The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*) Dóra Szinay 2010

The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*)

Dóra Szinay

The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*)

Dóra Szinay

Thesis committee

Thesis Supervisor

Prof. Dr. R.G.F. Visser
Professor of Plant Breeding
Wageningen University

Thesis Co-supervisors

Dr. J.H. de Jong
Associate professor at the Laboratory of Genetics
Wageningen University

Dr. Y. Bai
Assistant professor at the Laboratory of Plant Breeding
Wageningen University

Other members

Prof. Dr. H.J.M. Dons, Wageningen University
Prof. Dr. Ir. P.C. Struik, Wageningen University
Prof. Dr. A.G.M. Gerats, Radboud University Nijmegen
Dr. P. Fransz, University of Amsterdam

This research was conducted under the auspices of the Graduate School of Experimental Plant Sciences.

The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*)

Dóra Szinay

Thesis
submitted in fulfilment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 17 May 2010
at 1.30 p.m. in the Aula.

Dóra Szinay
The development of FISH tools for genetic, phylogenetic and breeding studies in tomato (*Solanum lycopersicum*)
147 pages

Thesis, Wageningen University, Wageningen, NL (2010)
With references, with summaries in Dutch and English
ISBN 978-90-8585-635-1

Contents

Chapter 1

General introduction:

FISH technology and its relevance to genomics, phylogenetics and plant breeding

A genome can be considered as all DNA sequences of an organism together. However, we also can sea it as all its chromosomes and all its genes. Most eukaryote species, including higher plants and vertebrates have two genomes in their nuclei, one from father and one from mother, and their size can vary enormously. Studying genomes can be done by genetic mapping, physical mapping, sequencing and chromosome analysis. Genetic or linkage maps are based on recombination frequencies between loci, which depend on the occurrence of crossovers between the homologous chromosomes. Loci are the positions of genes and molecular markers on the chromosome. One chromosome normally represents one linkage group. Physical maps show the position and true distances (base pairs, kilobases, megabases) of DNA fragments along chromosomes. The ratio between genetic and physical distances between DNA markers (*e.g.*, cM/kb) can vary enormously along chromosomes as crossovers are uneaully distributed along the chromosome, and even entirely absent in large heterochromatin blocks and around the centromere. DNA sequencing reveals the nucleotide information of certain chromosome or DNA fragments. With modern sequencing technologies DNA is now cloned in a vector and then attached to beads that are amplified by polymerase chain reaction (PCR). Then parallel sequencing follows of a maximum read length of around 500-700 bp either by synthesis or hybridization and ligation of oligonucleotides (Hall 2007). Chromosome analysis is carried out by cytogenetic mapping that is well-represented nowadays by DNA::DNA Fluorescence *in situ* Hybridization (FISH). In FISH experiments (Figure 1) single strand probe DNA, which is obtained from total genomic DNA sequences, repetitive sequences or single copy DNA, is directly or indirectly labeled with a fluorescent dye and hybridized on denatured chromosomes, nuclei, chromatin or DNA fibers. FISH provides unique and essential information about the location of single copy sequences, repetitive DNA and chromosome structure (Schmidt and Heslop-Harrison 1998). In plants as well as in mammals, FISH is a very versatile and wide-spread technology that is applied in genetics, genomics, molecular biology, plant breeding and phylogenetics.

In this chapter, I will review and discuss a number of important applications of FISH technology, focusing on crop plants, with special attention to tomato and potato, and related species of the *Solanaceae* family. I will describe the main characteristics of genome painting, repeat bar-coding FISH, chromosome painting and BAC FISH painting for detection of single copy sequences.

Chromosome banding and painting in plants

Differences between plants and mammals

Higher plants and animals have few, but major differences in the molecular organization of their chromosomes. Firstly, chromosome banding pattern is more complex and discriminative in mammals than it is in plants. The five most common banding methods are C-, N-, G-, Q- and

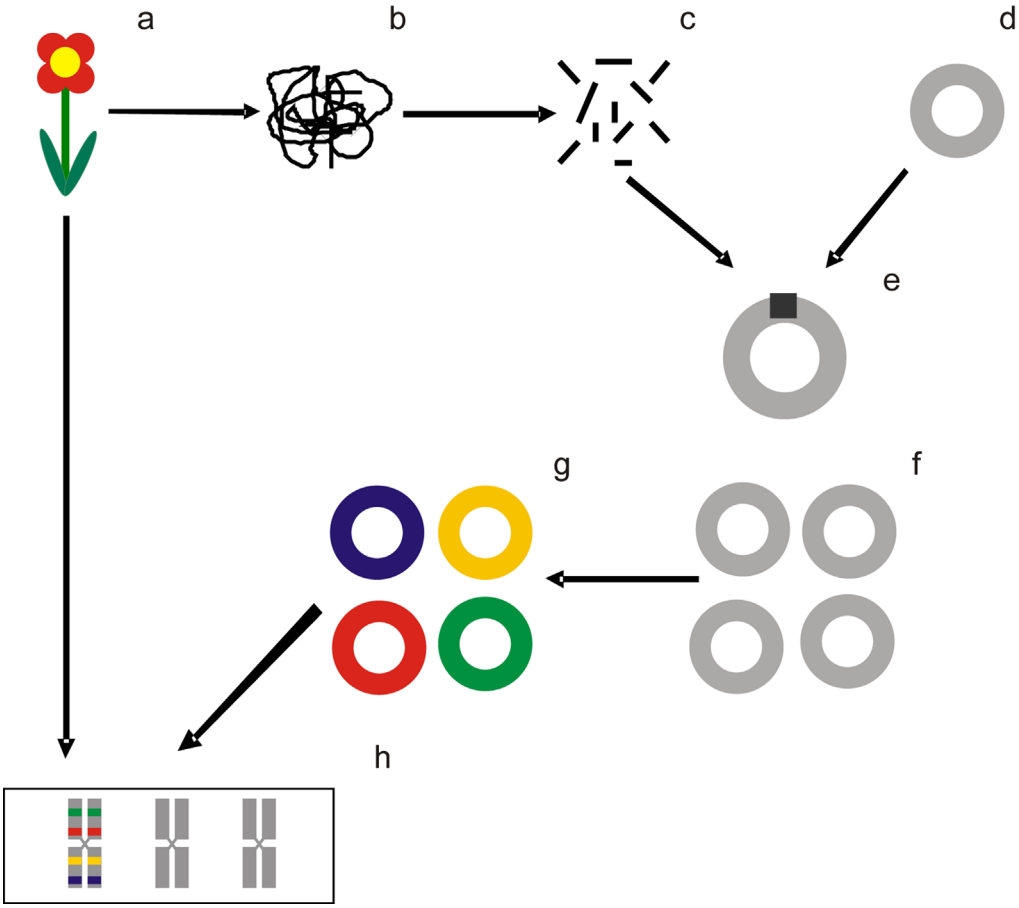


Figure 1. Scheduled overview of the FISH procedure. a = plant of interest; b = isolated plant DNA; c = fragmented plant DNA; d = bacteria plasmid; e = cloned plant DNA in bacteria plasmid (BAC); f = multiplied BACs; g = fluorescence labeled BACs; h = slide with spread plant DNA and hybridized with specific BACs.

R-banding (for overview, see Sumner 1991). C-bands show the big heterochromatin blocks around the centromere, distal blocks of the chromosomes and the nucleolar organizer region (NOR). It is performed by alkaline denaturation of the DNA and mild reannealing, followed by staining with Giemsa. N-banding distinguishes the nucleolar organizer region (NOR) or displays a banding pattern that resembles C-banding. The method can be carried out by DNA extraction in trichloracetic and hydrochloric acid, followed by staining in Giemsa (or relarted dyes) or silver nitrate. G-banding, which is based on mild trypsin digestion of the chromosomes, results in darkly stained heterochromatin and light euchromatin by Giemsa staining. Q-banding is a staining with the fluorescence dye quinacrine mustard dye, which binds to guanine (Caspersson et al. 1968). R-banding method uses heated (87 °C) phosphate buffer to

denature the chromosomes which are then stained with Giemsa. R-banding results in darkly painted bands that appear as weak bands in G- and Q-banding (Dutrillaux and Lejeune 1971).

Although most plants have simple C-bands in the centromere region, only few demonstrate more complex C- and N-banding (Greilhuber 1977). Most plants with large genomes, such as wheat (Gill and Kimber 1974), *Secale cereale* (Appels and Peacock 1978), *Vicia faba* (Cionini et al. 1985) and *Petunia hybrida* (Dietrich et al. 1981) have big C-bands. This banding patterns represent the sites of constitutive heterochromatin and exist of mostly satellite and other tandem arrays. Some plants, such as *Cestrum aurantiacum* (Berg and Greilhuber 1993) and *Vicia faba* (Schubert and Rieger 1979) have N-banding patterns as well. In most plants none of the banding technique can be used for chromosome identification, except in very few cereals such as *Aegilops ovata* (Landgeva and Ganeva 1996) and *Medicago sativa* (Masoud et al. 1991). Due to the most more powerful FISH technology, people have lost their interest in chromosome banding.

A second difference between mammals and plant chromosomes is the morphology of prophase chromosomes. In meiotic prophase (pachytene) mammals have fuzzy and relatively less extended chromosomes than plants. Comparing to plants, mammals have shorter pachytene complements,, which form lampbrush-like or diffuse chromatin (Klašterska 1978). In contrast, plants at pachytene stage form well distinguishable chromosomes with clearly defined euchromatin and heterochromatic regions (De Jong et al. 1999).

The third difference deals with the molecular organisation of dispersed repeats that in plant chromosomes are more homogenized than in mammals. As a result, plant chromosomes often have similar GC:AT ratios, whereas mammal chromosomes can be very different in that respect. Plant chromosomes carry similar nucleotide organization in every species, except the rDNA region, which is GC rich (Schmidt and Heslop-Harrison 1998). These differences give the opportunity in mammals to flow sort large numbers of identical chromosomes for the production of chromosome-specific DNA libraries. Flow sorting is a technique that can select specific chromosomes from a metaphase chromosome suspension in a small capillary flow, on the basis of different size and staining with AT and GC specific fluorochromes. In plants, the use of flow sorting is limited and could so far only be used in few cases like the selection of wheat chromosome 3B (Vrana et al. 2000) and chromosome 2 of tomato, that carries the NOR region (Arumuganathan et al. 1994). Isolation of specific chromosomes or chromosome parts by microdissection using a fine needle on a microscope slide was also successfully in a few studies, like in wheat to discriminate the long arm of chromosome 5 of the B genome (Vega et al. 1994).

The final goal of isolating large numbers of chromosomes by flow sorting or microdissection was the production of chromosome specific DNA libraries for chromosome painting studies. However, this goal has never become successful. Schubert et al. (2001) failed to compare genomes of morphologically very different *Asteraceae* species by using microdissected chromosomes. This was due to the similar dispersed repeat content, which is one of the major criteria that differ between plants and mammals. More discussion about the failure of chro-

mosome painting is given in Schubert et al. (2001). More successful for plants is the painting technique based on large insert genomic clones, like cosmids, YACs or BACs as probes that were selected for low repeat content. In practice the most common clones are BACs (Lysak and Lexer 2006).

Repeat painting

Cot DNA for total repetitive fraction of the genome

Isolation of Cot DNA is based on differential DNA reassociation of single stranded DNA, in which highly repetitive sequences reanneal first, then the middle repetitive fraction and finally the single copy sequences (Peterson et al. 1998). By isolating a reannealed DNA fraction after a certain time, followed by S1-endonuclease digestion of the single strand molecules, double stand DNA is obtained that contains a pool of repeats of a specific complexity. The experimental conditions for the different Cot fractions can be calculated using genomic DNA size, GC/AT ration and reannealing conditions. Short reannealing will give only highly repetitive sequences, reannealing of a few days the highly and middle repeat fraction etc. This method has been tested and applied extensively on tomato by Peterson et al. (1998) and Chang et al. (2008) showing that Cot 100 is the pool of repeats that has all the repetitive sequences to paint all heterochromatin regions.

There are two major applications of Cot 100 for FISH analysis. Firstly, unlabeled Cot DNA can be used to block off repeats from the BAC probes when single copy sequences have to be mapped. This application is now standard in our laboratory for all BACs containing lots of repeats (Szinay et al. 2008). Secondly, Cot DNA as a probe in FISH can be used to visualize all the repeats of the genome (Chang et al. 2008; Szinay et al. 2008). Tomato was one of the first crop plants where Cot analysis was carried out (Peterson et al. 1998).

Painting of specific genomic repeats

A substantial part of genomic DNA of most angiosperms is repetitive, even in species with a small genome size. There are two major repeat classes based on organization or position; tandem repeats and dispersed repeats. Examples of tandem repeats are rDNA, satellite DNA and telomere repeats. They mostly localize on large conspicuous heterochromatic DNA blocks at the distal ends and interstitial parts of the chromosome (Schmidt and Heslop-Harrison 1998). Microsatellites or simple sequence repeats are also examples of tandem arrays, but their motifs are very short, typically 2-6 bp, and they are distributed all along the chromosomes, both in heterochromatin and euchromatin (Cuadrado and Schwarzacher 1998; Cuadrado and Jouve 2007a,b; Chang et al. 2008).

Dispersed repeats that spread mostly on the heterochromatin can be subdivided into retroelements (amplified via RNA intermediates) and transposable elements (amplified via DNA). Retroelements includes retrotransposons, retroposons and retrosequences. Retrotransposons can change the structure of a genome dramatically (Zhang et al. 2004a), both in terms of significant genome size differences as well as by inserted into genes causing its inactivation, or influence its expression and recombination (Levin 2002). The types of retrotransposons can be very different between distantly related species, such as sorghum, rice and maize, although the gene order and sequence are highly conserved within the repeats (Schmidt and Heslop-Harrison 1998). Transposable elements contribute to mutations which might cause gene inactivation as well as chromosomal rearrangement. Transposable elements are mostly inactive in plants due to epigenetic silencing, but can be reactivated under conditions of heavy physiological or genomic including inbreeding or distant hybridization (McClintock 1984).

Distribution of repetitive sequences can be highly different between plant genomes, but in almost all cases they are concentrated in the pericentromere heterochromatin, like in tomato (Chang et al. 2008), *Arabidopsis thaliana*, *Medicago truncatula*, *Sorghum bicolor* etc. Number and size of repeat domains can even differ between close genotypes or accessions and so can be used as chromosomal markers in karyotype analysis and chromosome markers in a segregating population. For example, Kato et al. (2004) showed distinction between 14 maize lines using most of the main repeat classes, such as NOR, 5S rDNA, centromere specific repeat of maize, microsatellite, sub telomeric repeat. There is no clear correlation between repeat distribution and genome size compared to *A. thaliana*, *M. truncatula* and *S. bicolor* (Pedrosa-Harand et al. 2009). Tessadori et al. (2007) characterized the heterochromatin regions in different cells of *A. thaliana*. They observed that all major repetitive sequences decondensed, except for the 45S rDNA, as a consequence of heterochromatin reduction.

Repeat painting in the Solanaceae family

The most ubiquitous repeats in eukaryotes are the telomeres (TTTAGGG or sometimes TTAGGG) and the 45S rDNAs and 5S rDNAs. The telomeres have been studies extensively in species of the Solanaceae family, and show mostly the Arabidopsis-type telomere (TTTAGGG). They occur in *Nicotiana* species as well as in *S. tuberosum* and *S. lycopersicum*, however, they are missing in *Cestrum* (Sykorová et al. 2003), which has the human TTAGGG type. Schweizer et al. (1988) characterized tandem repeats from *S. acaule* and *S. lycopersicum* by FISH, which seemed to be good markers to distinguish the parental species in their somatic hybrids. Tek et al. (2005) showed that satellite repeats (such as Sobo) could change and develop quite quickly in *S. bulbocastanum*. Furthermore, Sobo was not found in any other *Solanum* species and was not present in some *S. bulbocastanum* accessions. Sobo has high sequence similarity with a retroelement (Sore1), suggesting that Sobo has gone through a fast and dramatic evolutionary development from that retroelement.

Lapitan et al. (1989) analyzed the major tomato repeat classes, which they referred to as Tomato Genomic Repeat TGRI, TGRII and TGRIII. Later Chang et al. (2008) described a new repeat class in the proximal part of the pericentromere and the centromere, TGRIV. In addition, his studied various microsatellites, rDNAs, Ty1-*Copia* and several Cot fractions. All these repeats were well characterized by FISH and organized in six chromatin classes including euchromatin, NOR region, distal and interstitial heterochromatin, pericentromere heterochromatin, chromomeres, structural centromere. The tandem repeat TGRI hybridized to most of the centromeres and proximal ends on both arms and to interstitial knobs along the long arms. TGRI has high sequence similarity within the tomato clade and in *S. lycopersicoides*, but only 3/4 similarity to *S. brevidens* and *S. circaeifolium* (Hemleben et al. 2007). TGRII and TGRIII hybridized mostly on the pericentromere region, TGRIV is a component of the structural centromere but also paints the flanking pericentromere. On chromosome 2 the NOR region where 45S rDNA localize was rich in [GACA] sequences. Other microsatellites, like [GA] and [GATA] and the retrotransposon Ty1-*Copia* localized in the pericentromere heterochromatin. A recent sequencing study of chromosome 6 BACs revealed that Ty1-*Copia* occur in both euchromatin and heterochromatin, in the sense that the ratio of Ty3-*Gypsy* and Ty1-*Copia* elements is 2:3 euchromatin and 3:2 in heterochromatin (Peters et al. 2009).

Single copy painting strategies in plants

Single copy sequences are mostly localized in the euchromatin part of the genomes. These regions are spread all over the chromosomes except for the highly condensed pericentromere heterochromatin and proximal ends. *Arabidopsis*, *Brassica*, cucumber, tomato and potato are few of the many examples of plant species where the euchromatin is clearly distinguishable from the highly condensed heterochromatin. In other species, like rice, maize, cereals, *Taraxacum* and *Petunia* the euchromatin is interrupted by many short stretches of heterochromatic knobs, chromocentres, most likely resulting from a different distribution of retrotransposons (de Jong: personal communication).

FISH mapping is usually carried out on interphase, metaphase or pachytene chromosome complements. Several *Diptera* species have long, well-differentiated polytene chromosomes that consists of a large number of tightly paired homologous interphas chromosomes, which give the opportunity to distinguish species just by their banding patterns and identify genes by deletion mapping and FISH studies. In some plants this type of interphase chromosomes does occur in the suspensor or related tissues of the developing embryo with high metabolic activity and short lifespan (Nagl 1978). Furthermore chromatids are not tightly paired and chromosomes are condensed in the pericentromere and the proximal ends. The polytene chromosomes of *Phasaeolus coccineus* and the closely related *Vigna unguiculata* were in this way successfully differentiated by FISH (Guerra 2001).

Metaphase chromosomes have a condensed structure with an equally low spatial resolution, and hence makes FISH mapping for species with small or middle sized chromosomes

limited. However, for the study of mammal chromosomes, it is still the most favorite stage for karyotype analysis and FISH as chromosomes have Q-or G-banding (discussed above) and miss a well-defined pachytene morphology. In plants the use of pachytene chromosomes is the best alternative for metaphase FISH, as chromosomes in this stage are on average 15 times longer than metaphase (Ramanna and Prakken 1967; de Jong et al. 1999). Here we can distinguish the condensed heterochromatin, the decondensed euchromatin areas and chromomeres (small condensed chromosome regions) along the chromosomes. However, the large genomes of wheat, lily and onion have so many dispersed repeats, that analysis of pachytene FISH is very difficult and cumbersome. One of the few large genome plant species with an exceptional pachytene morphology is maize. Lamb et al. (2007a) used successfully repeat free chromosome specific probes for FISH analysis on pachytene complements to investigate conserved regions in three maize introgression lines.

Applications of FISH in genomics, phylogenetics and plant breeding

FISH can support genome sequencing projects

Complex eukaryote genomes are sequenced by BAC-by-BAC or whole genome shotgun sequencing techniques. The first approach uses overlapping BACs, that are anchored on the genetic map by molecular markers, then each BAC is sequenced separately. With the second approach the whole genome is fragmented into huge number of small fragments at a particular high coverage of the genome, sequenced to obtain reads. In the next step multiple overlapping reads are obtained by doing several rounds of this fragmentation and sequencing, which are finally assembled by computer programs to bring the overlapping ends of different reads into a continuous sequence. So far, my FISH work focused on supporting the BAC-by-BAC strategy of the chromosomes 4, 6, 7, 9 and 12 of tomato. The principle of the technology is summarized in Figure 1. In most cases I mapped so called seed BACs that were anchored to the chromosomes by molecular markers with known position on the genetic map (Szinay et al. 2008). Later, I also mapped extension BACs where high repeat content of the clones raised doubt about their position on the chromosome.

As I explained briefly above, the essential basis of all sequencing techniques is to link sequence contigs to chromosomes. Molecular markers with known positions on the genetic map are helpful especially in small genomes with low repeat content, such as rice (Sequencing Project International Rice 2005), *A. thaliana* (The Arabidopsis Genome Initiative 2000), *M. truncatula* (Young et al. 2005) and with highly polymorphic genomes such as potato (Visser et al. 2009). However, there are also limitations by using genetic mapping approaches to link sequences to chromosomes, especially in those cases where genetic and physical maps show big discrepancies. This can happen in large heterochromatic blocks where crossover

recombination is suppressed and all markers in that region tightly linked. A second reason for such differences between genetic and physical maps can occur in homoeologous regions of introgression lines of tomato where severe suppression of recombination in the introgressed region resulted from lack of homology or chromosomal rearrangements (Liharska et al. 1996). In my thesis I will show and discuss several examples where FISH can be applied to show the physical chromosomal locations of BACs and/or whether BACs contain repeats that can potentially spread all over the genome. The latter is invisible by genetic mapping. In the sequencing of small genomes, like *A. thaliana*, FISH was not necessary due to its low number of repetitive sequences. In very large and complex genomes, like wheat, single copy BAC FISH is very difficult due to its high repeat content. Wheat has a large genome with low polymorphism and it was only recently published that using the sequencing of flow-sorted chromosomes (Doležel et al. 2007) it could be feasible to create a physical map and accelerate marker development (Paux et al. 2008).

