequencing Project, 2005), for example.

The alternative approach of whole-genome shotgun sequencing has a major advantage in that it can be accomplished in a short period of time and thus is faster and cheaper than the BAC-by-BAC approach. In whole-genome shotgun sequencing, an entire genome is fragmented and cloned into libraries with BACs having various insert sizes. Clones from each of these libraries are sequenced until a certain genomic coverage is reached. Then, all obtained sequences are assembled into contigs and larger supercontigs that are mapped to the genome using molecular markers identified in the contigs. A drawback of this method is that it quite difficult to progress from a good draft of the genome to a high-quality, completely finished genome. The whole-genome shotgun approach has been used to sequence, amongst others, the genomes of *Drosophila melanogaster* (Adams *et al.*, 2000), *Fugu rubripes* (Aparicio *et al.*, 2002) *Oryza sativa* ssp. *Indica* (Yu *et al.*, 2002), grapevine (Jaillon *et al.*, 2007) and poplar (Tuskan *et al.*, 2006). Hybrid approaches in which a partial whole-genome shot gun assembly was combined with a partial BAC-by-BAC assembly have been used in sequencing of the mouse genome (Waterston *et al.*, 2002), for example.

The BAC-by-BAC approach has also been chosen to sequence the euchromatin fraction of the tomato (*Solanum lycopersicum*) genome. However, in this particular case, the strategy used to sequence this partial genome differs from the 'classical' BAC-by-BAC approach as no minimal tiling path of BAC clones was established prior to the large-scale sequencing. Instead, a large number of so-called 'seed BACs' or small to medium-sized contigs of seed BACs were anchored to the genome using molecular markers prior to the actual sequencing. After sequencing of the seed BACs, new contigs are built and existing contigs are extended by identifying overlapping BACs in a database containing the sequences of approximately 400 000 BAC ends (http://www.sgn.cornell.edu). If no further extension BACs can be retrieved from the database, additional seed BACs are retrieved from the available BAC libraries by new rounds of marker screening. In this way, the minimal tiling path of BAC clones is constructed while sequencing, which can be

regarded as a 'map-as-you-go' strategy (Peters *et al.*, 2006). A key prerequisite in this strategy is a robust protocol to confirm seed BAC positions on the chromosomes, as mis-allocated BACs will result in the building and subsequent sequencing of contigs on the wrong part of the tomato genome. As the tomato sequencing project only covers the euchromatin part of the genome, such errors would seriously hamper the progress of the project.

We have shown that the use of FISH on pachytene complements is an outstanding method for ascertaining the physical position of BACs. The long and well-differentiated pachytene chromosomes allow accurate determination of chromosome size and identity, as well as unequivocal identification of centromeres, telomeres, and, to some extent, the borders of euchromatin and heterochromatin. Also, use of FISH at pachytene along with extended DNA fibres allows the measurement of physical distances, and thus can be used to estimate the physical distances between BACs and/or repeats (Zhong *et al.*, 1998).

In this study, we have developed some technical modifications, including reduced and sharpened grey display of the DAPI pachytene morphology for better accentuation of minor chromomeres and heterochromatin domains of the chromosomes. We have described the use of Cot-100, not only to block off excessive repeats in the BAC probes from hybridization to the chromosomal targets, but also for accurate FISH detection of repeat-rich regions (Cot-100 BAC FISH), which is far more accurate and versatile than identifying heterochromatin/euchromatin borders on the basis of DAPI fluorescence intensity. It allows a more objective assessment of repetitive DNAs and BACs on or near the euchromatin/heterochromatin borders. The third important improvement is the use of five-colour FISH, which strongly enhanced the efficiency of accurate mapping of larger numbers of BACs. A related procedure was developed, referred to as 'pooled-BAC FISH', which allowed hybridization of larger numbers of pooled BAC clones. This method directly reveals the gaps in the euchromatin not covered by previously confirmed BACs, and assesses positioning of newly acquired BACs in the gaps.