In general, FISH is indispensable in genomics and plant genetics, and breeding as one of the most direct methods to show DNA sequences directly on chromosomes, in spite of their genetic mapping and repeat content. In the tomato sequencing project FISH proved to be a useful and fast method to anchor BAC sequences to chromosomes (Peters et al. 2009; Szinay et al. 2008). Furthermore FISH has three advantages compared to genetic mapping: 1) mapping populations and DNA polymorphisms are not required; 2) suppression of recombination hampering genetic mapping in the pericentromere does not influence the accuracy of FISH mapping and 3) pooled BAC FISH technique can be applied to reveal the coverage of the BACs and the size of the gaps between the different contigs (Szinay et al. 2008).

Detection of chromosomal rearrangements

Chromosomal rearrangements, like inversions, translocations, duplications and deletions are major genomic processes and have profound effects on the genetics and evolution of related species. Inversions occur as the result of one of two DNA breaks and inverted reunion inside a chromosome arm (paracentric) or the two arms of a chromosome (pericentric). Translocation occurs when the broken chromosome segment conjugates to another chromosome. These types of rearrangements are the most common ones in plants. Occurrence of reciprocal translocations is significantly higher than inversions with increasing chromosome number (Levin 2002). However, in *Solanum* species the inversions seem to prevail. Duplication happens when chromosomes or segments are multiplied. Deletion often follows duplications (mostly polyploidization) resulting in complete loss of some parts of the chromosome complement. Koo and Jiang (2008) described such an event in *Tripsacum dactyloides*, which is an ancient tetraploid which lost a chromosome segment during evolution.

Chromosomal rearrangements mostly take place during interphase, however telomeres and partly centromeres can stay in telophase even in the interphase nucleus. 5S rRNA clusters are able to keep the telophase structure, which could be explained by the short distance to

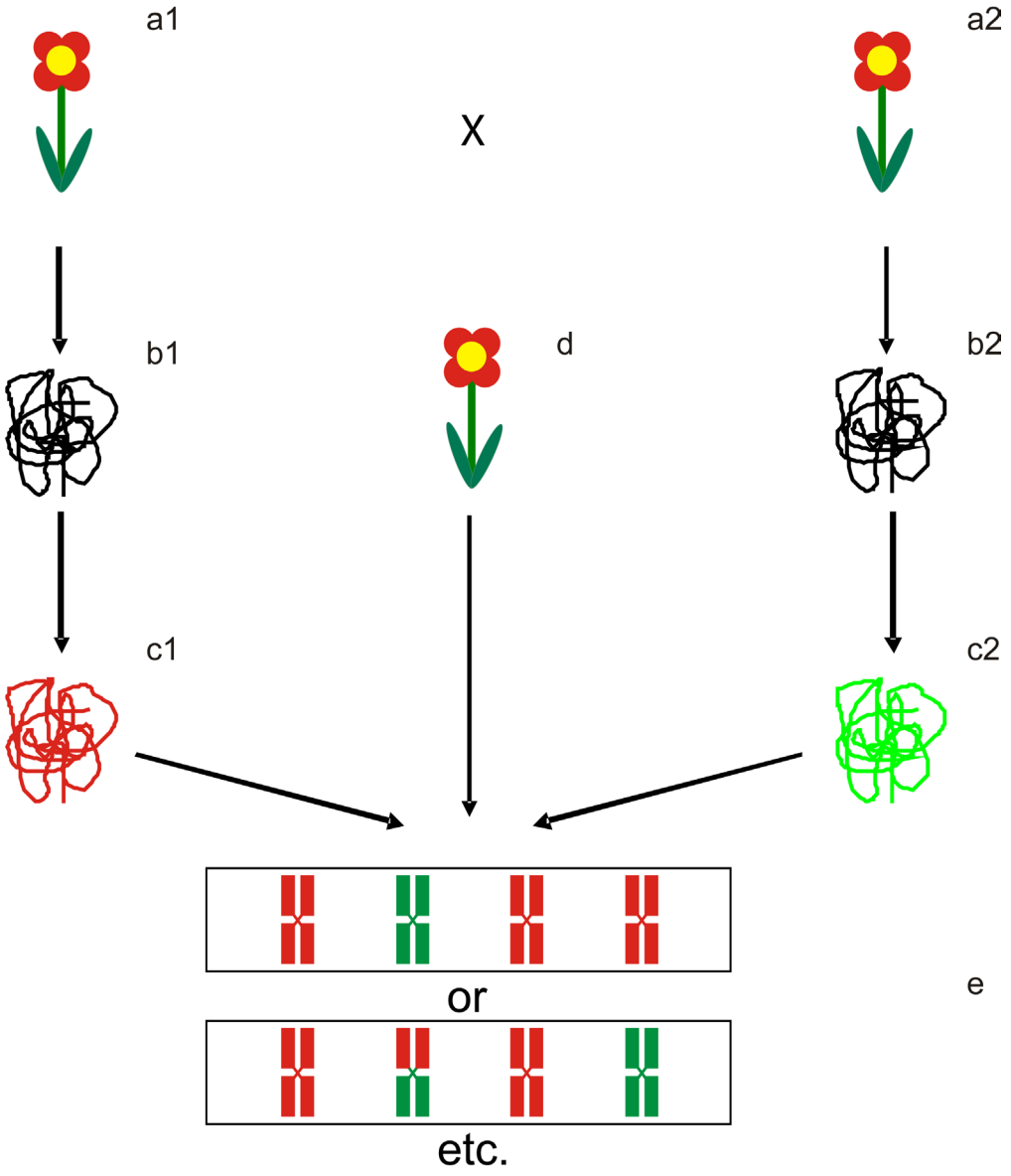


Figure 2. GISH procedure. a1,a2 = parental species; b1,b2 = isolated genomic DNA of the parents; c1,c2 = labeled DNA of the parents; d = species of interest; e = spread DNA of species in 'd' then hybridized with genomic DNA of the parents.

the centromere as shown on chromosome 2 of *Petunia hybrida* and *Crepis capillares* (Montijn et al. 1999). Genome painting or Genome *in situ* Hybridization (GISH), based on FISH with labeled genomic DNA from one parental species with unlabeled DNA from the other parental

species as blocking of common repetitive sequences, can identify homoeologous recombinations and translocations between the chromosomes of the parental genomes (Figure 2); however, they cannot demonstrate the occurrence of inversions as well as smaller duplications and deletions. The technique makes use of low sequence similarity between the dispersed repeats of the parental species. GISH analyses by Kosina and Heslop-Harrison (1996) showed chromosome number changes and structural differences among *Tripsacum durum*, *Thinopyrum distichum* and *Lophopyrum elongatum*. Intergenomic translocations were mostly seen in *T. durum* between its A and B genomes and less frequently translocations between *T. durum* and *T. distichum*. Jiang and Gill (1994) showed that intergenomic translocations can be species-specific in tetraploid wheat. Intergenomic translocations can involve one chromosome arm or often both arms as it is described in *Avena* species (Jellen et al. 1994). There are several other examples of translocation events observed by GISH in plants, such as raspberry and blackberry (Lim et al. 1998), *Allium* species (Friesen and Klaas 1998) and *Aegilops* species (Linc et al. 1999) using with combined FISH techniques (Coriton et al. 2009).

FISH with repeats and BACs containing single copy sequences offers different information and is highly complementary compared to genomic painting. Ren et al. (2009) used microsatellites as probes for FISH and observed a paracentric inversion on chromosome 5 between two *Cucumis sativus* varieties. In *Brassica*, A and C genomes have high sequence similarity, which requires a combined GISH and FISH approach to separate the two genomes with species specific repeats (Howell et al. 2008). There are not so many studies using single copy sequences to reveal chromosomal rearrangement, though this technique can show specific differences between species. For example, in the Brassicaceae family FISH with pooled chromosome specific BACs reveals several translocations between *A. thaliana* and chromosome arms of different *Brassica* species (Lysak et al. 2005).

Chromosomal rearrangements in Solanaceae

So far, genome-wide chromosomal rearrangements in Solanaceae have been demonstrated mostly on the basis of comparative genetic mapping. Examples of that are the studies referring to inversions between wild tomato species compared to *S. lycopersicum* (Pertuzé et al. 2002; Seah et al. 2004) and chromosome size and organization differences (Canady et al. 2006; Ji et al. 2004). Between tomato and its closest outgroup (S. sect. *Juglandifolia*) inversions and translocations were shown (Albrecht and Chetelat 2009). Between tomato and potato (*S. tuberosum*), only whole arm inversions were found on 5S (short arm), 9S, 10L (long arm), 11S and 12S (Bonierbale et al. 1988; Grube et al. 2000; Tanksley et al. 1992). The inversion on chromosome 10 demonstrates that *S. sitiens* and *S. lycopersicoides* are colinear with *S. tuberosum* (Pertuzé et al. 2002), suggesting that this inversion was fixed in the common ancestor of the tomato lineage. A paracentric inversion on chromosome 7 was mentioned on one of the accessions of *S. pennellii* compared to tomato (van der Knaap et al. 2004). The most detailed comparison of related *Solanum* species reveal inversions on *S. tuberosum* (chromo-

some 5, 6), *S. lycopersicoides* (chromosome 6, 7), *S. pennellii* (chromosome 6), *S. melongena* (chromosome 9) and some differences compared to genetic studies (Doganlar et al. 2002). Translocations as well as inversions were shown between eggplant (*S. melongena*) and pepper (*Capsicum annuum*) which are more distant relatives to tomato and potato (Doganlar et al. 2002; Livingstone et al. 1999). Further, (Doganlar et al. 2002) showed that eggplant and tomato genomes are differentiated by 28 rearrangements, which could be explained by 23 paracentric inversions and five translocations.

Comparative cytogenetic studies have demonstrated only very few rearrangements between the chromosomes of the *Solanum* species. The best classical study in this respect was from Gottschalk (1954) who compared pachytene chromosomes from a large number of tomato and potato species and other Solanacae genera. However, the details of the pachytene morphology was limited and later strongly improved by Ramanna and Prakken (1967) and Ramanna and Wageningvoort (1975) using aceto-carmine squash preparation of tomato and potato, respectively. Khush and Rick (1963) observed differences in chromosomal organization between tomato and related tomato species. In spread synaptonemal complexes of pachytene cells of somatic tomato (+) potato hybrids, de Jong et al. (1993) reported several chromosome pairing irregularities, indicating the occurrence of various inversions and translocations between tomato and potato. In a more recent SC analysis of interspecific hybrids between tomato and related species of the tomato clade, Anderson et al. (2010) demonstrated various pairing irregularities that could be interpreted as known paracentric inversions. In addition, they described several unmatched pericentromere regions between homoeologues, suggesting pericentric inversions, as well as few hitherto unknown translocations.

BAC FISH painting is also a very powerful method to detect chromosomal rearrangements between related species. In two recent studies of tomato and potato chromosome 6 Iovene et al. (2008) and Tang et al. (2008) demonstrated a previously unknown paracentric inversion on chromosome 6 between tomato and potato. Many more inversions on the basis of BAC FISH painting are described in Chapter 5 of my thesis.

Evolutionary aspects (phylogeny)

Genome evolution in plants takes place via four main levels, which are strongly correlated with each other. Plants can have multiple genomes (polyploidy), or can loose or produce individual chromosomes (aneuploidy). In those processes dynamics of repeated sequences play a role as well as chromosomal rearrangement. A chromosomal rearrangement causes problems in chromosome pairing that can lead to gamete sterility and may contribute to reproductive isolation, and consequently on speciation. In an extreme case speciation occurs with decreased genetic variability - known as founder effect - when few individuals separate and might create a new population. However new species can develop from interspecific hybrids by spontaneous and immediate chromosome doubling or by fixed viable recombinant events (Benavente et al. 2008).

Phylogenetic relationship can be studied by GISH where genomes are not complex and the parental genomes are distinguishable. In complex polyploid genomes a single data set is not enough to reveal a clear phylogenetic relationship between species. In plants polyploidy and introgression hybridization have a much larger evolutionary function than in mammals (Schmidt and Heslop-Harrison 1998). *Nicotiana* is a good example, where studies were reported about using ribosomal DNA to reveal the origin of the species. In a hybrid, GISH could reveal one of the parents (Kenton et al. 1993), but in amphidiploids due to similar parental genome sequences GISH is not capable to distinguish parental chromosomes (Chase et al. 2003; Clarkson et al. 2005). In maize the genomic DNA of all its ten wild relatives were used as probes for GISH to identify the ancestral genome donor of maize. The result was not satisfactory, as the probe DNA from all wild species hybridized all over the maize chromosomes equally except the DNA from the *Sorghum* species, that hybridized only to the NOR region of maize (Takahashi et al. 1999).

Centromeres are dynamical structures in the chromosomes and can be inactivated or replaced in plants as it was first reported in maize and rye. This phenomenon occurs mostly in smaller dicentric chromosomes (Han et al. 2006). Centromeres evolve faster than other parts of the genome and are mostly composed of satellite repeats, that can vary between closely related species, like in *Arabidopsis* (Lysak and Lexer 2006), but even between chromosomes within a genome, like in maize (Lamb 2009), *Brassica rapa* and *B. oleracea* (Lysak and Lexer 2006). Also epigenetic changes may contribute strongly to centromere identification. In eukaryotes a centromere is identified by a histone H3 variant, called CENH3, where CEN chromatin is hypomethylated compared with the same chromatin at the flanking region of the pericentromere. This phenomenon was found in *A. thaliana* as well as in maize (Zhang et al. 2008).

Plant breeding aspects

One of the major goals in plant breeding is broadening the genetic basis of crops. This is often done by longstanding introgressive hybridization in which a desired trait of the donor species is introduced into the recipient crop by recurrent backcrosses (Anamthawat-Jónsson 2001). Interspecific hybridization can take place in a sexual way, like triticale (wheat x rye) (Fernandez Calvin and Orellana 1994), beet x wild beet (Desel et al. 2002), *S. lycopersicum* x *S. pennellii* (Eshed and Zamir 1995), but for more distantly related species conventional plant breeding methods will fail due to incompatibility of the parental genomes. Therefore several *in vitro* technologies have been adapted for merging less related species by somatic hybridization through protoplast fusion and/or embryo rescue. Even though the hybrids are viable, meiotic abnormalities still can occur that can result fertility problems and/or suppression in recombination frequency (Levin 2002) between the alien and recipient homoeologues (Chang and de Jong 2005). Many factors influence the success of introgression, which include phylogenetic relationship, mating system, density and physical distribution of repeats as well as chromosomal rearrangements between the related species (Benavente et al. 2008).

Different FISH techniques, such as genome painting have been applied in plants to distinguish alien chromosomes in interspecific or intergeneric hybrids. In wheat and rye (Sepsi et al. 2008), C-banding and GISH techniques were compared and the results showed that C-banding was less efficient than GISH in detecting recombination frequency and introgressed segment size (Fernandez Calvin and Orellana 1994). By using *Alstroemeria* hybrids, chromosome association study was done successfully with GISH and FISH to identify parental chromosomes (Kamstra et al. 2004). *Lilium*, *Alstroemeria* and other lilies in the Liliaceae family belong to species with the largest genomes in the plant kingdom. Lim et al. (2001) successfully studied the genomes of *Alstroemeria* and *Lilium* by GISH. In another study between oriental and Asiatic lilies (Barba-Gonzalez et al. 2005) showed intergenomic recombination between the genomes, where the chromosome pairing was low. More studies have combined GISH and FISH analyses to obtain detailed information on the origin and composition of chromosomes in hybrids as well as the crossover events between the homoeologous chromosomes.

Combined technologies such as chromosome banding techniques (like C-banding, (Fernandez Calvin and Orellana 1994), FISH with repeats (like 45S and 5S rDNA, (Xu and Earle 1996) and chromosome pairing studies at meiotic prophase of the hybrids (de Jong et al. 1993) provide additional landmarks to identify and trace the behavior of individual chromosomes. Kopecky et al. (2008) studied *Festuca pratensis* and *Lolium multiflorum*, which are not very close relatives and have distinct repeat content as shown by GISH. Interestingly little differences were observed in pairing preferences between homologous and homoeologous chromosomes depending on the identity of the introgressed segment. In monosomic introgression lines the pairing was random, but in disomic introgression lines homologous pairing was slightly preferential.

In the Solanaceae family several examples were found in using somatic hybridization for plant breeding. Tomato and potato have 12 chromosomes, which are morphologically similar, but differ in repeat content (Ganal et al. 1988). Interspecific hybridization between tomato and potato is possible by producing somatic F1 hybrid and in vitro embryo rescue technology of next back cross generations. Although Solanaceous crops have rather small somatic chromosomes, they are always distinguishable from different parental species with GISH (Jacobsen et al. 1995). GISH analysis showed that the four tetraploid hybrids derived from the crosses between *S. lycopersicum* and *S. lycopersicoides* have equal numbers of chromosomes from their parental genomes, but the four hexaploid descendents carry two third of tomato and one third of the wild species chromosome sets (Escalante et al. 1998). In a study of Iovene et al. (2007), five *S. bulbocastanum* (+) *S. tuberosum* somatic hybrids were investigated by GISH to reveal genome constitution. In the two nearly tetraploid plants the parental chromosome proportion of the cultivated and wild parental genomes was equal. In the hexaploid hybrids the distribution of the parental genomes differed by parental genome dosage (4:2= cultivated : wild or 2:4= cultivated : wild). Within the genus *Capsicum*, karyotypic differences have been observed. Though *C. annuum*, *C. frutescens* and *C. chinense* are crossable, but their chromosomal constitution is structurally different (Moscone et al. 2007).

Scope of this thesis

The aim of this research was to apply FISH technology in tomato for support of the ongoing genome sequencing project, genetic studies, plant breeding and phylogeny, in which selected tomato/potato BACs were used to demonstrating chromosomal evolutionary processes and chromosomal rearrangements within Solanaceous crops. Firstly, we used novel technologies, such as 5-colour cross-species FISH, Cot 100 painting and pooled-BAC FISH. Secondly, we applied them to verify physical locations of BACs that were selected in the tomato sequencing project. Then, with 5-colour cross-species FISH method, we discovered an inversion between tomato and potato that was not detected genetically. Finally, we studied the overall physical organisation of tomato/potato BACs across a large number of *Solanum* and *Capsicum* species by using the available tomato and potato BACs and the newly developed cross-species multcolour FISH technique. We attempted to set up a BAC-synteny map to study the chromosomal collinearity and the evolution of chromosome complements among genomes of tomato, potato, pepper and eggplant. Also, we tried to analyse and interpret the FISH data in context with phylogenetic patterns and breeding aspects. The latter will provide a basis for a breeding strategy to introgress genes from wild species.

Chapter 2

In Chapter 2, the strategy of cytogenetic analysis of tomato chromosome 6 in the framework of the International Solanaceae Genome Project is described. In this study we developed 5-colour and pooled BAC FISH techniques and applied them to confirm BAC positions provided by the genetic map. Significant percentage of these BACs was discrepant with the linkage map mostly in the pericentromere.

Chapter 3

In Chapter 3, the repeat distribution along chromosome 7 of tomato in combination with bioinformatics and cytogenetic data is described. All tomato genome repeats were investigated according to their distribution, function, conformation and possible evolution. The results of the sequencing and cytogenetic mapping were mutually complementary and highly significant.

Chapter 4

In Chapter 4, the first application of multicolour cross-species FISH technique in the *Solanaceae* family was presented and applied to describe a novel chromosomal inversion between tomato and potato on the short arm of chromosome 6.

Chapter 5

In Chapter 5, large scale cross species FISH study is described using representatives of the tomato clade, *S. tuberosum*, *S. melongena*, *C. annuum* and tomato breeding lines. The focus of

the research was to show phylogenetic relationships between species and to reveal chromosomal rearrangement by FISH. Several novel chromosomal rearrangement and some differences compared to previously described results in literature were observed.

Supplementary table

Overview of cytogenetic techniques that I use in my thesis, including DNA labeling techniques with dyes and fluorescence spectra and microscopy. The first table contains descriptions about the main cytogenetic techniques. The second table shows various DNA labeling techniques. The third table contains information about fluorescence spectrum including wavelength of several dyes as possibilities. The fourth table describes different kind of microscopes.