The methods described above were effective in confirming and positioning 75 potential seed BACs for tomato chromosome 6. The BACs were derived either from the Sol Genomics Network (SGN) database at Cornell University (http://www.sgn.cornell.edu/), which forms the main seed BAC repository for the tomato genome project, or from AFLP screening of the available BAC libraries by the Dutch consortium sequencing tomato chromosome 6. Of the 75 BACs analysed, 51 BACs were confirmed as occurring in the euchromatin of chromosome 6, whereas 19 BACs were located on one of the other chromosomes, or FISH of these BACs resulted in signals on multiple chromosomes. These mis-allocations of BACs probably reflect false positives obtained in the screening of the BAC libraries rather than erroneously mapped markers on the tomato genetic map. The remaining five BACs were discarded due to their location in heterochromatin instead of their expected location within the euchromatin. These BACs were either located close to the telomeres or close to the centromere, suggesting that the mis-allocation of such BACs is probably due to mapping errors caused by suppression of recombination near the tomato centromere (Sherman and Stack, 1995) and telomeres. On average, a genetic distance of 1 cM on the tomato map corresponds to approximately 750 kb (Tanksley *et al.*, 1992), but different values have been reported for distinct fractions of the tomato genome. Ganal *et al.* (1989) calculated a value of 4 Mb cM^{-1} near the centromere of tomato chromosome 9, but higher ratios of 21.74 Mb cM^{-1} and 100 Mb cM^{-1} were found for the short- and long-arm pericentromere heterochromatin of chromosome 12, respectively (Budiman *et al.*, 2004). Tor *et al.* (2002) calculated a value of 330 kb cM^{-1} for the euchromatin regions of chromosome 2L, and this ratio is less than half the mean ratio of 750 kb cM^{-1} for the tomato genome (Tanksley *et al.*, 1992). Our own data showed Mb cM^{-1} ratios for chromosome 6 ranging from 0.41 for the short-arm euchromatin, 3.3 for the pericentromere heterochromatin, and 0.32 for the long-arm euchromatin.

Similar high variability of recombination rate along chromosomes has also been observed in other plant species, for example Arabidopsis and rice. For the Arabidopsis genome, maximum local recombination rates approximately 30–70-fold greater than the genome average have been reported (Drouaud *et al.*, 2006; Singer *et al.*, 2006). Analyses of rice chromosome 4 showed that recombination rates can vary up to approximately 30-fold along the chromosome (Zhao *et al.*, 2002). The available data for tomato also show that the recombination rate is highly variable along the chromosome, and many 'hot spots' and 'cold spots' for recombination seem to occur.

Using the pooled-BAC FISH protocol, it was shown that good coverage with seed BACs has already been accomplished for chromosome 6. The short arm in particular is covered to a great extent by seed BACs, and only a few small gaps remain. For the long arm, a number of major gaps still have to be bridged. The sizes of these gaps range from an estimated 1.36 Mb for gap III to 6.45 Mb for gap II (H.d.J., unpublished results). Computer simulations with BACs suggest that the observed gaps are probably not due to bias in the F2.2000 genetic map or in the constructed BAC libraries, but rather reflect the outcome of a random distribution of the 33 seed BACs over the long arm. As the observed physical gaps for the long-arm euchromatin of chromosome 6 coincide with gaps in the genetic map of tomato, novel seed BACs specific for these gaps can probably easily be obtained by increasing the marker density on the genetic map. The experiments described here show that this approach is feasible. However, for

regions in the tomato genome where cross-overs are absent or suppressed, large physical gaps will correspond to small gaps on the genetic map. For these kinds of physical gaps, it will be harder to identify novel seed BACs by molecular marker screening, and, as a consequence, these physical gaps probably can only be closed by BAC walking.

Using the FISH applications developed here, we have been able to build a reliable backbone to guide the sequencing of tomato chromosome 6. Many of the seed BACs and contigs of seed BACs have already been extended (S.P., unpublished results), which has resulted in a complete BAC tiling path of the short arm. Further extending of BAC contigs of the long arm, in combination with targeting BACs towards the remaining gaps, will complete the sequencing of the entire euchromatin of tomato chromosome 6.

Experimental procedures

Chromosome preparations

Young flower buds of tomato *Solanum lycopersicum* cv. VFNT Cherry (LA1221) were fixed in freshly prepared Carnoy's fixative (acetic acid:ethanol, 1:3) for 1 day, and could be stored in 70% ethanol at 4°C for several months. We selected buds with anthers containing pollen mother cells at meiotic prophase I, and rinsed them three times in distilled water and once in 10 mM sodium citrate buffer (pH 4.5) before transferring to an enzyme mix containing 1% pectolyase Y23 (Sigma P-3026), 1% cellulase RS (Yakult 203033, Yakult Pharmaceutical, Tokyo, Japan) and 1% cytohelicase (Bio Sepra 24970-014) in citrate buffer for 3 h at 37°C. The material was then left on ice until further use. We dissected two or three anthers from a flower bud, transferred them to 30–40 μl 60% acetic acid, and squeezed the anther tissue carefully with fine needles to release the pollen mother cells. Very clean grease-free slides were held face down into the steam of boiling water for 1–3 sec, and then turned over, and 8–10 μl of the cell suspension were dropped onto the humid surface of the slide and the liquid was spread gently with the pipette tip. The slide was then put on a 55°C hot plate and 10 or 11 drops of 60% acetic acid were dropped onto the cells for further maceration. After 2–3 min, the slides were covered with 50 μl Carnoy's fixative, air-dried, post-fixed in 1% formaldehyde solution (in PBS, pH 6.8), air-dried again and stored at 4°C until further use. We screened all slides under a phase-contrast microscope and selected late-pachytene pollen mother cells with little or no cytoplasm, good chromosome spreading and well-differentiated chromatin morphology. The steam treatment was found to be essential to obtain well-spread pachytene chromosomes, and the formaldehyde treatment makes the chromatin more resistant to degeneration during the FISH procedure.

Cot-100 DNA

Tomato Cot-100 DNA was prepared as described by Zwick *et al.* (1997) with some modifications. Total genomic DNA was isolated using the cetyltrimethylammonium bromide (CTAB) method and was sonicated to a fragment size of about 1 kb. We denatured 0.5 μg μl^{-1} of this fragmented DNA in 0.3 M NaCl at 95°C for 10 min, and then let it re-anneal at 65°C (Peterson *et al.*, 1998) for 37 h 40 min. The remaining ssDNA was digested with S1 endonuclease (Fermentas, http://www.fermentas.com, final concentration 1 U μg^{-1}) for 90 min at 37°C. The reaction was stopped and DNA was extracted by adding 300 μl chloroform:iso-amylalcohol (24:1). Then the DNA solution layer was transferred to a new tube, 2.5 volumes of ice-cold absolute alcohol were added, and the mixture was kept overnight at -20°C, before centrifugation at 14 000 ***g*** and 4°C for 30 min. The dry pellet was resuspended in 20 μl HB50, pH 8.0. For every new batch of Cot-100 DNA, we used a small batch of labelled DNA in a hybridization to check that the probe covered all heterochromatin.

BAC DNA isolation

Tomato BACs were obtained by screening the tomato Heinz 1706 *Hin*dIII BAC library (Budiman *et al.*, 2000) and the Heinz 1706 *Mbo*I BAC library (http://www.sgn.cornell.edu/about/tomato_sequencing.pl) with chromosome 6-specific markers derived from the EXPEN F2000 genetic map (Fulton *et al.*, 2002). BAC DNA was isolated using a standard alkaline extraction method and the High Pure PCR product purification kit (Roche, http://www.roche.com) for extra cleaning, and labelled by standard nick translation (Roche).

Fluorescence in situ *hybridization and image capturing*

For indirect detection of BACs in two-colour FISH experiments, we labelled the BAC DNA with biotin-16-dUTP or digoxigenin-11-dUTP, and visualized the probe using standard streptavidin–Texas Red and anti-digoxigenin–FITC detection protocols (Chang *et al.*, 2008), respectively. For direct labelling in the multi-colour FISH, we labelled BAC DNA with dUTP-DEAC (Perkin-Elmer, http://www.perkinelmer.com), dUTP-FITC (Perkin-Elmer), dUTP-Cy3 (Amersham, http://www5.amershambiosciences.com/), dCTP-Cy3.5 (Amersham) and dUTP-Cy5 (Amersham). For dUTP-Cy5, we used a tenfold lower concentration of dTTP in the nick translation mix. All further details of the FISH experiments have been described previously (Zhong *et al.*, 1996). Hybridization of the repetitive sequences in the BAC DNA was suppressed by adding unlabelled Cot-100 (10 times probe concentration). Chromosomes were counterstained in 5 μg ml^{-1} DAPI in Vectashield anti-fade (Vector Laboratories, http://www.vectorlabs.com). Slides were examined under a Zeiss Axioplan 2 imaging photomicroscope (http://www.zeiss.com/) equipped with epifluorescence illumination, and small band filter sets for DAPI, DEAC, FITC, Cy3, Cy3.5/Texas Red and Cy5 fluorescence. Selected images were captured using a Photometrics Sensys 1305 × 1024 pixel CCD camera (Photometrics, http://www.photomet.com). Image processing and thresholding were performed using Genus image analysis software (Applied Imaging Corporation, http://www.aicorp.com). DAPI images were displayed in dark- to medium-grey and sharpened using a Hi-Gauss high-pass spatial filter to accentuate minor details and heterochromatin morphology of the chromosomes. The remaining fluorescence images were pseudo-coloured in blue (DEAC), green (FITC), orange (Cy3), red (Cy3.5, Texas Red) and purple (Cy5), and merged in multichannel mode. Chromosome straightening was performed using with the 'straighten-curved-objects' plug-in of Image J (Kocsis *et al.*, 1991), and final image optimization was performed using Photoshop (Adobe Inc., http://www.adobe.com).