Cytogenetic techniques	Description	Sensitivity
Cross-species colour segmentation	FISH based banding technology	
	Flow sorted chromosomes hy-bridize to related region (gibbon, human)	
Comparative genomic hybridization (CGH)	DNA probe extracted from a test and a normal samples = differentially labeled	For single cel study
	Copy number differences can be distinguished	for microdissected smaller DNA study
	Limitations: does not provide information about ploidy, neither about structure	
Interphase cytogenetics	Euchromatin is decondensed; heterochromatin condensed as at meta-phase	probe order within 50 kb - 1 MB
	Chromosome organization is distinguishable	
Replication (R) banding	Differentiate between early and late replication regions	
	combined with Q-PCR and counting FISH in S phase = measure replication time	
Fiber FISH	Naked DNA, free from proteins. Detect gaps, overlapping contigs, copy number differences	1-500 kb; extended DNA fiber: 2.96 kb / µm
	Alkaline lysis or high-salt in SDS: different fibre length, not suitable for quantification	
	Molecular combing: high molecular weight DNA stretched, removing the cover slip at constant rate	
	Parallel, even DNA = suitable for quantification	

Tyramid FISH	Tyramide conjugated to a fluorophore - enzymatically	684-900 bp detectable
	Short detection time, highly sensitive (10-100 times more sensitive than conventional FISH)	In maize: semisensitive, 2.4 kb, resolution 3.3-8.2 Mb
Super-stretched chromosomes for FISH	up to 100 times longer than native metaphase chromosomes and 20 times longer than pachytene chromosomes	1 kb Spatial resolution of meighboring loci is less than 70 kb; in metaphase it is 5-10 MB
Multiplex-FISH (M-FISH)	based on combinaorial labelling	
	used for detecting chromosomal rearrangements	
	each chromosome has unique label	

Labeling technique	**Description**
PCR approach	Chomosome / chromosome regions are amplified and labeled in the same time
DOP-PCR (degenerate oligonucleotide primed PCR)	Flow sorted or microdissected chromosomes are labeled, used for reversed chromosome painting as well
	Reveal content and breakpoint of chromosome rearrangements
Combinatorial labeling	Probes are indicated by their absence or presence => calculation of the Boolean spectral
Ratio labeling	Fluorochromes are combined in probes by different proportion
Padlock probe	2 complementary segment covalently linked on the hybridization in a circular probe
	Resolution is on bp level
Rolling circle amplification	DNA amplified by DNA polymerase under isothermal conditions by linear or geometrical kinetics
	High throughput approach
Nick translation	DNAse I attach to the dsDNA, DNA polymerase act as endonuclease and polymerase. It incorporate with labeled nucleotides and DNA
Immunostaining	Antibody based for protein detection
End labeling	Template independent reaction, uses enzyme terminal deoxynucleotidyl transferase and DNA Polymerase that support the attachment of the nucleotides to the 3'-OH end of the DNA
End labeling (cont.)	Suitable for single and double strand DNA
Random prime labeling	Single strand DNA is amplified by Klenow fragment of DNA polymerase I with random oligonucleotides
	Suitable for small DNA fragments or single strand DNA

Fluorescence spectrum	Fluorescence dye	Wavelength
UV/Blue	DAPI	350
	AMCA	353
	CB	396
	DEAC	432
Blue/Green	FITC	491
	OG-488	495
	A-488	493
	RGr	515
Green/Yellow-orange	R6G	524
	Cy3	550
	TAMRA	547
Orange/Red	Texas-Red	583
	Cy3.5	581
Red/Far-red	Cy5	649
Infrared/Far red	Cy5.5	645
	Cy7	743
Hapten	Biotin	-
	Digoxigenin	-
	Dinitrophenol	-

Microscope	Description
Bright Field	Illumination is via transmitted white light
	Low contrast and resolution (200-300 nm on the xy axis, 500-800 nm on the z axis)
	Smallest distance between signals: 0.2 µm
Phase contrast	Difference in contrast and refractive index
	Not suitable for thick samples
	Circular annulus produce a cone of light in the condenser
	Objectives have different size ring
Differential interference contrast	Image created by the interference with the detected light
	Condensed aperture can be kept open, which decrease the depth of field and the resolution
	Special prism splits the light, then collect in the specimen
	In a homogeneous sample contrast is not generated
Fluorescence	Fluorescence: illuminate with high energy light and emit light in lower frequency
	Very sensitive: single molecule is detectable
	Use fluorescence dyes which can show structure of the sample
	Use epi-illumination to decrease the excitation light to the detector
Confocal	Use scanning point of light
	Blocks the out of focus light, that improves the optical sectioning and the resolution
	Used in 3D structure studies
Stimulated emission depletion (STED)	Based on confocal microscopy
	Use 2 laser pulses: second pulse is the depletion pulse, which goes through the phase modulator and the centre of the sample remains fluorescence. With saturation the resolution gets tens of nanometers

Chapter 2

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
5 Centre for Biosystems Genomics, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

Appeared in The Plant Journal (2008), doi: 10.1111/j.1365-313X.2008.03626.x

Abstract

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.

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

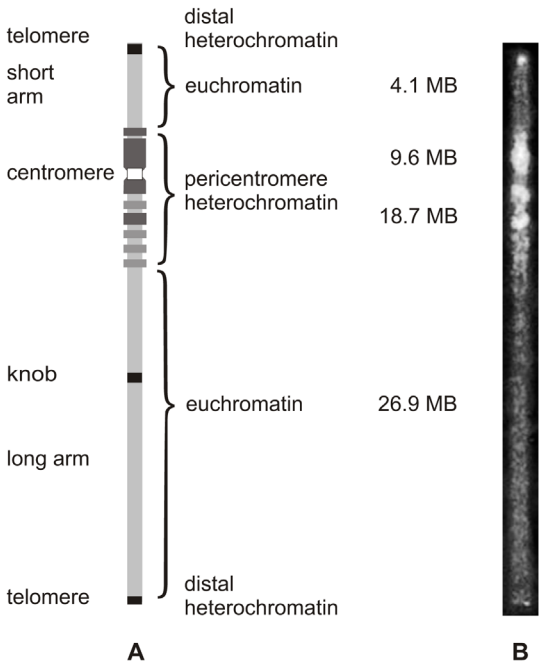


Figure 1. Genomic and cytogenetic characteristics of chromosome 6 of tomato (*Solanum lycopersicum*). (A) Schematic representation of the pachytene chromosome with estimations of DNA size in the euchromatin and heterochromatin regions. (B) Straightened pachytene bivalent stained with DAPI.

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 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

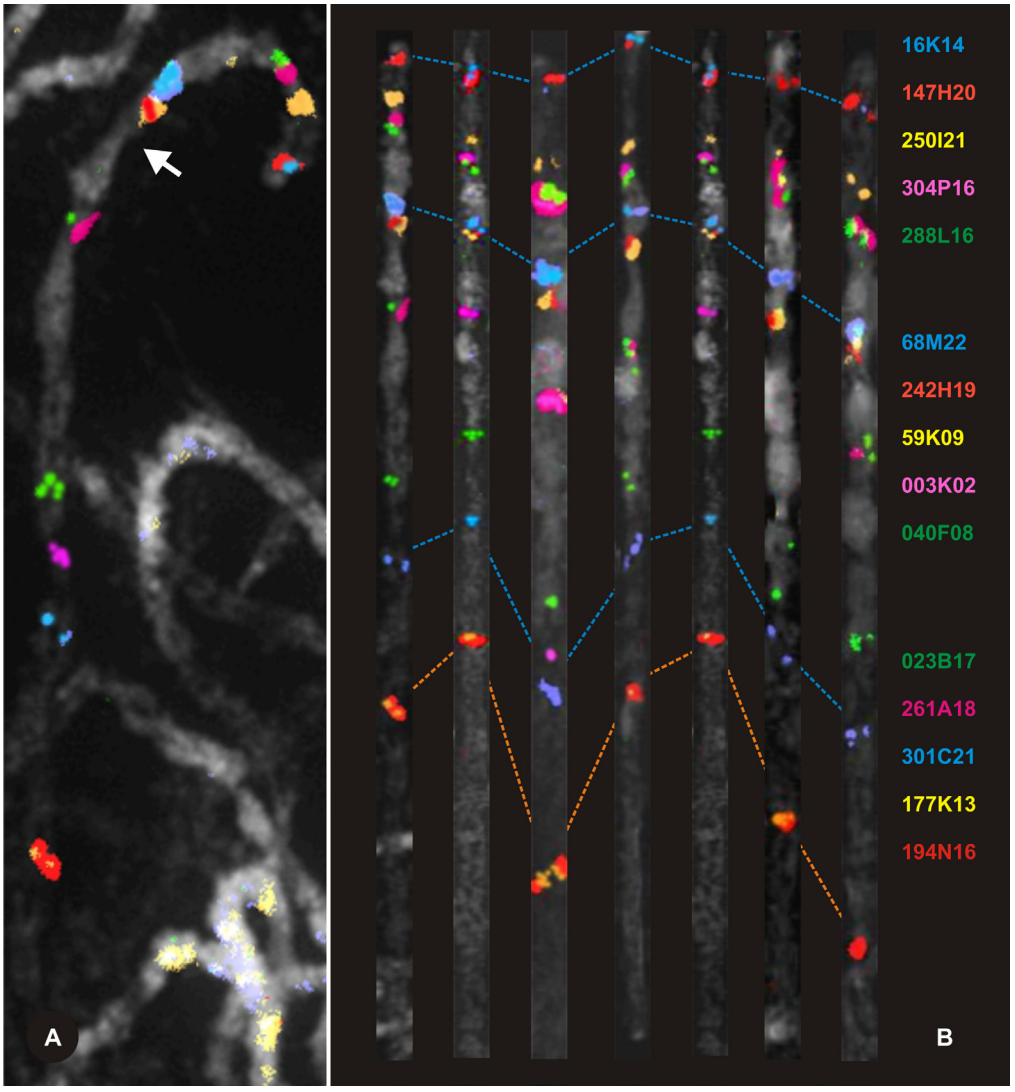


Figure 2. Multi-colour FISH using three sets of five seed-BACs. (A) Detail of chromosome pair 6 in a pachytene complement. The arrow indicates the position of the centromere. (B) Seven examples of straightened bivalents, with the colour scheme of the fifteen different BACs.

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.

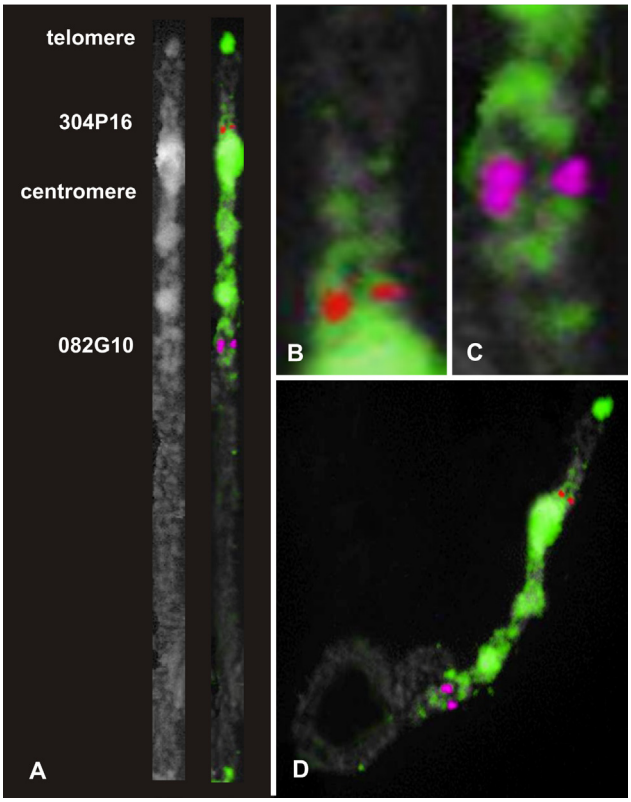


Figure 3. Labeling of the heterochromatic part of tomato chromosome 6 by FISH with the Cot-100 genomic DNA fraction (green signal). The differently labelled BAC clones resident in the heterochromatin / euchromatin borders of the short arm and of the long arm are pseudo-coloured in red and magenta respectively. The images (B) and (C) are detailed magnification of (A) and (D).

Previous studies on the composition of the pericentromere in tomato have revealed high amounts of various repetitive elements, including the TGRII and TGRIII 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.

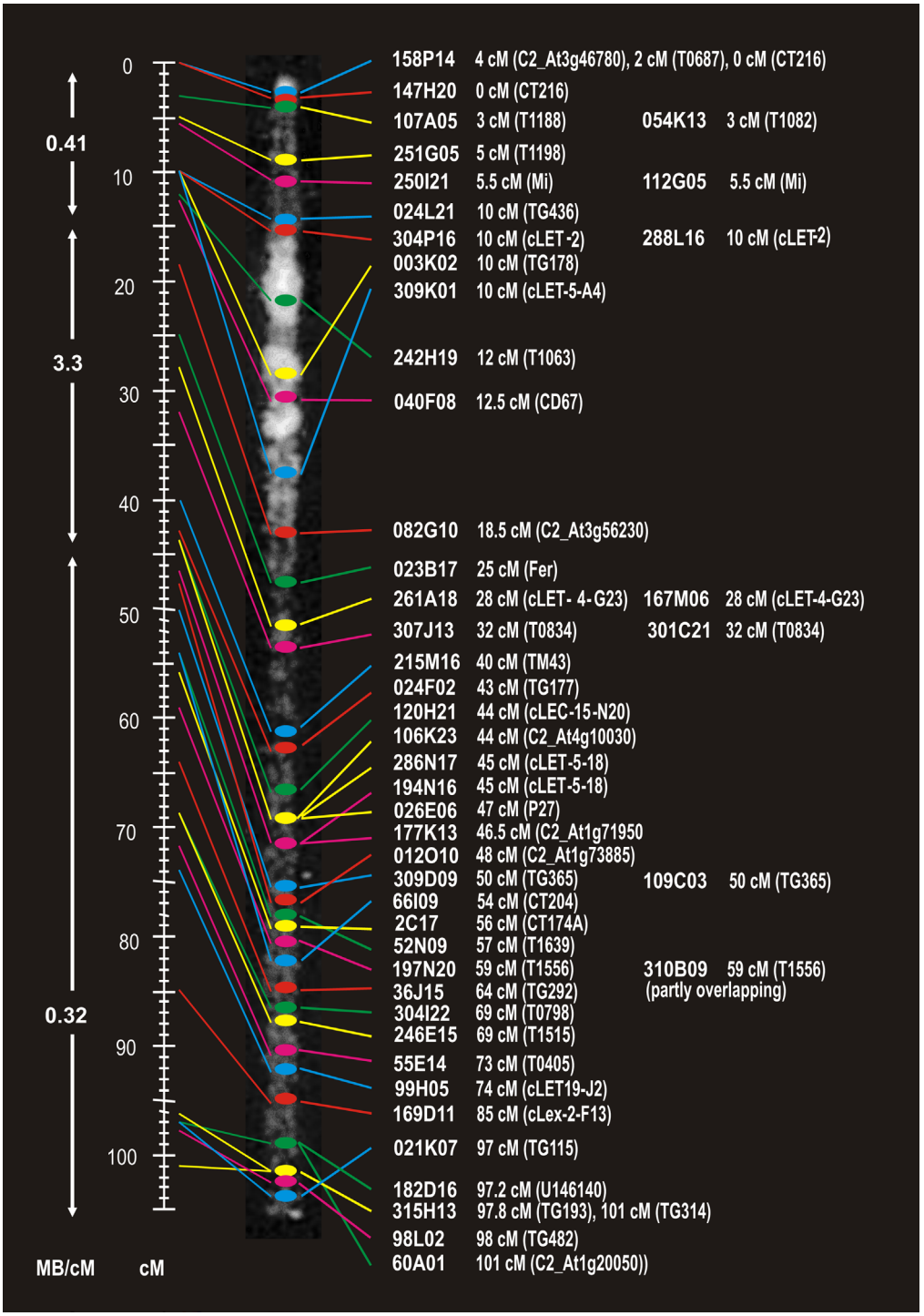
Confirmation and physical mapping of seed BACs

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 for heterochromatin and 0.6 Mb/µm for euchromatin (Budiman et al. 2004). From these measurements, it was estimated that recombination frequency is reduced approximately nine fold 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 crossovers in the pericentromere.

Determination of chromosomal coverage by pooled-BAC FISH

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 in the distal ends and in the 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. → → → →


0.41
3.3
0.32
MB/cM
cM
0
10
20
30
40
50
60
70
80
90
100
158P14 4 cM (C2_At3g46780), 2 cM (T0687), 0 cM (CT216)
147H20 0 cM (CT216)
107A05 3 cM (T1188)
054K13 3 cM (T1082)
251G05 5 cM (T1198)
250I21 5.5 cM (Mi)
112G05 5.5 cM (Mi)
024L21 10 cM (TG436)
304P16 10 cM (cLET-2)
288L16 10 cM (cLET-2)
003K02 10 cM (TG178)
309K01 10 cM (cLET-5-A4)
242H19 12 cM (T1063)
040F08 12.5 cM (CD67)
082G10 18.5 cM (C2_At3g56230)
023B17 25 cM (Fer)
261A18 28 cM (cLET- 4-G23)
167M06 28 cM (cLET-4-G23)
307J13 32 cM (T0834)
301C21 32 cM (T0834)
215M16 40 cM (TM43)
024F02 43 cM (TG177)
120H21 44 cM (cLEC-15-N20)
106K23 44 cM (C2_At4g10030)
286N17 45 cM (cLET-5-18)
194N16 45 cM (cLET-5-18)
026E06 47 cM (P27)
177K13 46.5 cM (C2_At1g71950
012O10 48 cM (C2_At1g73885)
309D09 50 cM (TG365)
109C03 50 cM (TG365)
66I09 54 cM (CT204)
2C17 56 cM (CT174A)
52N09 57 cM (T1639)
197N20 59 cM (T1556)
310B09 59 cM (T1556)
36J15 64 cM (TG292)
(partly overlapping)
304I22 69 cM (T0798)
246E15 69 cM (T1515)
55E14 73 cM (T0405)
99H05 74 cM (cLET19-J2)
169D11 85 cM (cLex-2-F13)
021K07 97 cM (TG115)
182D16 97.2 cM (U146140)
315H13 97.8 cM (TG193), 101 cM (TG314)
98L02 98 cM (TG482)
60A01 101 cM (C2_At1g20050))

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 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

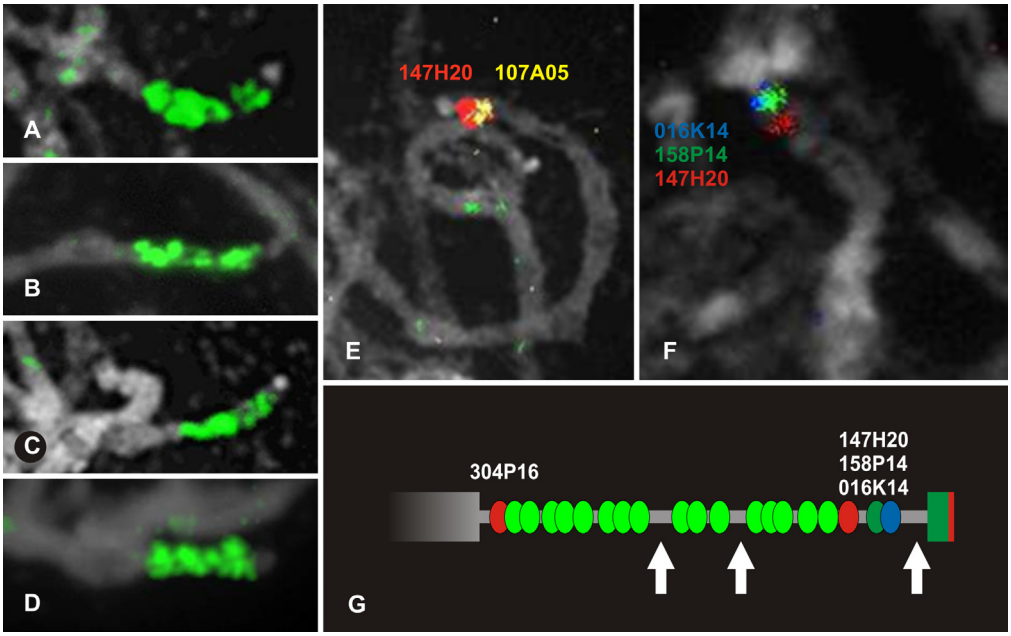


Figure 5. Pooled BAC FISH for the short arm of chromosome 6. (A)-(D): Compilation of four different pachytene spreads after hybridization with 18 seed BACs. (E, F): Multi-colour FISH showing proximal and terminal BACs for the euchromatin part of the short arm. (G) Schematic representation showing the gaps (arrows) between the seed BACs and the positions of proximal and terminal BACs.

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

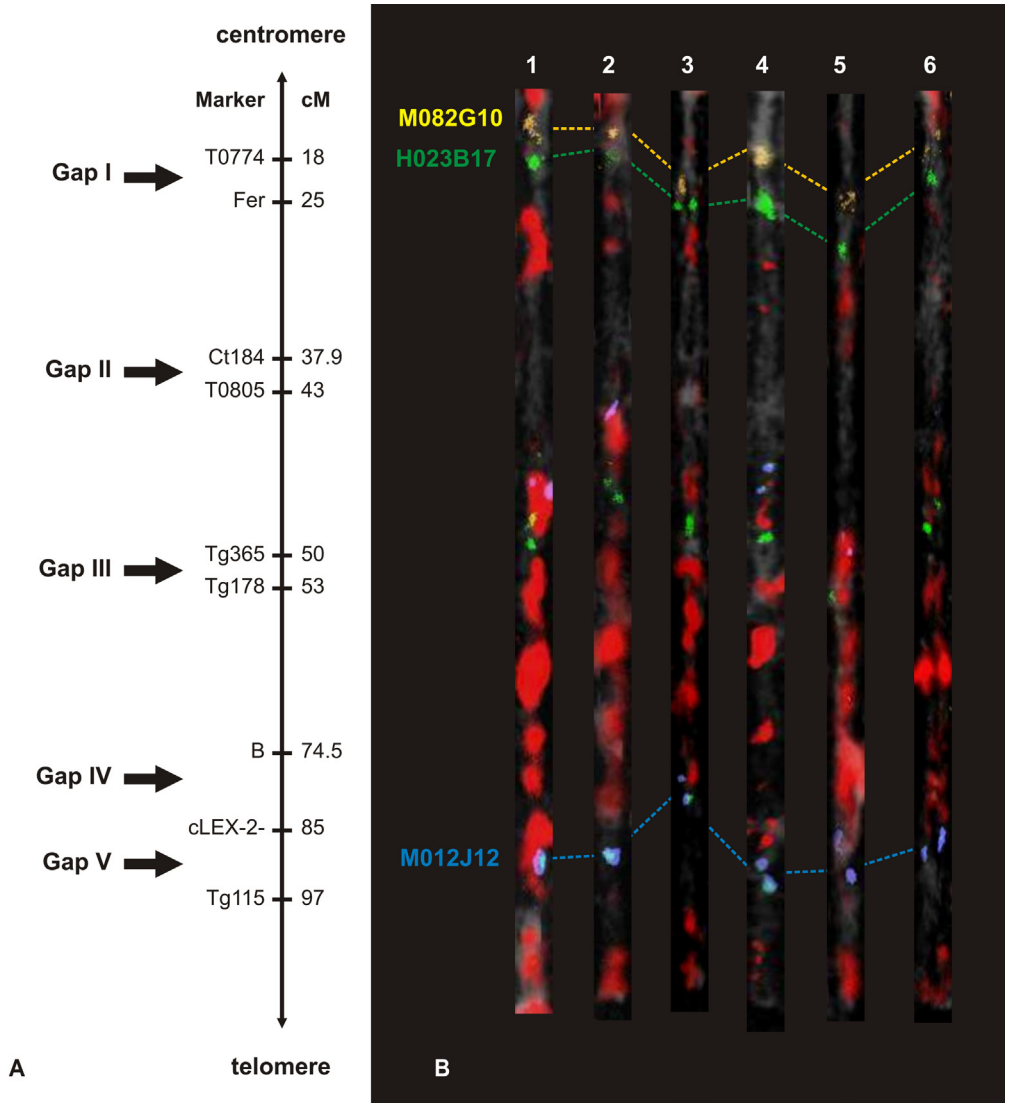


Figure 6. Pooled BAC-FISH of the long arm. (A) Genetic map of tomato chromosome 6 showing gaps lacking genetically anchored seed BACs. (B) Multi-colour pooled-BAC FISH. The red fluorescence signals are the pooled seed BACs of the long arm of chromosome 6. BACs targeted specifically towards gaps in the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

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 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(‘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 approx. 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 misallocated 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 misallocations 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 near the centromere of tomato chromosome 9, but higher ratios of 21.74 Mb/cM and 100 Mb/cM 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 for the euchromatin regions of chromosome 2L, and this ratio is less than half the mean ratio of 750 kb/cM for the tomato genome (Tanksley et al. 1992). Our own data showed Mb/cM 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 crossovers 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 of this fragmented DNA in 0.3 M NaCl at 95 °C for 10 min, and then let it reanneal 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 for 90 min at 37°C. The reaction was stopped and DNA was extracted by adding 300 µl chloroform:isoamylalcohol (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 HindIII BAC library (Budiman et al. 2000) and the Heinz 1706 MboI 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. 1996a). 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 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 x 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 ImageJ (http://rsb.info.nih.gov/ij: Kocsis et al. 1991), and final image optimization was performed using Photoshop CS4 (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.