Acknowledgements

This research was co-financed by the European Commission (EU-SOL Project PL 016214) and by the Centre for BioSystems Genomics (CBSG), which is part of the Netherlands Genomics Initiative/Netherlands Organization for Scientific Research.

References

Adams, M.D., Celniker, S.E., Holt, R.A. *et al.* (2000) The genome sequence of *Drosophila melanogaster*. *Science*, **287**, 2185–2195.

Ainscough, R., Bardill, S., Barlow, K. *et al.* (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science*, **282**, 2012–2018.

Ammiraju, J.S.S., Veremis, J.C., Huang, X., Roberts, P.A. and Kaloshian, I. (2003) The heat-stable root-knot nematode resistance gene *Mi-9* from *Lycopersicon peruvianum* is localized on the short arm of chromosome 6. *Theor. Appl. Genet.* **106**, 478–484.

Aparicio, S., Chapman, J., Stupka, E. *et al.* (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science*, **297**, 1301–1310.

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

Arumuganathan, K. and Earle, E. (1991) Estimation of nuclear DNA content of plants by flow cytometry. *Plant Mol. Biol. Rep.* **9**, 208–218.

Broun, P. and Tanksley, S.D. (1996) Characterization and genetic mapping in simple repeat sequences in the tomato genome. *Mol. Gen. Genet.* **250**, 39–49.

Budiman, M.A., Mao, L., Wood, T.C. and Wing, R.A. (2000) A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. *Genome Res.* **10**, 129–136.

Budiman, M.A., Chang, S.B., Lee, S., Yang, T.J., Zhang, H.B., de Jong, J.H. and Wing, R.A. (2004) Localization of *jointless-2* gene in the centromeric region of tomato chromosome 12 based on high resolution genetic and physical mapping. *Theor. Appl. Genet.* **108**, 190–196.

Chang, S.-B., Anderson, L.K., Sherman, J.D., Royer, S.M. and Stack, S.M. (2007) Predicting and testing physical locations of genetically mapped loci on tomato pachytene chromosome 1. *Genetics*, **176**, 2131–2138.

Chang, S-B., Yang, T-J., Datema, E., van Vugt, J., Vosman, B., Kuipers, A., Meznikova, M., Szinay, D., Klein Lankhorst, R., Jacobsen, E. and de Jong, H. (2008) FISH mapping and molecular organization of the major repetitive sequences of tomato. *Chromosome Research*, DOI: 10.1007/s10577-088-1249-z.

De Jong, J.H., Fransz, P. and Zabel, P. (1999) High resolution FISH in plants – techniques and applications. *Trends Plant Sci.* **4**, 258–262.

De Jong, J.H., Zhong, X.-B., Fransz, P.F., Wennekes-van Eden, J., Jacobsen, E. and Zabel, P. (2000) High resolution FISH reveals the molecular and chromosomal organisation of repetitive sequences of individual tomato chromosomes. In *Chromosomes Today*, Vol. 13 (Olmo, E. and Redi, C.A., eds). Basel, Switzerland: Birkhäuser Verlag, pp. 267–275.

Deberdt, P., Olivier, J., Thoquet, P., Queneherve, P. and Prior, P. (1999) Evaluation of bacterial wilt resistance in tomato lines nearly isogenic for the *Mi* gene for resistance to root-knot. *Plant Pathol.* **48**, 415–424.

Drouaud, J., Camilleri, C., Bourguignon, P.-Y. *et al.* (2006) Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination 'hot spots'. *Genome Res.* **16**, 106–114.

Fulton, T.M., Van der Hoeven, R., Eannetta, N.T. and Tanksley, S.D. (2002) Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. *Plant Cell*, **14**, 1457–1467.

Ganal, M.W., Lapitan, N.L.V. and Tanksley, S.D. (1988) A molecular and cytogenetic survey of major repeated DNA sequences in tomato *Lycopersicon esculentum*. *Mol. Gen. Genet.* **213**, 262–268.

Ganal, M.W., Young, N.D. and Tanksley, S.D. (1989) Pulsed-field gel electrophoresis and physical mapping of large DNA fragments in the Tm-2a region of chromosome 9 in tomato. *Mol. Gen. Genet.* **215**, 395–400.

Huang, C.C., de Putte, P.M.H.V., van der Meer, J.G.H., Dekens, F.M. and Lindhout, P. (2000) Characterization and mapping of resistance to *Oidium lycopersicum* in two *Lycopersicon hirsutum* accessions: evidence for close linkage of two Ol-genes on chromosome 6 of tomato. *Heredity*, **85**, 511–520.

International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

Jaillon, O., Aury, J.M., Noel, B. *et al.* (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. *Nature*, **449**, 463–467.

Kaloshian, I., Yaghoobi, J., Liharska, T. *et al.* (1998) Genetic and physical localization of the root-knot nematode resistance locus Mi in tomato. *Mol. Gen. Genet.* **257**, 376–385.

Khush, G.S., Rick, C.M. and Robinson, R.W. (1964) Genetic activity in a heterochromatic chromosome segment of the tomato. *Science*, **145**, 1432–1434.

Kocsis, E., Trus, B.L., Steer, C.J., Bisher, M.E. and Steven, A.C. (1991) Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment. *J. Struct. Biol.* **107**, 6–14.

Lapitan, N.L.V., Ganal, M.W. and Tanksley, S.D. (1989) Somatic chromosome karyotype of tomato based on in situ hybridization of the TGRI satellite repeat. *Genome*, **32**, 992–998.

Liharska, T.B., Koorneef, M., Van Wordragen, M., Van Kammen, A. and Zabel, P. (1996) Tomato chromosome 6: effect of alien chromosomal segments on recombinant frequencies. *Genome*, **39**, 485–491.

Liharska, T.B., Hontelez, J., Van Kammen, A., Zabel, P. and Koornneef, M. (1997) Molecular mapping around the centromere of tomato chromosome 6 using irradiation-induced deletions. *Theor. Appl. Genet.* **95**, 969–974.

Mueller, L.A., Tanksley, S.D., Giovannoni, J.J. *et al.* (2005) The Tomato Sequencing Project, the first cornerstone of the International Solanaceae Project (SOL). *Comp. Funct. Genomics*, **6**, 154–158.

Peters, S.A., van Haarst, J.C., Jesse, T.P., Woltinge, D., Jansen, K., Hesselink, T., van Staveren, M.J., Abma-Henkens, M.C.J. and Klein Lankhorst, R.M. (2006) TOPAAS, a tomato and potato assembly assistance system for selection and finishing of bacterial artificial chromosomes. *Plant Physiol.* **140**, 805–817.

Peterson, D.G., Price, H.J., Johnston, J.S. and Stack, S.M. (1996) DNA content of heterochromatin and euchromatin in tomato (*Lycopersicon esculentum*) pachytene chromosomes. *Genome*, **39**, 77–82.

Peterson, D.G., Pearson, W.R. and Stack, S.M. (1998) Characterization of the tomato (*Lycopersicon esculentum*) genome using *in vitro* and *in situ* DNA reassociation. *Genome*, **41**, 346–356.

Ramanna, M.S. and Prakken, R. (1967) Structure of and homology between pachytene and somatic metaphase chromosomes of tomato. *Genetica*, **38**, 115–133.

Rick, C.M. (1971) Some cytogenetic features of the genome in diploid species. *Stadler Symp.* **1**, 153–174.

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

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

Singer, T., Fan, Y., Chang, H.S., Zhu, T., Hazen, S.P. and Briggs, S.P. (2006) A high-resolution map of Arabidopsis recombinant inbred lines by whole-genome exon array hybridization. *PLoS Genet.* **2**, 1352–1361.