Chapter 3

Dynamics and characterisation of the two major repeat families in tomato (*Solanum lycopersicum*)

Dóra Szinay[1,3,4], Murielle Philippot[2,4], Simo Zouine[2], Pierre Frasse[2], Mondher Bouzayen[2], Hans de Jong[3,5]

1. Wageningen University and Research Centre (WUR), Laboratory of Plant Breeding, Wageningen, Netherlands
2. Laboratoire de Génomique et Biotechnologie des Fruits, UMR 990 INRA/INP-ENSAT, Chemin de Borderouge BP107, 31326 Castanet Tolosan Cedex, France
3. WUR, Laboratory of Genetics
4. These authors contributed equally to this study
5. Corresponding author: WUR, Laboratory of Genetics, P.O. Box 309, 6700 AH Wageningen, The Netherlands. E-mail: hans.dejong@wur.nl

Abstract

Tomato (*Solanum lycopersicum*, 2n=24) has a genome size of 980 MB and consists of about 70 % of repetitive sequences, of which most reside in the nucleolar organiser region, the large distal heterochromatin blocks at most chromosome arms and in all twelve pericentromeres. These repeatrich chromosomal areas are difficult to sequence and hence become a major challenge in the genome sequencing of the species. Here we present chromosomal repeat mapping by Fluorescence *in situ* Hybridisation (FISH) on pachytene chromosome 7 and extended DNA fibres in combination with the sequencing and bioinformatics analysis of 168 BACs of that chromosome. Long tandem arrays of the tomato genomic repeat TGRI were found in short arm distal BACs as well as in BACs that were mapped in small interstitial heterochromatic island in the long arm euchromatin. The phylogenetic tree of their sequences by Neighbour-Joining analysis revealed clustering of the distal and interstitial TGRI blocks suggesting that they may have originated from blocks of other chromosomes. The remaining Tomato Genomic Repeats in this study: TGRII, TGRIII and TGRIV are all members of the Ty3-*Gypsy* LTR class of retrotransposons, and were mapped by FISH on the pericentromere of the chromosome, each with different coverage in the sense that TGRII paints the whole pericentromere, TGRIII overlaps with TGRII except for the long arm distal regions, whereas TGRIV was found only in the most proximal parts of the heterochromatin. The sequencing of the BACs were in agreement with the FISH data but also revealed several solo LTRs of TGRII in the euchromatin regions. The pericentromere contained both complete TGRII and TGRIII, truncated elements and solo LTRs of the repeats. TGRIV could not be further sequenced as the repeat occurred in the most proximal part of the pericentromere close to the centromere that is too repeat-rich for sequencing. The transition of euchromatin - heterochromatin borders in the short arm was sharp and clearly defined and corresponded well with the FISH signals of TGRII and the BACs containing complete or truncated Ty3-*Gypsy* elements. The border of the long arm showed a far more gradual transition of heterochromatin and presence of repeats. The two heterochromatin blocks in tomato can now be distinguished into two entirely different heterochromatin types, each with their unique chromosomal distribution, molecular organisation and characteristics. The former type contains the distal TGRI tandem arrays and are supposed to be very dynamic chromosomal domains, that frequently transpose by extrachromosomal circular DNA in trans to other distal and/or interstitial chromosome arm positions during stages that chromosomes occupy their Rabl orientation. This process is a plausible explanation for the known equilocal distribution of distal and interstitial heterochromatin blocks and may also account for the observed genetic instability in these regions. The latter type of heterochromatin is populated by the Ty3-*Gypsy* LTR retrotransposons, in which the three members of this repeat class have unequal distribution towards the centromere. We also discuss the possible selective forces that are responsible for their distribution in the pericentromere region.

Introduction

Tomato is one of the most important vegetable crops worldwide and together with potato the first species of the Solanaceae family with moderate genome size that are being sequenced. Both species now become reference genomes for all other Solanaceae crops and wild relatives, and thus provide invaluable resources for various fundamental questions on gene expression, genome organisation and quality trait improvements (Mueller et al. 2005a; 2005b; http://solgenomics.net/; Visser et al. 2009, and references therein). With a genome size of about 980 MB, tomato was initially estimated to contain about 10-22 % of repetitive sequences (Zamir and Tanksley, 1988; Ganal et al. 1988) that are located in the heterochromatin regions of the chromosomes. The percentage of single copy sequences corresponds roughly with the 30 % of BAC clones that are repeat rich as they show strong hybridisation with Cot 100 DNA (Chang et al. 2008). However, Peterson (1996, 1998) and Van der Hoeven et al. (2002) estimated that 75 % of the nuclear DNA is noncoding, leaving about 50 % of repeats of the tomato genome undetectable by FISH or Southern hybridisation, and may possibly be part of uncondensed euchromatin.

The distinction between euchromatin and heterochromatin as domains of repeat poor and repeat rich chromatin is a matter of intense debates (Bennetzen 2000; Yasuhara et al. 2006; Yasuhara and Wakimoto 2008), and the difference between both types is in tomato not as obvious as it is in the well-differentiated *Arabidopsis thaliana* chromosomes (Fransz et al. 1998). In a comparative study on FISH analysis and sequencing data, Peters et al. (2009) revealed that the short arm euchromatin of tomato chromosome 6 has a repeat content of 13.4 %, whereas the pericentromere displays an unexpectedly high gene density of one gene per 36.7 kb. In addition, carmine-stained pachytene chromosomes displayed numerous chromomeres in the euchromatin (Ramanna and Prakken 1967), that were also visualised in contrast enhanced DAPI stained pachytene complements (Szinay et al. 2010). The presence of these repetitive sequences are a major challenge in the genome sequencing initiatives as long stretches of repeats produces gaps in the physical map and hence hampers building supercontigs of the euchromatin areas. In addition, most of the repeats are in the large distal and pericentromere heterochromatin blocks that are devoid of crossover recombination resulting in clustering of molecular markers in their linkage maps.

The first study of tomato repeats was the analysis of a 452 bp *Hind*III DNA motif, THG2 (Zabel et al. 1985), which was characterised as a member of a genomic dispersed repeat family present in all pericentromeres without the nucleolar organiser region of chromosome 2 (Zhong et al. 1996b). Other major Tomato Genomic Repeats are the TGRI, in the distal heterochromatin blocks and small interstitial long arm knobs, and the TGRII and TGRIII in the pericentromeres of all chromosomes (Schweizer et al. 1988: Ganal et al. 1988, 1991, Lapitan et al. 1989, 1991). A more detailed chromosomal mapping and molecular size estimation of the TGRI repeats were carried out on pachytene and extended DNA fibre FISH showing an overview of molecular sizes for these tandem repeats on all chromosomes (Zhong et al. 1996b, 1998).

TGRII and TGRIII were mapped on all twelve pericentromere heterochromatin areas (Ganal et al. 1988; Chang et al. 2008), together with the recently discovered centromere-specific TGRIV, as well as various microsatellites and Ty1-*Copia* (Yang et al. 2005; Chang et al. 2008). The plant telomere TTTAGGG repeat could be mapped on the distal ends of all chromosomes (Lapitan et al. 1989; 1991; Zhong et al. 1996b), although plaque hybridisation of the TTTAGGG probe on lambda clones of tomato also showed telomere-homologous sequences on eight of the twelve centromere regions (Presting et al. 1996). These telomere sequences have never been detected in our FISH experiment (unpublished observations). Ribosomal genes coding for the 5S rDNA were mapped on chromosome 1 (Lapitan et al. 1991), whereas the nucleolar organiser region of chromosome 2 is the domain of the 45S rDNA (Chang et al. 2008, and references therein), with their associated microsatellites (Chang et al. 2008) and tandem repeats (Jo et al. 2009).

The first publications about physical mapping of tomato BACs dealt with the FISH positioning of repeat poor BACs on chromosome 1 (Chang et al. 2007), chromosome 2 (Koo et al. 2008) and chromosome 6 (Szinay et al. 2008), and focused on the integration of genetic, cytogenetic and physical maps of single copy sequences in the euchromatin arms. More attention for the occurrence of repeats in these BACs came from Peters et al. (2009) showing a surprisingly high content of repeats in the short arm euchromatin and the conspicuous difference between Ty3-*Gypsy* / Ty1-*Copia* ratio between euchromatin and heterochromatin. Datema et al. (2008) made a comparative genome-wide analysis of tomato and potato using BAC end sequences based on representative samples of 19 % of the tomato and 10 % of the potato genome, respectively. The 17 % larger genome size of tomato reflect the higher repeat content that are explained by a higher number of retrotransposon insertions of tomato compared to potato. However, simple sequences are more abundant in potato whereas both species also differ in their composition of microsatellite motifs. Also on the level of DAPI stained pachytene chromosomes, few albeit noticeable differences between tomato and potato can be detected, in which tomato chromosomes have larger and denser pericentromere regions and less euchromatin chromomeres than potato (Tang et al. 2008; Iovene et al. 2008; unpublished observations).

Here we present a comparative study on the molecular organisation of the major tomato repeats TGRI, TGRII, TGRIII and TGRIV based on a comprehensive analysis of 168 BACs of chromosome 7 of tomato. We compare chromosomal positions of these repeats on pachytene chromosomes using diagnostic BACs for unequivocal identification of the chromosome and present the first tentative results of repeat painting on extended DNA fibres. All BACs were sequenced and analysed for the presence and organisation of the repeats. Our primary interest was the distribution and organisation of repeats at the borders of euchromatin and heterochromatin. We also compared the sequences of the TGRI motifs in the BACs from the distal short arm heterochromatin block and the long arm interstitial knobs, and compared the structures of the different Ty3-*Gypsy* TGRII, TGRIII and TGRIV retrotransposons in the pericentromere heterochromatin.

Material and methods

Plant materials

Heinz 1706 plants were grown under standard greenhouse conditions. Young flower buds were collected and their anthers fixed in freshly prepared Carnoy solution (acetic acid : ethanol = 1:3), and the next day transferred to 70 % ethanol for further storage at 4 °C. Anther selection and slide preparation of spread pachytene complements were described in Szinay et al. (2008). Extended DNA fibres were prepared according to Zhong et al. (1998) with the exception of the way how we spread the dissolved nuclear material. Here we used the suggestion from the Plant Cytogenetics Lab of Madison University (Prof. Jiming Jiang, personal communication) to spread the drop by toughing the drop with the nuclear mix using the short edge of a 24x32 mm cover slip and moving the coverslip over de glass surface of the slide without touching the slide itself.

Repeats, Cot 100 and BAC DNA isolation and labelling

The repeats that we used for our experiments were Cot 100 (Chang et al. 2008), the tomato genomic repeats TGRI (Schweizer et al. 1988), TGRII (Vosman and Arens, 1997; Yang et al. 2005; Wang et al. 2006), TGRIII (Ganal et al. 1988) and TGRIV (Chang et al. 2008). In addition we selected 60 repeat-poor BACs, aiming to detect borders of heterochromatin and euchromatin.

Genomic DNA was isolated by the CTAB method according to Szinay et al. (2008). Cot 100 isolation was carried out as described by Zwick et al. (1997), with the exception that the phenol step during DNA extraction was skipped. The four repeats and the BACs were isolated by High Pure Plasmid Isolation Kit (Roche 11754785001) and then labelled by the Nick Translation method following the instructions of the manufacturer (http://www.roche.com). The repeats were labelled with digoxigenin and amplified with anti-digoxigenin-FITC and anti-sheep-FITC. BACs were labelled by biotin and amplified three times with Streptavidine-Cy5 and biotinilated-anti-streptavidin (for fibre FISH; biotin labelled probes were amplified with Avidine Texas Red and biotinilated-anti-avidine) and direct labelled Cy3-dUTP (Amersham, http://www5.amershambiosciences.com), Cy3.5-dCTP (Amersham) and Diethylaminocoumarin-5-dUTP, DEAC (Perkin Elmer, http://www.perkinelmer.com).

The FISH procedure followed the description by Szinay et al. (2008), but in a few experiments we decreased hybridisation / washing stringency from 82 % to 64 % (Schwarzacher and Heslop-Harrison, 2000) to get a stronger fluorescent signal. Microscopy and image processing were performed according to Szinay et al. (2008) with the following additional modifications: all repeats were labelled with FITC, but the grayscale images of the fluorescence signals were pseudo-coloured in different ways using the multichannel mode in Photoshop CS4 (Adobe Inc., http://www.adobe.com).

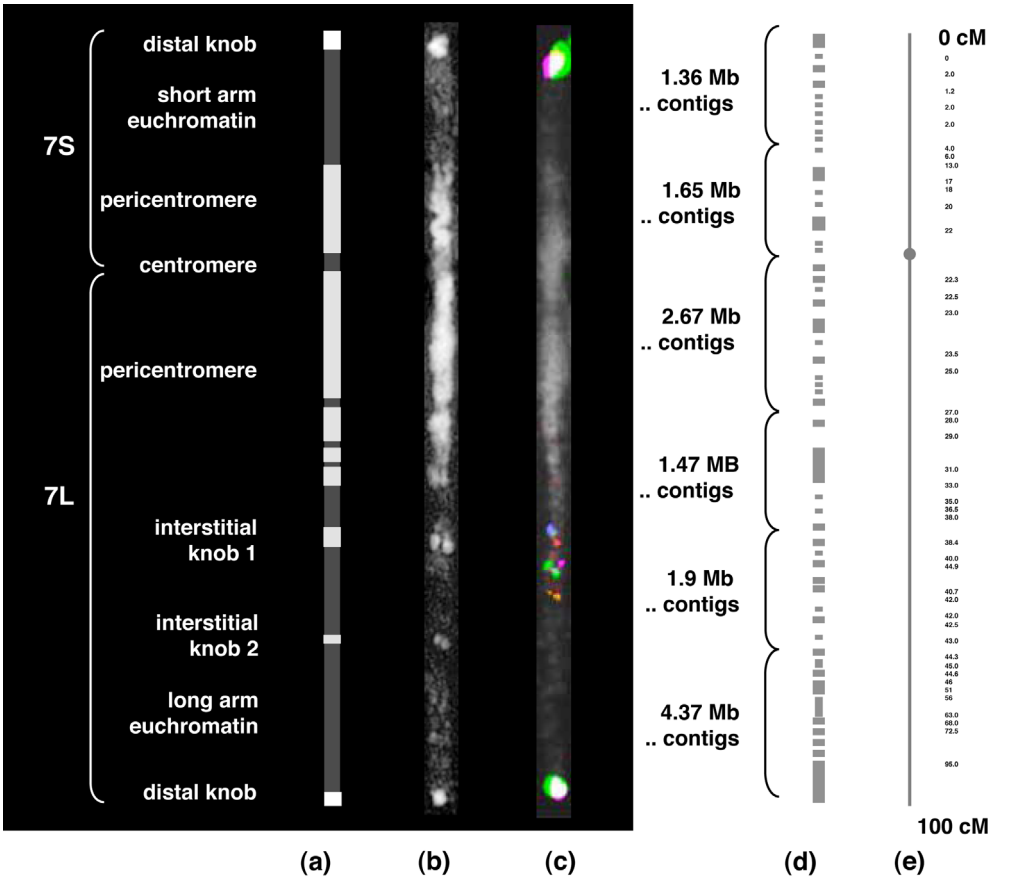


Figure 1. Overview of chromosome 7. a. Ideogram. b. DAPI fluorescence of a straightened pachytene chromosome. c Example of a BAC-FISH painting around the interstitial knob. d. Physical/Cytogenetic map showing the rough positions of the cotigs and singletons. e. Length indications for the genetic markers.

Phylogenetic analysis

The evolutionary history of related repeat sequences was inferred using the Neighbor-Joining method (Saitou and Nei, 1987). Evolutionary distances were computed using the Kimura 2-parameter method (Kimura 1980) and are in the units of the number of base substitutions per site. All positions containing gaps and missing data were eliminated from the dataset (Complete deletion option). Phylogenetic analyses were conducted in MEGA4 (Tamura et al. 2007.

Results

Chromosome identification and features

Chromosome 7 has an asymmetric centromere position with a centromere index of 37 % and displays well distinguishable telomere blocks and pericentromeric regions. An ideogram and straightened version of the pachytene chromosome with the nomenclature for the different chromosome regions are given in Figure 1a-c. One of the unique features of this chromosome is that it has two condensed heterochromatin knobs halfway the long arm euchromatin (Figure 1a-b). The short and long arm euchromatin measure 6.9 µm and 20.9 µm respectively, whereas the pericentromere heterochromatin is 10.1 µm (Ramanna and Prakken, 1967). DNA amount is estimated at 79 MB of which 68.2 % is heterochromatic (Chang et al. 2008).

For establishing the repeat-rich regions with respect to the borders of heterochromatin, euchromatin and centromere, we hybridised the chromosomes with Cot 100 that includes high, moderate and low copy sequences of the genomic DNA. FISH showed that the Cot 100 DNA covered the pericentromeric heterochromatin, the telomere ends and the proximal heterochromatic knob on the long arm euchromatin on chromosome 7 (Figure 3). The border of the euchromatin - heterochromatin on the short arm was sharp, in contrast to the long arm where the transition of heterochromatin and euchromatin was more gradual. These borders follows essential the characteristics of chromosome 6 (Szinay et al. 2008).

In addition, we mapped 60 repeat-poor BACs on chromosome 7, aiming to anchor BAC contig sequences (see below) on the genetic map of this chromosome. Fifty-one out of the 60 BACs had chromosome positions that corresponded to their positions on genetic maps; eight gave a FISH signal on one of the other chromosomes, whereas one BAC could not be mapped due to high repeat content. These cytogenetically mapped BACs were used further to anchor BAC contigs (see below) on the genetic map of this chromosome.

BAC sequencing and contig assembly of chromosome 7

The sequencing of chromosome 7 is part of the international initiative aiming at sequencing the tomato genome (http://sgn.cornell.edu/about/tomato_sequencing.pl, Mueller et al. 2009). The present study is partly based on the use of 15.25 Mb of non redundant sequences generated on chromosome 7. Among the 168 sequenced BACs and Fosmids analyzed for chromosome 7, 143 are distributed over 33 contigs, whereas the remaining 43 are singletons (Figure 1d). The BAC sequences are available under NCBI Genbank database (http://www.ncbi.nlm.nih.gov/Genbank/GenbankSearch.–html).

By BAC-FISH mapping, these BAC contigs were cytogenetically mapped and also anchored onto the genetic map of chomosome 7. Molecular markes with known genetic positions were identified on the BAC sequences of established BAC contigs, which revealed inverted map positions between the genetic positions of few markers and their physical positions (Figure 2). In the most distal part of the short arm, close to the telomere, 13 molecular markers within

2.3 cM were identified in the sequence of three BAC contigs covering about 300 kb of these markers, and the genetic orders of six markers are in agreement with their physical orders, wherease seven show wrong map positions (Figure 2A). In the euchromoatin region on the long arm, a 400 kb contig presented in Figure 2B covers the genetic distance of about 6.5 cM

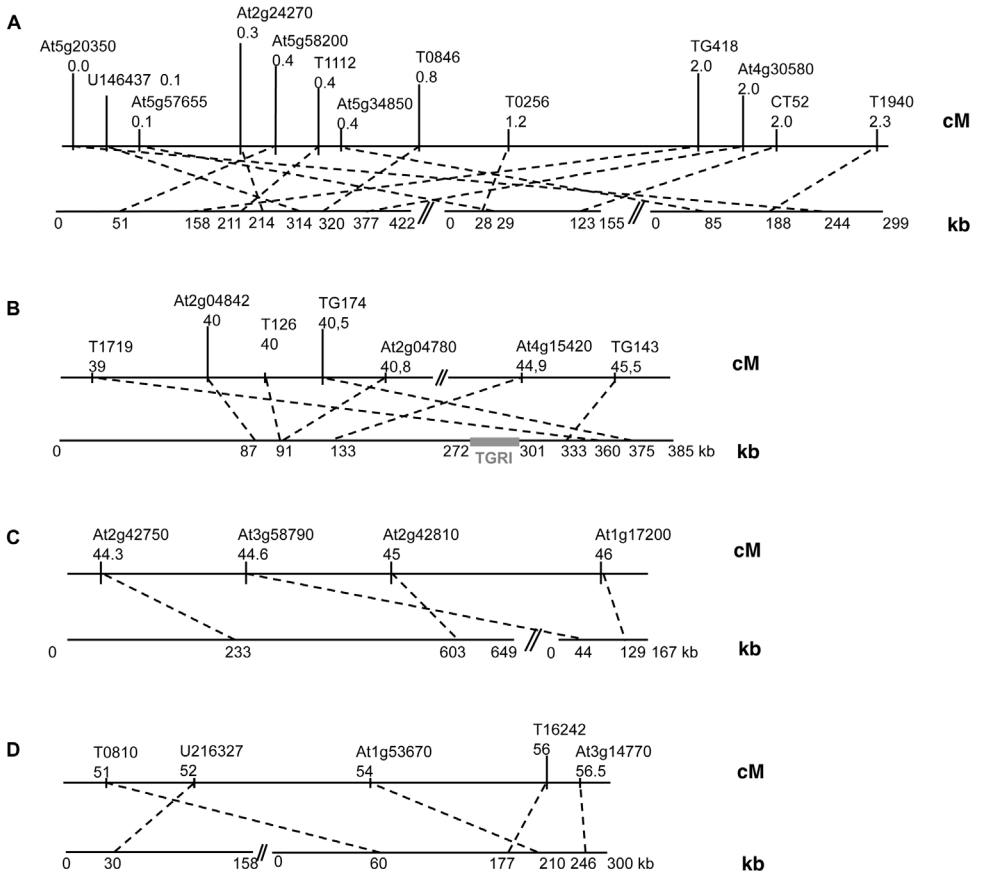


Figure 2 : Comparison of the genetic and physical maps for certain parts of Chromosome 7. Top bar represents the genetic pap positions of molecular markers in cM. Lower bar shows the physical positions of the same markers in the established BAC contigs (in kilobases). A : First contig close to the short arm telomere is composed of five BACs : SL_EcoRI0110K10, SL_EcoRI0111B06, LE_HBa0002D20, LE_HBa0166N19 and LE_HBa111F22. Second contig is composed of two BACs and two Fosmids : SL_FOS0014I11, LE_HBa0162M15, SL_FOS0085P03 and LE_HBa0062O11. Third contig is composed of 3 BACs: LE_HBa0033O01, SL_MboI0046H06 and LE_HBa0127J08. B : The contig is composed of five BACs : SL_MboI0093E04, LE_Hba0066L12, SL_MboI0126F03, LE_Hba0043E14 and SL_MboI0034N13. C : First contig is composed of 9 BACs :LE_HBa0076O09, SL_MboI0096B05, SL_MboI0075L20, SL_EcoRI0020F06, LE_HBa0166A09, LE_HBa0049P16, SL_MboI0104D24, LE_HBa0023C09 and SL_EcoRI0095F20. Second contig is composed of one Fosmid and one BAC : SL_FOS0095A17 and LE_HBa0018L21. D : First contig is composed of two BACs : LE_HBa0220H02and LE_HBa0221C04. Second contig is composed of 3 BACs : SL_FOS0014I11, LE_HBa0162M15, SL_FOS0085P03 and LE_HBa0062O11. Third contig is composed of 3 BACs: SL_EcoRI0124H12, LE_HBa0059A10 and LE_Hba0102J11.

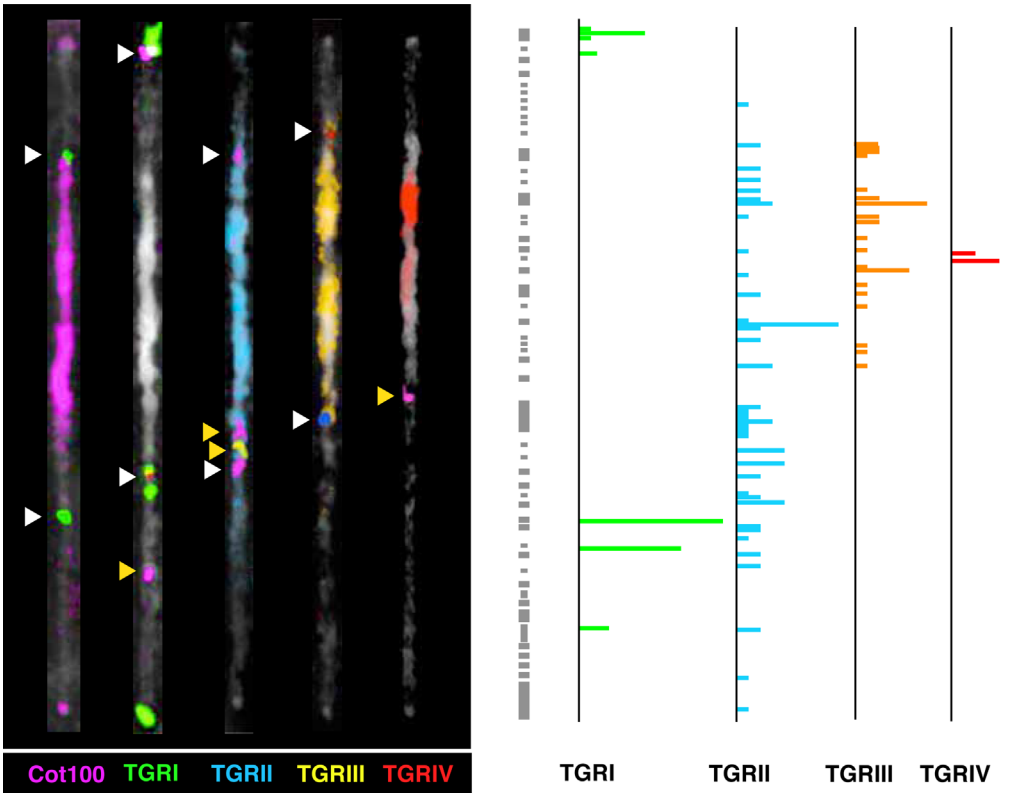


Figure 3. Major repeat positions on tomato chromosome 7. Border BACs were indicated by white arrow heads. Yellow arrow heads point at diagnostic BACs needed for the identification of the chromosome. The right part shows the occurrence of the same repeats as detected in the sequenced BACs.

(from 39 to 45.5 cM). Seven molecular markers could be identified in this contig and only two of them are correctly mapped in the genetic map. Interestingly, markers that are genetically mapped at 44.9 and 45.5 cM are flanking a 29 kb block of 110 TGRI repeats and a second large block of 27 TGRI repeats. These repeats may influence recombination crossovers and explain the incorrect genetic positions of these marekrs. Figure 2C and D displays the chromosomal regions from 44.3 to 46 cM and from 51 to 56.5 cM, in which the molecular markers are also indentified with inverted genetic and physical map positions.

BAC-FISH painting for repeat demarcation

For an accurate position of the repeats we decreased stringency in some of our experiments (see Material and Methods) to visualise weak repeat signals under the microscope. A lower stringency gives a stronger signal and helps to detect small chromosomal targets that were otherwise invisible in FISH under standard conditions. The reason is that the small and important

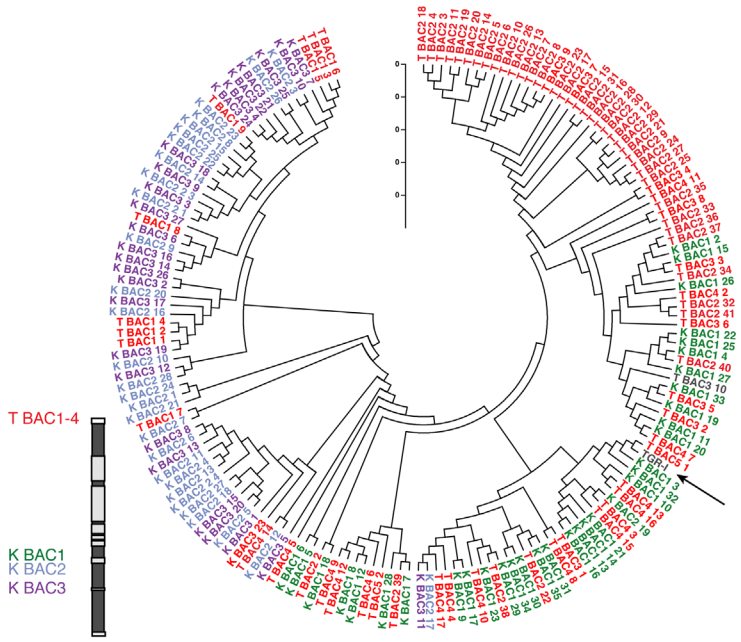
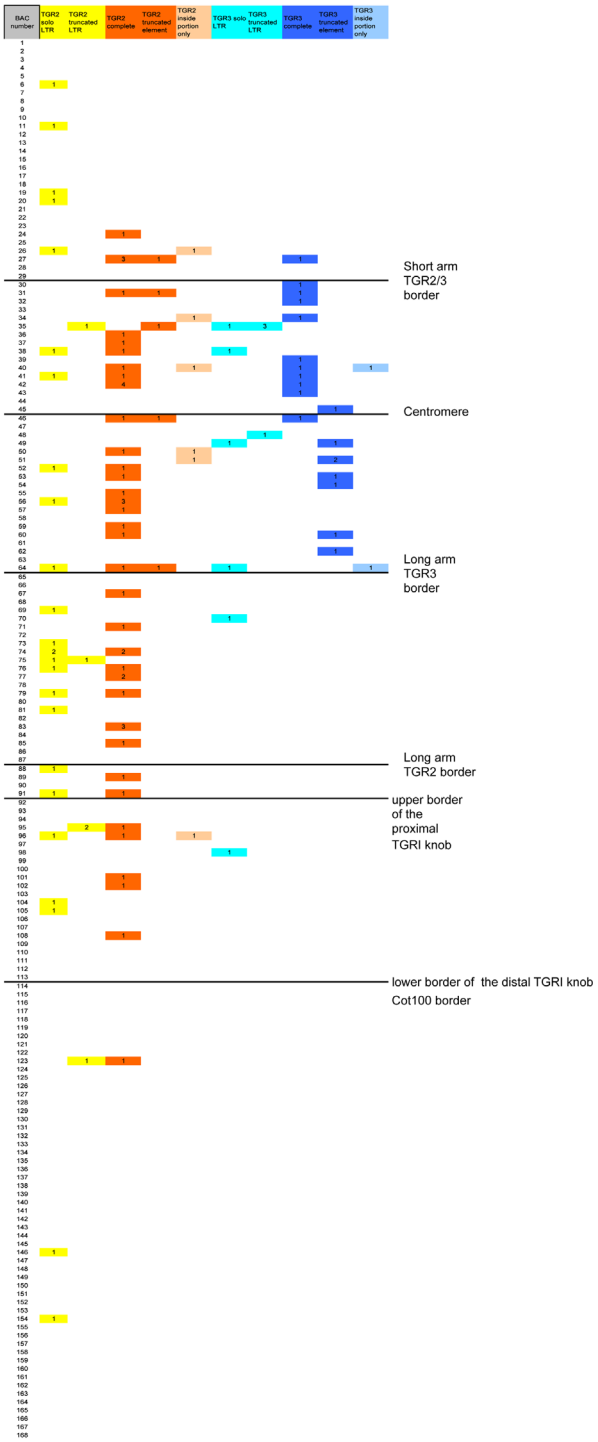


Figure 4. Evolutionary relationships of TGR1 sequences using the Neighbor-Joining method. Sequences were derived derived from 7 BACs as indicated in the small ideogram of chromosome 7. The telomeric short arm BACs are T BACs1-4, the interstitial knobs with TGRI are T BAC1-3. All positions containing gaps and missing data were eliminated from the dataset leaving a total of 52 positions in the final dataset. The Phylogenetic analyses were conducted in MEGA4. The black arrow points at the original published 162 bp TGRI sequence (Schweizer et al. 1988).

repetitive regions remain invisible with the high stringency especially around the euchromatin and heterochromatin borders. The DAPI images were displayed in grey and sharpened with a Hi-Gauss high-pass spatial filter to achieve more details in the pachytene chromosome morphology.

The combination of repeats and BACs using FISH is shown in Figure 3. FISH hybridization with Cot 100 and BACs revealed that BAC LE_HBa0030C22 is the best marker for the short arm euchromatin - pericentromere border and BAC SL_Mbo0137M12 for the long arm pericentromere - euchromatin border. The short arm border BAC in the Cot 100 hybridization locates at the border of the TGRII and TGRIII dispersed repeats. TGRII hybridised to most of the pericentromeric region including the centromere. On the long arm close to the euchromatin / heterochromatin border the fluorescence signal of TGRII is weaker and not continuous, and the border BAC is LE_Hba0241K09. In the short arm heterochromatin there is an almost perfect overlap of the Cot 100, TGRII and TGRIII. The coverage of TGRIII in the long arm pericentromeric region is continuous and about 20 % less than that of TGR II, and thus it ends within the heterochromatin. Therefore the border was easier to define with BAC LE_Hba0025K09. The third pericentromere LTR repeat, TGRIV, hybridised strongly on the proximal half of the short arm pericentromere region and weaker at the centromere region itself, and give a weak signal on the proximal half of the long arm pericentromere.

FISH revealed on the long arm two heterochromatin knobs that were expected to carry TGRI repeats. The results show the previously described TGRI domain at the short arm tel-



omere (Zhong et al. 1998), as well as on the long arm telomere. According to the BAC sequencing TGRI islands were expected on the long arm, but our FISH showed only two signals in that part of the chromosome. TGRIV could be clearly demonstrated on the chromosome 7, showing a big brightly fluorescing block on the short arm and a shorter in the long arm (Figure 3). However, BACs with TGRIV elements could not be analysed as their repeat content was too high for sequencing.

Extended DNA fibre analysis

The aim of extended fibre FISH was to analyse the molecular organisation of DNA sequences under the fluorescence microscope at the highest possible spatial resolution. Fransz et al. (1996) and Jackson et al. (1999) estimated the DNA size / microscopic length ratio at 2.7 - 2.44 kb/µm. Zhong et al. (1998) used the extended fibre FISH to study the molecular organisation of telomere repeats in tomato and determined DNA sizes for the TTTAGGG telomere and TGRI tandem arrays. As extended DNA fibres loose their chromosomal integrity and identity, extra chromosome specific markers are required to elucidate molecular size and patterns of specific TGRI loci on one of the twelve chromo-

Figure 5. Detailed map of repeat analysis of the chromosome 7 BACs, for the occurrence of complete and incomplete TGRII and TGRIII elements.

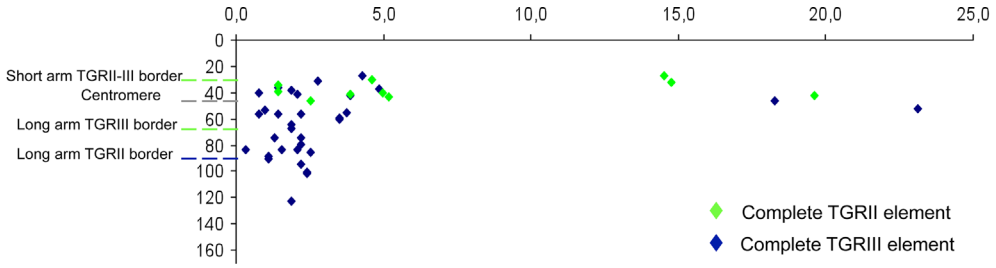


Figure 7. Age of TGRII and TGRIII element in the pericentromere of chromosome 7, based on the comparison of nucleotide substitutions in their sequences. The X-axis displays the estimated age of the insertion of the element in milions of years. The Y-axis shows the chromosome 7 BACs involved and presented by their order number (BACs are arranged from top to bottom of chromosome 7).

somes. So, for the domains of TGRI on chromosome 7 we did a 2-colour FISH experiment in which TGRI and one of the BACs (SL_EcoRI0110K10) adjacent to the distal TGRI tandem array were chosen as probes. The BACs were selected on the basis of their position in chromosome 7 pachytene FISH. It was argued that the BAC signal can serve as a landmark to identify the DNA fibres with TGRI signal from the specific chromosome 7 domain. Pilot experiments were successful for the TGRI signal (Figures 6A), but we could not establish a clear BAC signal in the extention of one of the TGRI strings. In addition, the BACs as probes showed specific signals, but weakly hybridised all over the fibres due to the presence of unknown repetitive sequences. Cot 100 DNA could not be used for blocking as it would also block the TGRI probe from hybridisation. In future experiments subclones of these BACs will be isolated that are free of repetitive sequences and can be used as better FISH markers for the identification of TGRI arrays in the extended fibre slides.

A second application of extended fibres that we tested is the FISH hybridisation of TGRII, TGRIII and TGRIV. Only two repeat classes could be tested at the same time as we were confined to the more sensitive indirect DNA labelling. Here we addressed the question how TGRII, TGRIII and TGRIV are intermingled with each other in different parts of the pericentromere. Marker BACs for chromosome 7 positioning are not available as the repeat concentration of the whole pericentromere is too high. Figure 6B and 6C give examples how TGRII, TGRIII and TGRIV are organised in fibres where two of them showed up. We found short stretches of the three pericentromere Ty3-*Gypsy* elements but still miss the essential information of their organisation with respect to the position between euchromatin - heterochromatin border and functional centromere.

Sequence analysis of the TGR repeats

TGRI analysis

The TGRI repeats retrieved from chromosome 7 BAC sequences have an average size of 179 bp, a size that is 17 bp longer then reported previously (Schweizer et al. 1988). They are present in

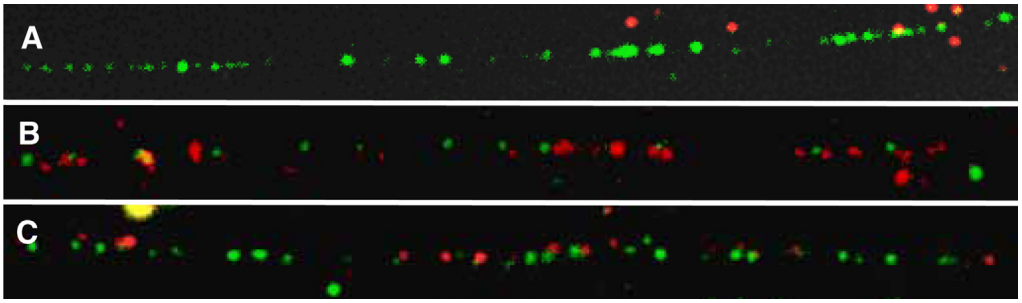


Figure 6. Extended fibre FISH of repetitive sequences in tomato. A. Fibres hybridised with TGRI; B. 2-colour FISH with TGRII (red) and TGRIV (green); C. Same with TGRIII (red) and TGRIV (green).

tandem arrays of 8 to 127 repeat elements (Table 1). On the short arm TGRI repeats are present close to the telomere, while on the long arm they are present in the heterochromatin knobs that are located in the middle of the euchromatin ocean of the long arm. As shown by Zhong et al. (1998) tomato TGRI arrays measure often more than 100 kb (Zhong et al. 1998), so the TGRI blocks found in the chromosome 7 BACs ranging from 1.5 to 29 kb cover only the small border parts of the long arrays. The four BACs situated in the vicinity of the telomere region of the short arm contain arrays of 8 to 42 TGRI elements (T BAC1-4). Three other BACs located near the interstitial heterochromatin knob in the long arm contain tandem arrays of 27 to 127 repeats (Table1, K BAC1-3). Notably, the number of repeats is much higher in the knob BACs than in the short arm distal region BACs. We have no BACs at the distal end of the long arm euchromatin containing TGRI sequences, however FISH on pachytene chromosomes reveal a clear TGRI focus on the long arm telomere.

In terms of sequence similarity, the 166 TGRI repeats could be grouped into few groups. The evolutionary history was inferred using the Neighbor-Joining method (Figure 4). The first group gathers TGRI repeats from blocks 2, 3 and 4 of the short arm (T BAC 1-4) and from block 1 of the long arm (T BAC1), all showing high sequence similarity to each other (88 to 94 % identity). The second group contains TGRI repeats present in T BAC 1 of the short arm and K BACs 2 and 3 of the long arm showing only 82 to 86 % identity both to each other and to blocks from the other group. In terms of sequence similarity, only one TGRI repeat has been described in

Table 1. Features of TGRI repeat sequences on chromosome 7.

name of the BAC	number of TGRI repeats	length of the TGRI block	name of the block
C07SLe0111B0	68 and 1 incomplete	1550	T BAC1
C07HBa0002D20	40 and 2 incomplete	7386	T BAC2
C07HBa0111F22	10 and one incomplete	1836	T BAC3
C07HBa0127J08	16 and 2 incomplete	3018	T BAC4
C07HBa0117J06	127	23751	K BAC1
C07HBa0043E14	110 (several truncated)	28980	K BAC2
C07HBa0076O09	27 and 1 incomplete	4941	K BAC3

the literature (Schweizer et al. 1988) and its sequence is closer to the first group described in the present study.

TGRII analysis

TGRII was defined as a 780 bp long repeat that could be found in both LTRs of transposable elements classified as a Ty3-*Gypsy* LTR transposable elements (Chang et al. 2008, Wang et al. 2006, Yang et al. 2005). The two LTRs encompass an internal region containing an unknown ORF followed by a polyprotein with a GAG domain, a reverse transcriptase domain, an RNaseH and an integrase domain. Chromosome 7 TGRII repeats match the previously described 780 bp long repeat TGRII (Chang et al. 2008) only in the region spanning 187 to 780 bp and no match is found with the first 187 bp. The present analysis revealed that TGRII corresponds to the region 1 to 593 bp and beyond this sequence, the two LTR sequences of each transposable element are very conserved and range in size ranges from 1594 to 2069 bp. Noteworthy, the sequence repeated was not restricted to the LTR region but extends to the whole transposable element. Fifty-seven of the 168 BACs were found to contain either a complete or a partial sequence the transposable element (Table 2). The elements ranged from 7706 to 10287 bp and display similarity with the Solanum lycopersicum clone BAC LE_HBa0040B13 Jinling retrotransposon (DQ445619.1, 8834 bp, Wang et al. 2006), PCTR1a-2 and PCTR1a-1 (AY850394 and AY850394, Yang et al. 2005).

Figure 5 shows an overview of all TGRII and TGRIII sequences in the sequenced BACs of chromosome 7. Most (83 %) of the complete TGRII elements are located in the heterochromatin compartment delimitated by FISH hybridizations with the repeat as probe, and cover the genetic region between 13 cM and 36 cM. If we consider the Cot 100 hybridization result, the heterochromatin compartment extends up to 42.5 cM on the long arm and only one complete TGRII element is located outside this part of chromosome 7 at 44.3 cM. Next to solo-LTRs we also detected complete elements and incomplete LTRs lacking genes or sequences of the retrotransposon. Solo-LTR and truncated Solo-LTR were located in BACs between 1.2 to 80 cM. A greater part (24 of 30) of the solo-LTRs are in the pericentromere containing strong TGRII fluorescence. The remaining six Solo-LTRs were found in the short or long arm euchromatin. We also found five truncated TGRII elements and eight sequences that we could not classify because they were cut off at the end of a

Table 2. Features of TGRII and TGRIII repeat sequences on chromosome 7.

	TGRII transposable element	TGRIII transposable element
LTR size	1594 to 2069 bp	1528 to 1599 bp
Complete element size	7706 to 10287bp	7867 to 14070 bp
Occurrence in chromosome 7 BACs		
solo-LTRs	25	6
truncated solo-LTR	5	4
complete element	46	11
truncated element	5	8
cut sequences	8	-
only internal portion, no LTR	6	2

BAC sequence or were present in a BAC sequence that was not finished yet (phase1). Wang et al. (2006) estimated the total number of LTR Jinling elements at 2000, whereas Ganal et al. (1988) came to a much higher estimation of around 4000. This discrepancy can be explained assuming that two TGRII should be counted for each transposable element. According to our data, by extrapolating the number of TGRII in chromosome 7, considering a sequencing bias due to our aim to sequence only euchromatin, is about 2875 complete elements per genome and 2678 supplementary LTR (Solo-LTR, truncated elements).

TGRIII analysis

The second class of the Ty3-*Gypsy* LTR transposable elements are the TGRIII repeats. Their LTRs are remarkably conserved in size and sequence in contrast to the corresponding internal regions that display high variation. The TGRIII repeats in chromosome 7 show high similarity with the PCRT2 element of Yang et al (2005), which was described as an *Athila* element. According to Wright et al. (2002) *Athila* belongs to a Ty3-*Gypsy* group retrotransposon with an env-like ORF. However, such an env-like sequence was not found in the TGRIII repeat in this chromosome. Out of 168 analysed BACs, 24 BACs contained TGRIII repeats among which 11 contain complete elements and 13 only showed a partial sequence of the element (Table 2). All complete transposable elements are concentrated in the heterochromatin compartment located between 13 and 22.3 cM (Figure 5). They are in the short arm of chromosome 7 except for one that was found in the long arm heterochromatin close to the centromere. Several truncated elements are present in the heterochromatin long arm between 22.3 and 28 cM. Solo-LTR and truncated solo-LTR are all found in the 20 to 38.4 cM heterochromatin region. Noteworthy, two out of 10 solo LTR could not be detected by FISH hybridization and are therefore located outside the heterochromatin compartment defined by FISH hybridization using a TGRII probe. However, these two solo-LTRs are located within the heterochromatic part of chromosome 7 as defined by FISH analysis using Cot 100 probe.

Based on hybridization to a Heinz 1706 tomato BAC library Yang et al. (2005) estimated more than 1,200 copies of PCRT2 (synonymous to part of TGRII), while Ganal et al. (1988) estimated the total number at 2100. According to our study based on extrapolation of TGRIII counts in chromosome 7 and with a sequencing bias due to our aim to sequence only euchromatin, we came to an estimation of 685 complete elements for the whole genome and 1121 supplemental LTR.

Ageing of transposition of LTR in TGRII and TGRIII

We noticed a significant difference in the distribution of TGRII and TGRIII transposable elements along the chromosome 7. To test an ageing effect of the insertions compared to their chromosomal position we want compared insertion events of these two types of Ty3-*Gypsy* LTR transposable elements. Ageing a transposition event can be done by aligning the two LTR of one complete element and comparing the nucleotide substitution in their sequences: the

two LTR sequences are used as an internal clock as the insertion of these LTR sequences was done on the same moment, and they were identical at this time. The estimation of the age of insertion was done using the following formula : t=K/2r where K is the Kimura parameter, and r is the synonymous substitution rate. The value of 4.54 × 10-9 substitutions per site per year was used for r (de Sa et al. 1996). Clearly, we can see that the TGRII elements localised from the TGRIII border down to the long arm are younger than the one located in the short arm heterochromatin and in the long arm heterochromatin (Figure 7). As to the TGRIII elements this was not that clear, as they are fewer (only 11 complete elements) and substitutions pointing at different ages. We could not conclude that TGRIII elements are older then TGRII elements.

Comparing FISH and molecular data of TGRII and TGRIII

Dispersed repeats – such as TGRII and TGRIII - have not been easy to handle due to their distribution across the genome. FISH reveal approximate positions and rough details compared to sequencing data. While FISH distinguished the same border of both repeats on the short arm (by BAC 29), molecular data reveal complete and truncated TGRII motifs in 3 more BACs above it. Moreover solo-LTRs of TGRII are present up to BAC 6. These sequences could not be detected by FISH. TGRIII repeats are less deviating between the compared methods. Also here the two solo-LTRs could not be demonstrated by FISH.

Discussion

This study showed the benefit of a combined FISH mapping, DNA sequencing and bioinformatics of the two major repeat types in heterochromatin regions of tomato. The comparative analysis was possible thanks to the large number of 168 sequenced BACs covering a greater part of chromosome 7 including all border regions of heterochromatin and euchromatin except the long arm distal block. FISH and BAC sequence data are mutually complementary: the former allows mapping of DNA sequences on chromosomes and shows the relation to telomeres, centromeres and heterochromatin, even when these regions are very repeat-rich. The latter reveals the precise DNA information but is limited to BACs with a repeat content that is low enough for sequencing.

In the first part a comparison was made between the positions of the genetic markers, and their position on the physical map. Clear discrepancies were observed in the top of the short arm, between 0 and 2.3 cM, and in the middle of the long arm, with various reversed marker orders between genetical and physical maps (Figure 2). This result shows a striking correspondence with the mapping errors that were described earlier for tomato chromosome 6 (Szinay et al. 2008). Although mapping inaccuracies in chromosome ends cannot be avoided, the discrepancies in the middle of the long arm are surprising and not easy to explain. What all regions have in common is the presence of TGRI sequences at or near the inaccurate mapped

loci. In the discussion below we will further hypothesise about a putative effect of repeat dynamics and their putative role in causing deviating genetic marker positions.

Borders of heterochromatin and euchromatin

The borders of the heterochromatin domains, although sharp on the chromosomal level, have intriguing transition zones on the molecular scale (Figure 3). Firstly, the top border, marking the transition of distal heterochromatin and euchromatin in the short arm is characterised by a sharp decrease of TGRI in the BACs. More distal BACs containing longer and more continuous stretches of TGRI array are lacking, but the extended fibre FISH of the repeats for chromosome ends in tomato showed that proximal ends of some of the distal TGRI domains have an interrupted irregular transition (Zhong et al. 1998). The short arm euchromatin - heterochromatin border shows a slight albeit interesting difference. On the microscopic level, the brightly fluorescing heterochromatin begins at the BACs 29 and 30 (Figure 5), whereas molecularly, a sudden transition can be observed from the euchromatin containing only solo-LTRs (from BAC 0 to 23) and the region where complete and truncated TGRII emerge, so starting at BAC 24. The centromere is between BAC 45 and 46. The borders in the long arm are more complex, both at the chromosomal level as well as on the molecular scale. The transition in the DAPI stained chromosomes is more gradual, a phenomenon that was also reported for chromosome 6 (Szinay et al. 2008). The BACs containing complete TGRII were found in the whole heterochromatin region (until BAC 108, plus a single more distal BAC 123). Interestingly, the border of the long arm telomere was determined by FISH on pachytene that we considered to be the border of the TGRI repeat as well. For more detailed analysis fibre FISH is necessary especially in this region. In the transition zone of the long arm heterochromatin - euchromatin we observed a few heterochromatic knobs that are rich in TGRI sequences.

Organisation of TGRI

A closer look at the chromosome 7 repeats showed that the molecular organisation and distribution of TGRI is essentially different from the Ty3-*Gypsy* retrotransposons, TGRII, TGRIII and TGRIV. Zhong et al. (1998) studied repeat size and distribution of TGRI and the telomere TTTAGGG using FISH on pachytene and extended DNA fibres of VFNT Cherry tomato. The telomeres of all but three chromosome arms display clear TGRI domains, whereas the chromosomes 4, 6, 8 and 9 have interstitial TGRI containing knobs in the long arm. The length of the distal TGRI tandem arrays varied from 223 and 1330 kb, whereas the lengths of the interstitial segments were estimated at 88, 157, and 634 kb. Interstitial knobs are well documented in maize (Ananiev et al. 1998, and references therein), are highly variable and known to contain tandem repeats. Also in tomato interstitial knobs varies strikingly between genotypes and cultivars, and have been noticed in the middle of the long arms of tomato chromosomes 3, 4, 6, 7, 8, 9 and 12 in other cultivars (Zhong et al. 1998), but more chromosome arms may be involved as

well. Also, to our knowledge no other report claimed two interstitial knobs in the long arm of chromosome 7, but that can be explained with the more sensitive TGRI painting that detects even smaller heterochromatic knobs.

The general accepted explanation for interstitial knobs is formulated in the theory of 'equilocal' (equal distance) distribution of Heitz (1932), and has since then be confirmed for many species showing interstitial heterochromatic knobs or C- bands (e.g., Greilhuber and Loidl, 1983; overview in Guerra, 2000). Bennett (1982) postulated a mechanistic model, in which the repeats are transferred between chromosomes that are spatially organised through the anaphase conformation (Rabl 1885), in which non-homologous sites can occupy adjacent domains in the nuclear matrix. If one assumes that chromosome arm lengths at pachytene are proportional to their length in interphase nuclei and that a copying mechanism between adjacent short arm telomeres and long arm interstitial knobs can occur between any combination of chromosomes, than the most proximal interstitial TGRI island may have originated from the arm telomeres of 8S, 9S, 10S or 12L. In a similar way, the more distal interstitial knob can have originated from the arms 5S, 5L, 11L and 12L.

The model of equilocal repeat or heterochromatin distribution has few, yet very fascinating aspects that can be envisioned in the frame of modern nuclear repeat dynamics and genomics. Recent studies revealed a vast body of evidence that tandem arrays can be transposed by Helitron transposons (Kapitonov and Jurka 2007), producing extrachromosomal circular DNA molecules (Pont et al. 1987; Cuzzoni et al. 1990; Cohen et al. 2008; Navrátilová et al. 2008). It is tempting to assume that such transpositions also hold true for TGRI and allow these sequences to footprint other adjacent chromosome domains in trans while chromosomes occupy their anaphase configuration in the nuclear matrix. However, the footprints are highly variable between genotypes and cultivars and this may reflect that such repeat transpositions result from recent dynamic processes. Moreover, it has been shown that such unstable knobs may influence meiotic pairing, homologous recombination and meiotic drive (Buckler et al. 1999; Kikudome 1959; Rhoades and Dempsey 1966) and hence can explain the difficulties in genetic mapping of these regions in the chromosomes 2, 6, 7, and 12 (Szinay et al. 2008; this paper; Drs. D. Choi and G. Giuliano, pers. comm.).

The sequencing of TGRI arrays in different BACs also shed light on the origin of the repeats. We compared complete TGRI sequences in BACs from four chromosome regions, i.e., five BACs of the short arm Telomere (T BAC1-5), and three interstitial (Knob) BACs in the long arm (K BAC 1, 2, 3) in a phylogenetic tree. As shown in Figure 4, a clustering is apparent leaving most of the distal short arm sequences clearly separated from those of the long arm interstitial K BAC 2 and 3, whereas interstitial K BAC 1 is mixed up with the T BAC 1-4 sequences. This result is surprising and suggests that TGRI domains can have slightly different motifs and if so, that part of the interstitial K BAC TGRI's originated from TGRI sites of other chromosomes. However, future analysis of TGRI domains from other chromosome arms are necessary to strengthen this hypothesis.

Organisation of the pericentromere repeats TGRII, III and IV

The most important classes of LTR retrotransposons in tomato are the Ty1-*Copia* and the Ty3-*Gypsy*. Various studies on plant chromosomes have documented that both repeat classes have essentially different chromosomal domains, in which Ty1-*Copia* resides more in euchromatin, whereas Ty3-*Gypsy* dominates heterochromatin areas (Belyayes et al. 2001; Brandes et al. 1997; Mroczek and Dawe, 2003; Lamb et al. 2007b; Presting et al. 1998; Heslop-Harrison et al. 1997; Pearce et al. 1996, 1997; Pich and Schubert, 1998). Peters et al. (2009) found in their BAC sequencing of tomato chromosome 6 that the ratio of Ty3-*Gypsy* and Ty1-*Copia* in euchromatin is 2:3 and in heterochromatin 3:2. Additional evidence for chromatin-specific retrotransposon integration came from studies on yeast (Kim at al., 1998) showing that Ty1, Ty3, and Ty5 integrate into specific chromosomal sites. Chalker and Sandmeyer (1992) and Devine and Boeke (1996) found that the preferred targets for Ty1 and exclusive targets for Ty3 are near tRNA genes or other genes transcribed by RNA polymerase III (Pol III) transcription. Here targeted integration by Ty3 is more precise, and typically occurs within two bases of the start of Pol III transcription, whereas Ty1 has a more regional target preference and inserts within about a 1 kb window up-stream of target genes. Another surprising observation was that the overall density of retrotransposon insertions varies enormously among the yeast chromosomes, and is in general, higher for the smaller chromosomes. This result corresponds to the relatively higher heterochromatin content in the smaller chromosomes of tomato as well as in several other species. The study of Gao et al. (2008) revealed a dynamic interplay between retrotransposons and heterochromatin, in which the mobile elements recognise heterochromatin at the time of integration and then stabilises the heterochromatic mark by epigenetic reprogramming. Pareira (2004) studies LTR retrotransposons in *Arabidopsis* and observed that Ty1-*Copia* elements insert randomly along the chromosomes, have been recently active and show no insertion bias, but insertions tend to be lost from euchromatin regions. In contrast, the Ty3-*Gypsy* preferentially targeted to heterochromatin, and was more active in the past.

Less conceivable are the differences in distribution between the three Ty3-*Gypsy* LTR retrotransposons (TGRII, TGRII and TGRIII) within the pericentromere. FISH of the repeats on pachytene chromosomes showed that all occupy the pericentromere heterochromatin, but overlap only partly: the most distal parts, flanking the borders with euchromatin we find a small region in the short arm and a longer region in the long arm with only TGRII. Then closer to the centromere both TGRII and TGRIII are detected, where the most proximal parts display co-localisation of TGRII, TGRIII and TGRIV. The sequence analysis of these repeats in the BACs are in agreement with that, and also the pilot experiments with repeats hybridised on the extended DNA fibres (Figure 7) showed intermingled patterns of TGRII, III and IV, resembling the complex organisation of repeats in rye (Alkhimova et al. 2004). Further extended fibre FISH in tomato will be needed to better understand how these repeats are organised in the pericentromere.

This study did not yet reveal a clear explanation for the selective forces for integration, and loss by recombination. One of the theories is that mobile elements have target integrations to specific chromosomal sites (Bushman et al. 2003), hinting at a common targeting mechanism of tethering of integration complexes to proteins bound at favourable sites. However, additional mechanisms may contribute to integration site selection, including age of the insertion and chromosome domain dispositions. As Pereira (2004) observed, younger Ty3-*Gypsy* elements have a preference further away from the centromere. In our study, this relation was true only for the TGRII elements distal from the TGRIII long arm border that are younger than the TGRII and TGRIII in the short arm and the TGRIII in the long arm. Additional factors for successful integration or loss of repeats one can speculate about are the chromatin organisation of the centromere regions in the nuclear matrix, the epigenetic status of the heterochromatin and the lack of meiotic recombination closer to the centromere, but additional research is needed for a better understanding of these processes.

A final remark about the significance of repeat distribution is in tomato for the ongoing genomics programs. As we have shown clearly in this paper, repeats or combinations of repeats occupy their own specific chromosomal domain, and so can contigs better be assigned to chromosome positions, even without a priori knowledge of the genetic or chromosomal positions. It means that BACs rich in TGRII, III and IV are close to centromeres, BACs with TGRI and TGRII are likely parts of interstitial long arm domains, whereas contigs with only Solo-LTR of TGRII are likely euchromatin regions. It is likely that this assignment can be further improved if more repeats with specific distributions will become available.

Chapter 4

Cross-Species Bacterial Artificial Chromosome-Fluorescence in Situ Hybridization Painting of the Tomato and Potato Chromosome 6 Reveals Undescribed Chromosomal Rearrangements

Xiaomin Tang*,[1], Dóra Szinay†,[1], Chunting Lang†, Munikote S. Ramanna*, Edwin A. G. van der Vossen*, Erwin Datema‡, René Klein Lankhorst‡, Jan de Boer*, Sander A. Peters‡, Christian Bachem*, Willem Stiekema‡, Richard G. F. Visser*, Hans de Jong† and Yuling Bai*,[2]

* Wageningen-UR Plant Breeding, 6708 PB Wageningen, the Netherlands
† Laboratory of Genetics, Wageningen University, 6703 BD Wageningen, the Netherlands
‡ Centre for Biosystems Genomics, 6708 PB Wageningen, the Netherlands

1. These authors have contributed equally to this work.
2. Corresponding author: Wageningen-UR Plant Breeding, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands. Email: bai.yuling@wur.nl

Appeared in Genetics 180: 1319–1328 (November 2008). Reproduced with permission of the publisher.

Abstract

Ongoing genomics projects of tomato (*Solanum lycopersicum*) and potato (*Solanum tuberosum*) are providing unique tools for comparative mapping studies in *Solanaceae*. At the chromosomal level, BACs can be positioned on pachytene complements by fluorescent *in situ* hybridization (FISH) on homoeologous chromosomes of related species. Here we present results of such a cross-species multicolour cytogenetic mapping of tomato BACs on potato chromosomes 6 and *vice versa*. The experiments were performed under low hybridization stringency, while blocking with Cot-100 was essential in suppressing excessive hybridization of repeat signals in both within-species FISH and cross-species FISH of tomato BACs. In the short arm we detected a large paracentric inversion that covers the whole euchromatin part with breakpoints close to the telomeric heterochromatin and at the border of the short arm pericentromere. The long arm BACs revealed no deviation in the colinearity between tomato and potato. Further comparison between tomato cultivars Cherry VFNT and Heinz 1706 revealed colinearity of the tested tomato BACs, whereas one of the six potato clones (RH98-856-18) showed minor putative rearrangements within the inversion. Our results present cross-species multicolour BAC-FISH as a unique tool for comparative genetic studies across *Solanum* species.

Keywords: chromosome painting, cross-species FISH, inversion, tomato, potato

Introduction

The first cornerstone of the International *Solanaceae* Genome Project (SOL) launched in November 2003 is the sequencing of the euchromatin part of the tomato (*Solanum lycopersicum*) genome by an international consortium of 10 countries (http://www.sgn.cornell.edu/). In The Netherlands, the Centre for BioSystems Genomics (CBSG) is in charge of the sequencing of tomato chromosome 6. This chromosome contains various genes for economically important traits, including resistance genes for *Oidium neolycopersici* (Ol-4 and Ol-6), *Cladosporium fulvum* (Cf-2 and Cf-5), root-knot nematode, aphids and whitefly (Mi-1 and Mi-9) and the tomato yellow leaf curl virus (Ty-1, Ty-3 and Ty-4) (Weide et al. 1993; van Daelen et al. 1993; van Wordragen et al. 1994, 1996; Milligan et al. 1998; Ammiraju et al. 2003; Bai et al. 2004; Seah et al. 2004; Ji et al. 2007).

The Dutch tomato sequencing project of chromosome 6 follows the BAC (Bacterial Artificial Chromosome) walking procedure or BAC-by-BAC approach, which involves the anchoring of a limited number of BAC clones (seed BACs) to the genome and further BAC contig building via BAC extension (Peters et al. 2006). For this approach, genetic mapping data, FISH and BAC sequencing analysis provide a framework, in which 84 anchored seed and extension BACs were positioned, covering a total of 2.4 Mb for the short arm and 10.2 Mb for the long arm (Peters et al. in prep.). Similarly, BAC-FISH map has been published for tomato chromosome 1 showing the relation of the linkage map to pachytene chromosome structure (Chang et al. 2007). Moreover, a BAC-FISH map is being completed at http://www.sgn.cornell.edu/cview/ to illustrate the orders and locations of tomato BACs on pachytene chromosomes. Such a full set of BACs forms a chromosomal scaffold along the chromosome and can be used to compare chromosomal colinearity between related species and to unravel chromosomal rearrangements by cross-species BAC-FISH painting. Large-scale genomic changes involving chromosomal inversions and/or interchanges can be important for species isolation and might also contribute to the phenotypic differences between species through possible effects on gene structure or expression (Tanksley et al. 1992; Livingstone et al. 1999; Doganlar et al. 2002).

FISH allows the simultaneous localization of different target sequences on chromosomes, depending on the number of fluorochromes with different excitation and emission wavelengths and the use of COmbined Binary RAtio (COBRA) as well as related labelling technologies (Raap and Tanke 2006). For basic FISH only red and green fluorochromes for probe detection are used, together with 4', 6-diamidino-2 phenylindole (DAPI) for counterstaining of chromosomal DNA. Advanced multi-colour FISH can involve up to 12 different fluorescent dyes together with DAPI as counterstain in a single experiment (Muller et al. 2002). In tomato, 5-colour high resolution FISH mapping has been successfully applied to process large number of BACs on chromosome 6 (Szinay et al. 2008).

Cross-species FISH painting was first applied to mammalian chromosomes and human chromosome probes have now been hybridized to metaphases of over 100 species (Rens et al. 2006b). In plants, it has been accomplished in *A. thaliana* and related species of the Brassi-

caceae family (Lysak et al. 2003, 2005 and 2006). In addition, the small genome of Sorghum has been used as a basis for integrating genetic and physical maps across grass genera with larger genomes (Draye et al. 2001; Koumbaris and Bass 2003).

Although the *Solanaceae* represents one of the best-studied and attractive plant systems for comparative genetics (Bonierbale et al. 1988; Fulton et al. 2002; Grube et al. 2000; Doganlar et al. 2002; Tanksley et al. 1992), applications of BAC-FISH for studying chromosomal evolutionary processes and chromosomal rearrangements have not been undertaken except for the very recent study by Iovene et al. (2008, accompanying article this issue), due to the lack of defined BAC libraries. So far, genome-wide colinearity within Solanaceae has been studied only with genetic maps. For example, comparative maps (Tanksley et al. 1992; Fulton et al. 2002; Doganlar et al. 2002) have revealed that tomato (*S. lycopersicum*) and potato (*S. tuberosum*) are differentiated by a series of paracentric inversions (inversions that do not involve the centromere) of chromosome 5, 9, 10, 11 and 12 (Bonierbale et al. 1988; Grube et al. 2000). The inversion on chromosome 10 demonstrated that *S. lycopersicoides* and *S. sitiens* are co-linear with *S. tuberosum* (Pertuze et al. 2002), suggesting that this inversion was fixed in the common ancestor of the tomato lineage. The limitations in such comparative genetic linkage mapping studies are that (1) mapping populations are needed; (2) deviations can occur between genetic and physical chromosome maps and (3) the large pericentromere regions contain markers with low genetic resolution in mapping because of the absence of crossovers. The latter two limitations were encountered in the tomato genome sequencing project, in which BACs were selected on the basis of genetic markers and the physical positions of these BACs were validated by FISH prior to sequencing. Discrepancies have been observed between the actual chromosomal positions of some of these BACs and the positions of their corresponding markers on the genetic map (Chang et al. 2007 and our unpublished data). This was most notable in the repeat rich domains in highly condensed pericentromere heterochromatin where crossovers were almost absent (Sherman and Stack 1995).

With the aims to study chromosomal colinearity between tomato and potato, we developed a cross-species multi-colour BAC-FISH technique for the *Solanum* species (Szinay et al.

table 1. Plant material used in this study.

	genotype	genetic background
tomato	Cherry VFNT (LA1221)	*Solanum lycopersicum* with introgressed *S. peruvianum*
	Heinz 1706	*S. lycopersicum*
potato	G254	Diploid Gineke
	RH88-025-50	F1 *S. tuberosum* × *S. phureja*
	RH98-856-18	F1 *S. sparsipilum* × *S. tuberosum*
	RH90-038-21	BC1 (*S. tuberosum* × *S. microdontum*) × *S. tuberosum*
	RH97-654-15	F1 *S. tuberosum* × *S. spegazzinii*
	CD1015	(*S. phureja* × *S. tuberosum*) × (*S. tuberosum* × (*S. phureja* × *S. tuberosum*))

Table 2. Overview of the tomato and potato chromosome 6 BACs used in this study.

	BAC[a]	Genetic map position[b] (cM)	Molecular markers	BAC size (kb)	Chromosome position by FISH[d]
6S/PCtomato	H107A05	3	T1188	166	6S/EU
	H054K13	3	T1182	160	6S/EU
	H153O03	5	T1198	NA[c]	6S/EU
	H251G05	5	T1198	98	6S/EU
	H112G05	5.5	*Mi*	91	6S/EU
	H073H07	5	*Mi*	82	6S/EU
	H250I21	5.5	*Mi*	148	6S/EU
	H024L21	6.5	TG436/SSR47	75	6S/EU
	H288L16	10	cLET-2-H1	112	6S/EU-PC
	H304P16	10	cLET-2-H1	124	6S/EU-PC
	H309K01	10	cLET-5-A4	102	6L/PC
	H295L11	10	T0244	110	6L/PC
	H003K02	10	TG178	110	6L/PC
	H242H19	12	T1063	98.2	6S/PC
	H023B17	25	FER	111.9	6L/PC
	H261A18	28	cLET-4-G2	106	6L/EU
	H059K09	41.3	NA	NA	6S/PC
	H106K23	44	C2_At4g10030	89	6L/EU
	H194N16	45	cLET-5-C8	93	6L/EU
	H176D13	45.6	NA	NA	6S/PC
	H026E06	47	P27	130.3	6L/EU
	H097D13	47.7	NA	NA	6S/PC
	H012O10	48	C2_At1g73885	80	6L/EU
	H309D09	50	TG365	142	6L/EU
	H060A01	101	Ct_At1g20050	168	6L/EU
potato	112M11	NA[b]	NA	NA	6S/EU
	RH026H24	1.6	EACAMAGG_94	NA	6S/EU
	67P23	NA	CT119	NA	6S/EU
	RH034P18	7	EACCMACT_286	NA	6S/EU
	RH069B12	10.7	EAACMCCT_377	NA	6S/PC
	RH084A13	12.2	EACGMCTA_215 EACGMCTA_215	NA	6S/PC

[a] All tomato BACs are from the Heinz 1706 HindIII library; the four RH potato BACs are from the RHPOTKEY BAC library; the other two potato BACs were kindly donated by Edwin A.G. van der Vossen (van der Vossen et al. 2005). [b] Tomato map position was adopted from the tomato-EXPEN 1992 map (Tanksley et al. 1992); the potato map position was adopted from the ultra-dense RH genetic map (van Os et al. 2006). [c] NA=not available. [d] FISH, fluorescent in situ hybridization; S = short arm; L = long arm; PC = pericentromere heterochromatin; EU = euchromatin; Cent = centromere; EU-PC = border between euchromatin and pericentromere heterochromatin.

2008). By applying this technique, we painted tomato BACs of chromosome 6 on potato chromosomes and *vice versa* and discovered a new paracentric inversion in the short arm euchromatin. Our results show that the cross-species multi-colour FISH strategy provides a powerful tool with the potential application for comparative genetics in the genus *Solanum*.

Material and methods

Plant materials and BAC clones

For preparing cell spread preparations we used anthers of the tomato (*S. lycopersicum*) cultivars Cherry VFNT (LA1221) and Heinz 1706, the potato (*S. tuberosum*) diploid genotype G254, and five diploid potato clones, RH88-025-50, RH98-856-18, RH90-038-21, RH97-654-15 and CD1015. Detailed genetic background of these plant materials presented in Table 1. For tomato, 25 tomato BACs were included in this study (Table 2). At the time of this study only six potato BACs were available for the short arm (Table 2).

Cot-100 DNA

Cot-100 fractions of tomato genomic DNA was prepared according to Zwick et al. (1997) with some modifications. Total genomic DNA was isolated and sonicated to a fragment size of about 1kb. The fragmented DNA (0.5 μg/μL) was denatured in 0.3 M NaCl at 95 °C for 10 min, and then let it reanneal at 65 °C for 37 hr 40 min. The remaining single strand DNA (ssDNA) was digested with S1 endonuclease (Fermentas, final concentration 1 U/μg) for 90 min at 37 °C. The reaction was stopped and extracted by adding 300 μL Chloroform-Isoamylalcohol (24:1). Then the DNA solution layer was transferred to a new tube, 2.5 volume of ice cold absolute alcohol was added to precipitate DNA, and the dry pellet was resuspended in 20 μL HB50 (pH8.0).

Fluorescence in situ hybridization (FISH)

Pachytene chromosome preparations were made as described by Zhong et al. (1996a) with few minor modifications. BAC DNA was isolated using a standard alkaline extraction and labelled by standard Digoxigenin or Biotin nick translation mix according to the instructions of the manufacturer (Roche Diagnostics). Two-colour FISH of BAC clones to pachytene chromosomes was performed according to the FISH protocols (Zhong et al. 1996b). Probes labelled with digoxigenin-dUTP, which were detected by digoxigenin-FITC, gave the green colour, biotin-dUTP-labelled probes, detected by Avidin-Tex-Red showed the red colour and Streptavidin-Cy5 the purple colour. For direct labelling in multi-colour FISH, five fluorescent nucleotides were used. They are Fluorescein-12-dUTP (FITC), Cy3-dUTP, Cy3.5-dCTP, Cy5-dUTP and Diethylaminocoumarin-5-dUTP (DEAC). Cy5 was also used in an indirect labelling with Biotin-

dUTP-Streptavidin-Cy5 detection (see below). The labelling methods followed the protocols of Amersham Bioscience (GE Healthcare, Sweden).

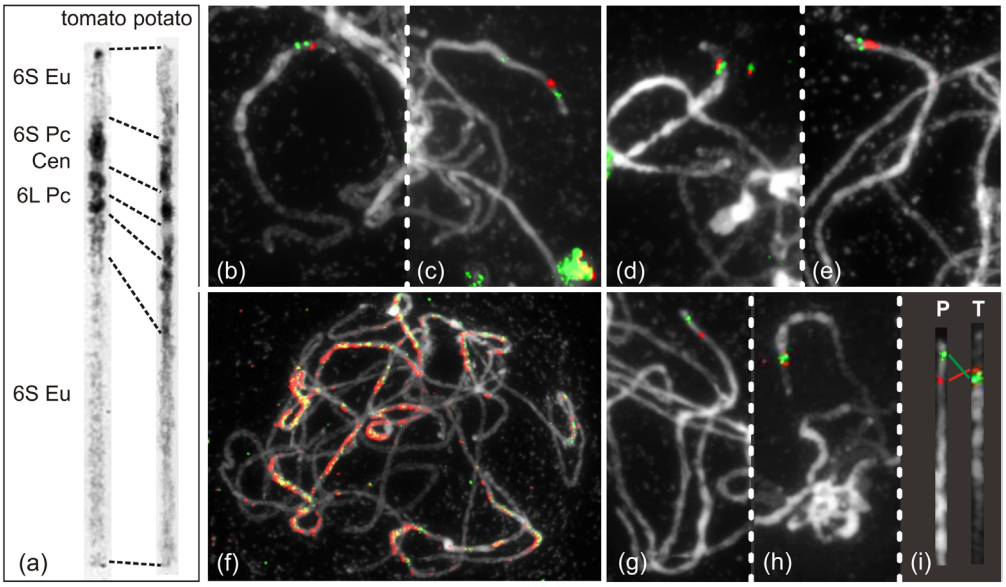
Cross-species was adapted with some minor modifications following the published protocol for cross-species chromosome painting (Rens et al. 2006a). For those BACs inside heterochromatin, 2µg (100x probe concentration) of Cot-100 DNA is sufficient for blocking if 20ng of a BAC probe is used per slide. Slides were examined under a Zeiss Axioplan 2 Imaging Photomicroscope equipped with epifluorescence illumination, filter sets for DAPI, FITC, Cy3, Cy5, DEAC and Cy3.5 fluorescence. Selected images were captured by a Photometrics Sensys 1305×1024 pixel CCD camera. Image processing and thresholding was performed with the Genus Image Analysis Workstation software (Applied Imaging Corporation). DAPI images were separately sharpened with a 7×7 Hi-Gauss high pass spatial filter to accentuate minor details and heterochromatin differentiation of the chromosomes. The different FISH signals were captured consecutively by double or multiple exposures and combined in a multichannel mode. Fluorescence images were displayed in dark grey for DAPI and pseudocoloured for the other colours. Further brightness and contrast improvement were done on the whole image in Adobe Photoshop. We used ImageJ (http://rsb.info.nih.gov/ij) for measurements and for straightening of the chromosomes (plugin of Kocsis et al. 1991).

Results

Chromosome 6 of tomato and potato at pachytene stage

We first compared the morphology of the DAPI stained pachytene chromosomes 6 of tomato and potato. Figure 1a displays converted black-white images of these chromosomes, which were straightened and stretched to equal length and slightly sharpened for better heterochromatin differentiation. The tomato chromosome 6 has an asymmetric centromere position and characteristic heterochromatin blocks in the long and short arms (Figure 1a). In addition, the short arm has the shortest euchromatin region in the complement constituting approximately 4.1 Mb of euchromatin (Chang et al. 2008). Potato chromosome 6 has a submedian centromere and its diagnostic heterochromatin blocks are less condensed than those of tomato (Figure 1a). The borders between euchromatin and heterochromatin are also gradual in the short and long arms of the potato chromosome. Besides, many tiny chromomeres in the euchromatin were observed. The short arm has a small distal knob that was seen in most chromosomes (Figure 1c and d), but sometimes was absent in the straightened chromosome of potato (Figure 1a). A second small knob just below the distal knob in the short arm euchromatin of potato, which was described as a diagnostic heterochromatic knob for chromosome 6 by Ramanna and Wagenvoorst (1976), was not visible here. Conceivably, the knob is polymorphic and not visible in the potato clones that we used for our work, or it could not be detected in our DAPI stained preparations. Similarly, we did not find a knob on the long arm of potato chromosome 6 as reported by Iovene et al. (2008).

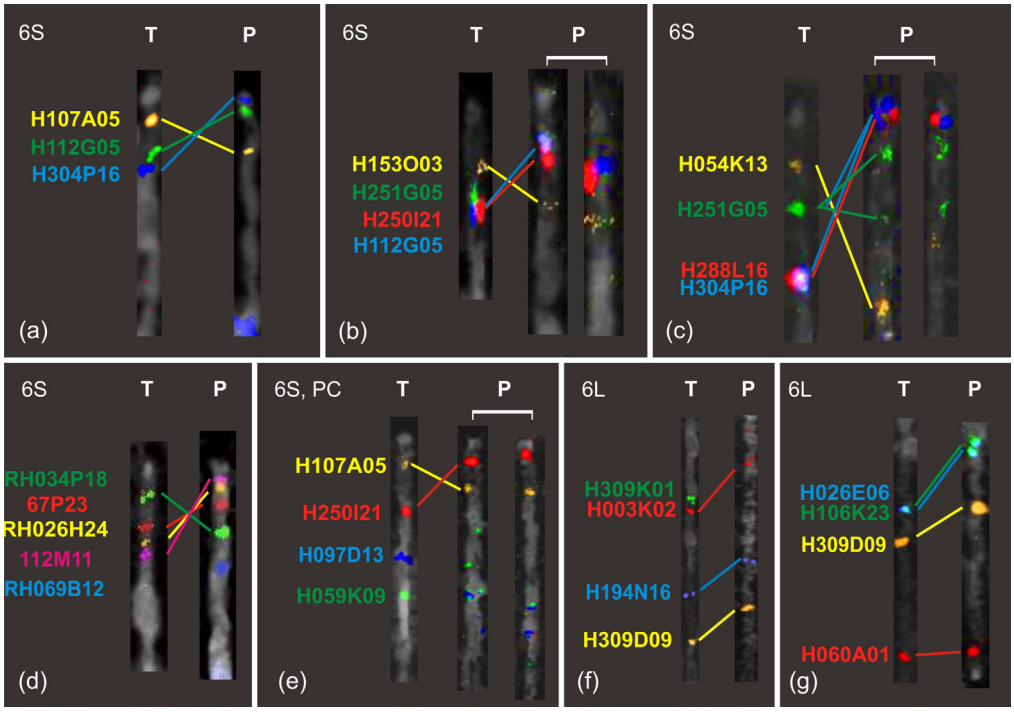
Figure 1. (a) Chromosomes 6 of tomato and potato at the pachytene stage. S, short arm; L, long arm; Eu, euchromatin; Pc, pericentromere heterochromatin; Cen, centromere. (b) FISH of the tomato BACs H153O03 (red) and H073H07 (green) on tomato pachytene chromosome 6. (c) Cross-species FISH of the same BACs on potato chromosome 6. (d) FISH of H112G05 (red) and H24L21 (green) on tomato chromosome 6. (e) Cross-species FISH of the same BACs on potato chromosome 6. (f) Cross-species FISH of the tomato BACs H003K02 (green) and H309K01 (red) on potato chromosomes without Cot-100 blocking. (g and h) FISH of the potato BACs 67P23 (red) and 112M11 (green) on potato RH98-856-18 chromosome 6 (g) and tomato chromosome 6 (h). (i) The straightened part of chromosome 6 of potato (P) and tomato (T), showing the orientation and relative distance of the two potato BACs 67P23 (red) and 112M11 (green).



Selection of BACs and cytogenetic maps for tomato and potato BACs on chromosome 6

We focused on the regions of tomato chromosome 6 which are rich for resistance genes, as preliminary comparative mapping studies within the Solanaceae genera showed that resistance genes occurred at syntenic positions in cross-generic gene clusters more frequently than expected by chance (Grube et al. 2000). Moreover, we selected potato BACs that are in the chromosomal region for which previously published data on genetic colinearity was controversial or doubtful. For example, the chromosomal region where the *Mi*-1 gene is located (Tanksley et al. 1992; van Wordragen et al. 1994; van der Vossen et al. 2005). For tomato we selected 14 BACs for the short arm (6S) and 11 for the long arm (6L) (Table 2 and Figure 3), on the basis of known positions on the chromosome maps (our unpublished data). The physical positions of the BACs in the 6S euchromatin were in agreement with their relative orders on the genetic map, except for H153O03 that was genetically mapped at 5 cM but was located closer to the distal telomere knob of tomato 6S than other BACs which had the same genetic positions of 5 cM (Figure 1b). BACs H250I21, H112G05 and H073H07, which had been mapped

Figure 2. Examples of FISH and cross-species FISH of tomato BACs on pachytene chromosome 6 of tomato (T) and potato (P). The chromosome regions of interest were straightened and oriented with the signals close to the short arm telomere upward. (a) FISH of H107A05 (orange), H112G05 (green), and H304P16 (blue) on the short arms of tomato and potato showed a clear inverted arrangement of the BAC signals. (b) FISH of H153O03 (orange), H250I21 (red), and H112G05 (blue) showed an inverted order between the homeologs. (c) FISH of H054K13 (orange), H251G05 (green), H288L16 (red), and H304P16 (blue) showed an inverted order between the homeologs. Notably, here we used potato clone RH98-856-18, and the H251G05 (green) BAC produced a large and a small focus on the potato chromosome, suggesting a breakpoint in this BAC for a putative chromosomal rearrangement. (d) FISH of RH034P18 (green), 67P23 (red), RH026H24 (orange), 112M11 (pink), and RH69B12 (blue) showed an inverted order between the homeologs; RH069B12 did not give a signal on tomato. (e) FISH of H107A05 (orange), H250I21 (red), H097D13 (blue), and H059K09 (green) on the short arm. The two pericentromere heterochromatin BACs H097D13 (blue) and H059K09 (green) showed weak and variable foci on the potato short arm. (f) FISH of H309K01 (green), H003K02 (red), H194N16 (blue), and H309D09 (orange) on the long arm showed the same order of the BACs except H309K01 (green) hybridization that gave no signal in potato. (g) FISH of H026E06 (blue), H106K23 (green), H309D09 (orange), and H060A01 (red) on the long arm showed the same order on the tomato and potato chromosomes.



genetically around 5 cM and assembled in the BAC contig containing the *Mi*-1 gene, were partly overlapped in the euchromatin region of pachytene chromosome 6 (Figure 2b T and 3). Although the signal of H073H07 showed a single focus in most of the pachytene complements, we also observed cases of clear double signals (Figure 1b). BACs H107A05 and H054K13 overlapped by FISH mapping and were identified as the most distal BACs in the euchromatin region so far (Figure 2a and c (T) and 3 (tomato)). BACs H288L16 and H304P16 co-localized at the border of 6S euchromatin and heterochromatin, which were shown to be the most proximal BACs on 6S euchromatin (Figure 2a and c (T) and 3 (tomato)). In the heterochromatin region of

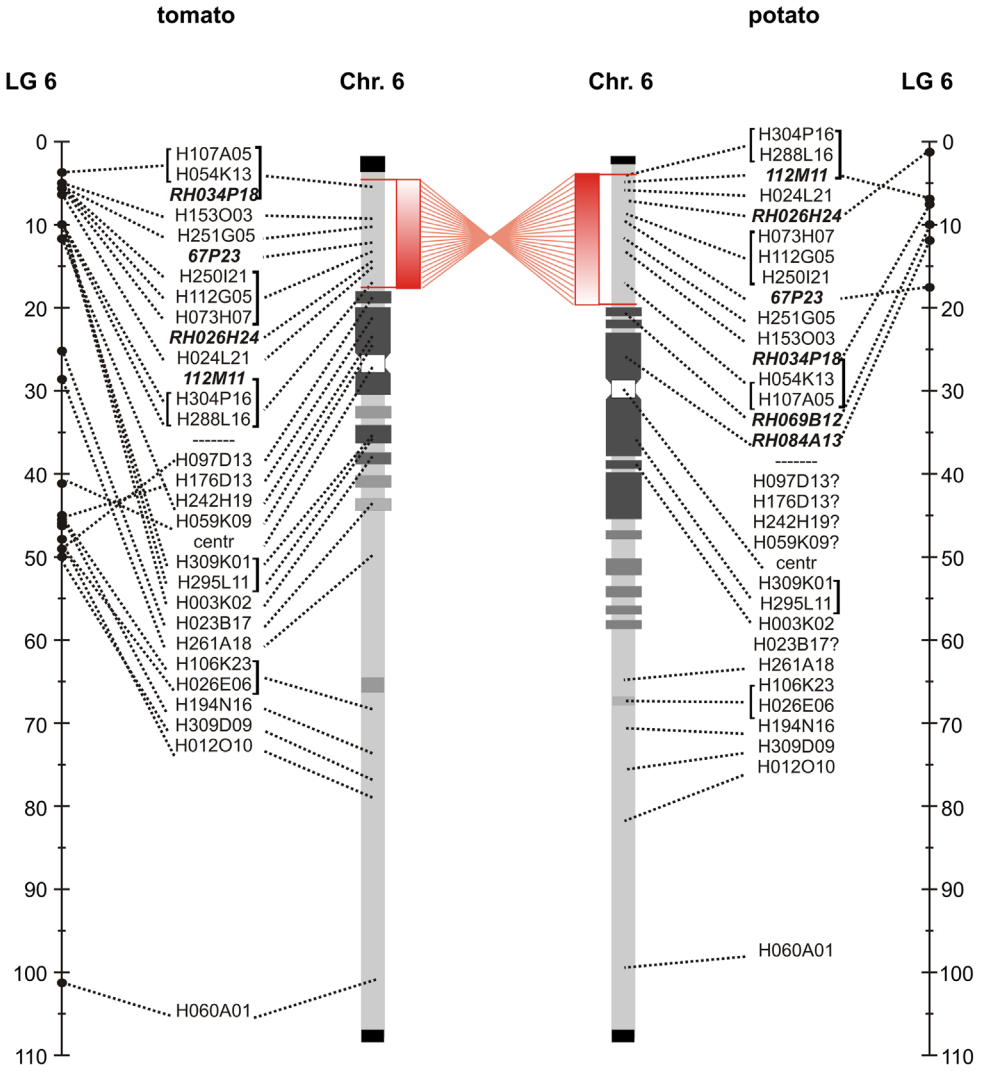


Figure 3. A comparison of genetic and physical maps for tomato BACs and cross-species FISH of tomato and potato BACs on pachytene chromosomes. The schematic drawings of the chromosomes are based on pachytene morphology. Black and dark gray blocks are heterochromatin regions; the dark blocks represent the dense brightly fluorescing heterochromatin regions, whereas the lighter regions are lighter and more variable; the white blocks are the centromeres. The BACs are positioned in sequence of FISH position. Brackets on the left of the BACs have the same genetic map positions; brackets on the right have overlapping FISH signals on the chromosome. The dotted lines show the position of their markers on the genetic map. BACs in bold italics are the potato BACs. The red bar between the tomato and potato chromosomes represents the short arm paracentric inversion; the green block indicates the positions of minor chromosome rearrangements between potato clone RH98-856-18 and the other five remaining potato lines. The question marks point at weak, variable, or no signals of the tomato BACs on potato.

tomato 6S, four BACs were selected. While, three of them (H097D13, H176D13 and H059K09) have been genetically mapped on the long arm and the fourth one (H242H19) mapped around the centromere (Table 2 and Figure 3). On the tomato 6L, we positioned four BACs in the pericentromere heterochromatin and seven in the euchromatin. The pericentromere BACs, except H023B17, harbor markers that are genetically located at the same locus as the 6S BACs H288L16 and H304P16. Clearly, discrepancies between genetic positions and relative physical positions exist, especially when the candidate BACs come from chromosomal regions near or in heterochromatin (Figure 3). For potato, only six BACs were available for the short arm (Table 2). BAC 112M11 was located in the very distal and RH034P18 was in the proximal euchromatic region of the short arm of potato (Figure 1g and 2d P). The cytogenetic and genetics map orders of all the potato BACs were in agreement (Figure 3).

Within- and Cross-species BAC-FISH in tomato and potato

A crucial factor for cross-species FISH is the post-hybridization controlling stringency. When the same washing stringency (50% formamide, 2×SSC at 42 °C for 15 min) as within-species FISH was used, the signals were found to be not highly specific or reproducible (data not shown). In order to enhance hybridization efficiency, post-hybridization washes under conditions of low stringency were carried out for 3×5 minutes in 20% formamide, 2×SSC at 42 °C. With this stringency, nearly every tested BAC could be painted across species.

FISH signal intensity depends on the chromosomal target size and repeat content. The tomato BAC clones in this study have genomic inserts of 75-168 kb (Table 1), which are large enough to produce bright fluorescent foci on the pachytene chromosomes. However, BAC clones harbouring high amounts of tandem and dispersed sequences, produce abundant fluorescence signals over all chromosomes, mostly in the pericentromeric regions. We therefore used Cot-100 to suppress hybridization by the highly and middle repetitive DNA sequences of the BAC probes when the BACs were hybridized to chromosome targets. Examples of such tomato BACs FISH with Cot-100 were shown in the Figure 1b, 1d, 2a-2c T and 2e-2g T. However, the potato BACs in this study produced specific foci even in the absence of Cot-DNA, suggesting a very low repeat content in these BACs (Figure 1g and 2d P).

Initial cross-species FISH with tomato BACs on potato chromosomes at a low hybridization stringency and without Cot-100 blocking demonstrated excessive cross-hybridization signals (Figure 1f). Then repeat signals could effectively be suppressed in the presence of tomato Cot-100 to paint tomato BACs on potato chromosomes (Figure 1c, 1e, 2a-2c P and 2e-g P). To hybridize the potato BACs on tomato chromosomes, potato Cot-100 was still not needed to produce clear signals (Figure 1h and 2d T).

Colinearity and rearrangements of BACs between tomato and potato chromosome 6

Upon mapping of several 6S euchromatin tomato BACs to potato, we found that the orders of any two BACs were inverted between tomato and potato (Figure 1b-1c and 1d-1e). The most distal tomato 6S BAC H107A05 (or H054K13) flipped to the proximal euchromatin of potato 6S, whereas H304P16 (or H288L16), a tomato 6S euchromatin/heterochromatin boundary BAC, mapped at the most distal region on potato 6S nearly covering the telomere knob (Figure 2a P and 2c P). Similarly, painting the four potato 6S euchromatin BACs on tomato (Figure 2d) also showed an inverted order of these BACs between tomato and potato. The most distal potato BAC 112M11 was clearly positioned at the border of euchromatin and heterochromatin on tomato 6S (Figure 1h and 2d T). Further FISH mapping of 112M11 together with H304P16 (or H288L16) demonstrated that they co-localized. In addition, H107A05 (or H054K13) showed an overlap with RH034P18 (Figure 3). These results suggest that an inversion involving the whole 6S euchromatin and possibly even involving telomere exists between tomato and potato.

To further investigate whether the 6S inversion involved also the pericentromere heterochromatin, we analyzed four tomato and two potato BACs from the pericentromere region. BACs H097D13 and H059K09, which produced single foci on the tomato chromosome 6, appeared in two or more copies on the potato homeolog with weak and not reproducible signals in separate experiments (Figure 2e). Moreover, two more tomato and two potato pericentromere BACs gave no signals on their potato and tomato homeologs, respectively (Figure 2d). As for the long arm, we observed that the 11 used tomato BACs mapped to tomato and potato at comparable positions, with the exception of H023B17, a pericentromere heterochromatin BAC of 6L, which gave no signal in potato. This reflects that the long arm of chromosome 6 is well conserved between these two species (Figure 2f, 2g and Figure 3).

Sequence feature of BACs in heterochromatin region

Of the seven BACs that could not be efficiently painted cross-species, three have been sequenced. These are the tomato BACs H023B17 (6L), H242H19 (6S) and potato BAC RH069B12 (6S). For the long arm, sequence analysis of H023B17 revealed a lower gene content compared to BACs (H309K01, H295L11 and H003K02) that did paint the heterochromatin of both tomato and potato 6L. H023B17 contains only one putative gene, covering 1.8% of the BAC sequence. In contrast, there are four putative genes (16.3%) in H309K01; three (11.4%) in H295L11 and two (6.8%) in H003K02. Since Cot-100 was used to block the repeat sequences in painting the tomato BACs on potato, a high repeat content in H023B17 is likely the cause of the loss of its signal in potato. Alternatively, this might be caused by the lack of homolog sequences from this BAC in the corresponding potato region.

In the 6S heterochromatin, potato BAC RH069B12 and tomato BAC H242H19 were found to be highly repetitive. The majority of repeats in these two BACs were similar to the *Gypsy*-type GYPSODE1_I retrotransposon. The putative gene content of BACs RH069B12 and H242H19 was low (3.1 and 3.2% of the BAC sequence length, respectively). Since Cot-100 was not applied

to paint potato BACs on tomato, it seems that gene content and homology as well as repeat sequences played a role in the failure of cross-species FISH of BACs in 6S heterochromatin region. Alternatively, this region could be involved in the 6S inversion which has lead to chromosomal rearrangements and/or loss of chromosome fragments.

Colinearity of chromosome 6 BACs within tomato and potato

Most of our tomato FISH work was done with the Cherry VFNT (LA1221) cultivar containing an introgression of *S. peruvianum* in 6S; while the BACs were from Heinz 1706 cultivar that does not have this introgression. To verify the colinearity of the chromosome 6 BACs between VFNT and Heinz 1706, we used 26 tomato BACs covering the whole linkage group from 0 cM to 101 cM (Supplementary Table). The results showed that the order of the BACs on chromosome 6 of Cherry tomato is the same as that of the Heinz 1706 (data not shown), demonstrating that the *S. peruvianum* introgression does not contain large scale chromosomal rearrangements.

As to the question of colinearity of the selected BACs within potato we compared the potato (*S. tuberosum*) diploid genotype G254 with the diploid clones RH88-025-50, RH98-856-18, RH90-038-21, RH97-654-15 and CD1015 (Table 1). In general, all BACs that we used for the comparison displayed comparable positions on the short arms of chromosome 6. However, clone RH98-856-18 containing an introgression of *S. sparsipilum* showed a striking difference in the relative distance between the potato BACs 67P23 and 112M11 covering the *Mi* region in the inversion between tomato and potato (Figure 1g-1i). Measurements of the BAC distances in five pachytene complements demonstrated that the distance between the two potato BACs is about 1/3 shorter in tomato than in potato clone RH98-856-18, while such differences do not exist with the other potato clones. We also observed that tomato BAC H251G05 produces two signals in this potato clone (Figure 2c) in contrast to one signal in the other clones. Both observations suggest the existence of a second nested inversion or other minor rearrangement in the middle of the short arm of RH98-856-18 with probably one breakpoint in the chromosomal target area of BAC H251G05.

Discussion

Cross-species multi-colour BAC-FISH: a powerful tool for comparative genomics across *Solanum*

Cross-species BAC-FISH was previously applied to *Arabidopsis* and related *Brassicaceae* species for demonstrating chromosomal evolutionary processes and rearrangements (Fransz et al. 2000; Jackson et al. 2000; Lysak et al. 2005, 2006 and 2007). In *Solanum* species chromosomal rearrangements have not been cytologically studied so far. Within the scope of the ongoing tomato and potato sequencing projects, chromosome specific BACs have been ob-

tained by genetic and physical mapping and by contig construction. In this study, we present a multi-colour cross-species BAC-FISH painting for directly displaying synteny between related *Solanum* species on the chromosomal level. It facilitates the simultaneous detection of more than two BACs in one pachytene preparation and enables colinearity studies on BACs while avoiding laborious re-probing of FISH experiments. We applied this technology successfully to paint a set of tomato and potato BACs onto chromosome 6 of potato and tomato. These experiments revealed both agreements and discrepancies between genetic and physical locations of tomato BACs, the paracentric inversion on 6S and the colinearity of BACs on 6L between tomato and potato as well as some minor rearrangements in one of the potato lines. Our results show that cross-species multi-colour BAC-FISH is a powerful tool to connect genetic and physical maps. This tool can be used for comparative genetic and evolution studies to reveal genome colinearity between tomato and potato and most likely also among different genomes across the Solanum species. Without using mapping populations, high-density BAC-maps can be readily obtained for many species and accessions within one species.

Structural chromosome rearrangements may exist among different species (interspecific) as well as within accessions of the same species (intraspecific). Examples are the 7S paracentric inversion between a distant wild relative of tomato *S. pennellii* and *S. esculentum* / *S. pimpinellifolium* (van der Knaap et al. 2004) and an inversion of the two clusters of *Mi*-1 homologues between *S. esculentum* and *S. peruvianum* (Seah et al. 2004). The possible minor chromosomal rearrangements in one of the potato lines used in this study and the absence in our potato lines of an interstitial heterochromatin knob of 6L as described in Iovene et al. (2008) indicate the existence of intraspecific variation. The extent of such rearrangements may parallel morphological diversity or taxonomical groupings and would contribute to our understanding of the importance of this type of mutation in the evolution of fertility barriers in the whole *Solanaceae* family (Perez et al. 1999). Currently, we have ongoing projects to construct such a BAC-synteny-map for potato and tomato (including several related wild species). This BAC-synteny-map will enable to identify chromosomal linearity/rearrangement(s) between tomato and potato as well as among different wild species. The latter will help to identify chromosomal regions showing deviation and to provide a basis for breeding strategies to introgress genes from wild *Solanum* species into cultivated crops. Furthermore, this BAC-map can be used to study chromosomal evolutionary processes within *Solanum* at a variety of taxonomic levels and to understand biodiversity with genomics data.

The 6S inversion encompasses a hot-spot of resistance genes

Initial macro-synteny studies of genetic linkage maps between tomato and potato did until now not clearly show an inversion in the 6S chromosome arm, though the marker order between *GP*164 and *GP*79 was reported to be inverted (Tanksley et al. 1992 and van Wordragen et al. 1994). Recently, the Rpi-blb2 gene conferring late blight resistance in potato was mapped as a Mi-1 gene homolog on 6S of potato (van der Vossen et al. 2005). Both *Mi*-1 and *Rpi-blb*2

are tightly linked to a common RFLP marker CT119, with *Mi*-1 proximal to CT119 in tomato and Rpi-blb2 distal to CT119 in potato (van der Vossen et al. 2005), indicating that a hidden inversion may exist between tomato and potato. In this study, by using cross-species multi-colour BAC-FISH analysis, we provide firm evidence for a conspicuous paracentric inversion between tomato and potato covering the entire short arm of chromosome 6.

In general, inversions do not change the phenotype of the individual unless a breakpoint of the inversion lies within the regulatory or structural region of a gene. However, with the reshuffling of gene order within chromosome arms, the changed context may also affect the functions of genes involved in the inversion to some extent (Hoffmann et al. 2004). The paracentric inversion of 6S that differentiate tomato and potato probably has moved genetic loci from regions of low recombination (*e.g*. centromeres) to regions of higher recombination (and *vice versa*) and therefore has changed the evolutionary perspective for those loci. Interestingly, in tomato, 6S is a chromosomal region where many resistance (R) genes reside. This R gene hot-spot contains the *Cf*-2/*Cf*-5, *Ol*-4/*Ol*-6, *Mi*-1/*Mi*-9 and *Ty*-1 genes, which confer resistance to several unrelated pathogens. R genes in plants are most frequently members of multigene families and locate in tandem arrays, like the *Mi* or *Cf* genes. Inter- and intragenic recombination at R gene loci has been described extensively and is thought to be a major mechanism for generating novel resistance specificities (reviewed in Hulbert et al. 2001). Previously, the physical position of the *Mi*-1 gene and its six homologs was mapped also using FISH at the border of euchromatin and heterochromatin regions of tomato chromosome 6S (Zhong et al. 1999). Suppression of recombination frequency is apparent in the Mi-1 gene region in tomato. However, the *Rpi-blb*2 gene in potato is mapped on the euchromatin of potato 6S, and resides in a gene cluster that is twice as big as the *Mi*-1 gene cluster in tomato (van der Vossen et al. 2005). We speculate that the 6S inversion and observed expansion of the locus reflects the opposite evolutionary potentials of the interacting pathogens. Root-knot nematodes have little potential for gene flow and thus exert little evolutionary pressure on the host to generate new specificities. *Phytophthora infestans* on the other hand is a high risk pathogen with both sexual and asexual reproduction systems resulting in dynamic spread of genetic variation (McDonald and Linde 2002). Availability of more genome sequences of both tomato and potato in the near future will allow us to study the sequence composition near the breakpoints, possibly by virtue of cross-species microarray painting (Ferguson-Smith et al. 2005). This will shed light on the molecular mechanisms underlying chromosome rearrangements in plant genomes and will thus form a basis for investigating mechanisms of gene regulation, evolution, signaling, disease resistance and defense, the phenotypic diversity and comparative biology in the Solanaceae.

The molecular nature of the chromosomal inversion

Studies of chromosomal inversions were pioneered in *Drosophila* more than 60 years ago (Sturtevant 1919; Dobzhansky 1970). Of the two types of inversions, paracentric and pericentric

inversion, it has been assumed that the former is less affecting fitness and thus be the most likely form of chromosomal rearrangements to survive through evolutionary time. Although the number of BACs was insufficient to establish the precise breakpoints of the inversion, the 6S inversion discovered in this study is most likely a paracentric inversion covering the entire euchromatin region and possibly involves the telomere. This can be deduced from the physical position of potato BAC 112M11 and tomato BAC H304P16 (or H288L16). These three BACs hybridized to positions close to the border of euchromatin and heterochromatin of tomato 6S, while on potato 6S, it was found at the most distal region of short arm euchromatin and nearly covering the telomere (Figure 2a, c and d). The putative proximal breakpoint of this inversion likely occurs in the short arm pericentromeric heterochromatin as tomato/potato BACs in these regions produced weak and variable signals or no signals at all in the cross-species FISH. Evidence on chromosome breakpoints in these regions came from Khush and Rick (1963, 1968) and Liharska et al. (1997) on radiation-induced deletion mapping studies.

Whether the inversion involves the most distal short arm euchromatin and heterochromatin block is still to be confirmed as BACs or sequence data between the most distal BACs (potato BAC 112M11 or tomato BAC H304P16 / H288L16) and the subtelomere repeats (Figure 2a, c and d) are lacking. Although telomere sequences have been shown in (peri)centromere regions of tomato and potato (Ganal et al. 1991; Presting et al. 1996; Tek and Jiang 2004; our unpublished data), it is not clear as to whether these telomere sequences resulted from inversion events. So, the 6S inversion between tomato and potato may be explained in two ways, either as the result of a single break in the pericentromere where the broken end formed a new telomere de novo (Blackburn 1991; Werner et al. 1992), or through the simultaneous incidence of breaks in the distal part of the short arm euchromatin and in the pericentromere itself. Further, it has to be proved in material showing the subdistal heterochromatic knob as mentioned by Ramanna and Wagenvoorst (1976) whether this knob represents one of the breakpoints of the inversion, as is the case with the short arm heterochromatic knob in the Arabidopsis representing a paracentric inversion between the accessions Col and WS and the knobless accessions Ler, C24, Zh and NoO (Fransz et al. 2000; Lysak et al. 2002).

Acknowledgements

This research was done in the framework of the EU-SOL Project PL 016214 (financed by the European Commission), of the Centre for BioSystems Genomics (CBSG) which is part of the Netherlands Genomics Initiative / Netherlands Organisation for Scientific Research and of the Dutch Potato Genome Sequencing Consortium projects (funded by Netherlands Genomics Initiative and Ministry of LnV via the Fonds Economische Structuurverstersterking).

Chapter 5

Chromosomal rearrangements among Solanum species and their relatives visualized by BAC-FISH

Dóra Szinay[1,2,3], Erik Wijnker[1], Ronald van den Berg[3], Jan de Boer[2], Richard GF Visser[2], Hans de Jong[1] and Yuling Bai[2,4]

1. Wageningen University and Research Centre (WUR), Laboratory of Genetics
2. WUR, Laboratory of Plant Breeding
3. WUR, Biosystematic Group, Centre for BioSystems Genomics
4. Author for correspondence: WUR, Laboratory of Plant Breeding, PO Box 386, 6700 AJ Wageningen and bai.yuling@wur.nl

Abstract

Chromosomal rearrangements are rare evolutionary events and can be used as markers in phylogenetic analyses. In the genus *Solanum* such rearrangement mainly involve inversions. We visualized these inversions by using multicolour fluorescence in situ hybridization (FISH) with selected tomato and potato BACs as probes. The BACs that were evenly distributed over the chromosome arms under study were hybridized to a range of *Solanum* species, including tomato breeding lines and related wild species, and representatives of the crops potato, egg-plant and pepper. In this way we studied chromosome synteny between these species. Our results confirmed previously reported inversions, but also showed five new inversions. We compared our synteny study with published *Solanum* phylogenies, and concluded that this approach can be a promising tool for the study of phylogenetic relationships, especially for resolving higher order relationships.

Introduction

The evolution of plant genomes is determined by a combination of four basically different processes (Coghlan et al. 2005). Firstly, plants can scale up and down their ploidy level, often in response to natural hybridization between taxa. Such changes mostly involve complete genomes (polyploidy), but can also lead to deletion or addition of single chromosomes (aneuploidy). Secondly, plant species can dramatically differ in their amounts of nuclear DNA, in spite of their similarity in genetic complexity. This phenomenon, known as the C-value paradox, can be explained by huge dynamic changes in different classes of tandem and dispersed repeats during the formation of new species and hybrids. In most cases, such dramatic events are accompanied by the third major counterpart of genome evolution, the resetting of epigenetic states of (part of the) genomes. This can be generated by genomic stress in newly formed hybrids, or sudden large-scale demethylation of the DNA and changes in the histone code, leading to transitory reactivating of previously silenced transposable and retro elements, and transpositions of genes, regulators, promoters and other DNA sequences. The fourth process, which is the subject of this chapter, includes chromosomal rearrangements such as paracentric inversions (inversions excluding the centromere) and pericentric inversions (inversions including the centromere), translocations, centric split and fusion, duplications and deletions of parts of chromosomes.

Chromosome rearrangements can be directly seen, or inferred from aberrant pairing of polytene chromosomes as was described for insects by Painter, (1933). In *Drosophila* and related Diptera insects, chromosome rearrangements can be characterized by the detailed and unique banding patterns of their polytene chromosomes. Such polytene chromosomes are rare in plants, and can be observed in the suspensor cells of few plants, such as wheat (Chojecki et al. 1986) and maize (Kowles et al. 1990), but are not informative for elucidating chromosome rearrangements as they lack parasynapsis between the homologues. A general applicable alternative of studying chromosome variants is chromosome painting, the staining by FISH of specific chromosomes using the pooled single copy sequences for each of the chromosomes as probes in the hybridization. Details about the chromosome technology are further outlined in the chapters 1 and 6 of this PhD thesis. Chromosome painting in *Drosophila* is possible now (Fuchs et al. 1998; Schubert et al. 2001), but does not replace the power of polytene banding patterns by which all chromosome rearrangements could be demonstrated in full detail.

The best application of chromosome painting in elucidating chromosome rearrangements was multicolour chromosome painting, where combinations of DNA from different chromosomal DNA libraries are used as hybridization probes. The technology allows higher resolution (few mega base pairs) than previous chromosome banding techniques (