Tanksley, S.D., Ganal, M.W., Prince, J.P. *et al.* (1992) High density molecular linkage maps of the tomato and potato genomes. *Genetics*, **132**, 1141–1160.

Tor, M., Manning, K., King, G.J., Thompson, A.J., Jones, G.H., Seymour, G.B. and Amstrong, S.J. (2002) Genetic analysis and FISH mapping of the colourless non-ripening locus of tomato. *Theor. Appl. Genet.* **104**, 165–170.

Tuskan, G.A., DiFazio, S., Jansson, S. *et al.* (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). *Science*, **15**, 1596–1604.

Van Daelen, R.A.J.J., Gerbens, F., Van Ruissen, F., Aarts, J., Hontelez, J. and Zabel, P. (1993) Long-range physical maps of two loci (*Aps-1* and *GP79*) flanking the root-knot nematode resistance gene (*Mi*) near the centromere of tomato chromosome 6. *Plant Mol. Biol.* **23**, 185–192.

Van der Hoeven, R., Ronning, C., Giovannoni, J., Martin, G. and Tanksley, S.D. (2002) Deductions about the number, organization and evolution of genes in the tomato genome based on analysis of a large EST collection and selective genomic sequencing. *Plant Cell*, **14**, 1441–1456.

Van Wordragen, M.F., Weide, R., Liharska, T., Van der Steen, A., Koornneef, M. and Zabel, P. (1994) Genetic and molecular organization of the short arm and pericentromeric region of tomato chromosome 6. *Euphytica*, **79**, 169–174.

Van Wordragen, M.F., Weide, R.L., Coppoolse, E., Koornneef, M. and Zabel, P. (1996) Tomato chromosome 6: a high resolution map of the long arm and construction of a composite integrated marker-order map. *Theor. Appl. Genet.* **92**, 1065–1072.

Waterston, R.H., Lindblad-Toh, K., Birney, E. *et al.* (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature*, **420**, 520–562.

Weide, R., Van Wordragen, M.F., Klein Lankhorst, R.M., Verkerk, R., Hanhart, C., Liharska, T., Pap, E., Stam, P., Zabel, P. and Koornneef, M. (1993) Integration of the classical and molecular linkage maps of tomato chromosome 6. *Genetics*, **135**, 1175–1186.

Weide, R., Hontelez, J., Van Kammen, A., Koornneef, M. and Zabel, P. (1998) Paracentromeric sequences on tomato chromosome 6 show homology to human satellite III and to the mammalian CENP-B binding box. *Mol. Gen. Genet.* **259**, 190–197.

Young, N.D., Cannon, S.B., Sato, S., Kim, D., Cook, D.R., Town, C.D., Roe, B.A. and Tabata, S. (2005) **Se**quencing the gene spaces of *Medicago truncatula* and *Lotus japonicus*. *Plant Physiol.* **137**, 1174–1181.

Yu, J., Hu, S., Wang, J. *et al.* (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). *Science*, **296**, 79–92.

Zhao, Q., Zhang, Y., Cheng, Z. *et al.* (2002) A fine physical map of rice chromosome 4. *Genome Res.* **12**, 817–823.

Zhong, X.B., de Jong, J.H. and Zabel, P. (1996) Preparation of tomato meiotic pachytene and mitotic metaphase chromosomes suitable for fluorescence in situ hybridization (FISH). *Chromosome Res.* **4**, 24–28.

Zhong, X.-B., Fransz, P.F., Wennekes van Eden, J., Ramanna, M.S., Van Kammen, A. and Zabel, P. (1998) FISH studies reveal the molecular and chromosomal organization of individual telomere domains in tomato. *Plant J.* **13**, 507–517.

Zhong, X.-B., Bodeau, J., Fransz, P.F., Williamson, V.M., van Kammen, A., de Jong, J.H. and Zabel, P. (1999) FISH to meiotic pachytene chromosomes of tomato locates the root-knot nematode resistance gene Mi-1 and the acid phosphatase gene Aps-1 near the junction of euchromatin and pericentromeric heterochromatin of chromosome arms 6S and 6L, respectively. *Theor. Appl. Genet.* **98**, 365–370.

Zwick, M.S., Hanson, R.E., McKnight, T.D., Islam-Faridi, M.N., Stelly, D.M., Wing, R.A. and Price, H.J. (1997) A rapid procedure for the isolation of Cot-1 DNA from plants. *Genome*, **40**, 138–142.

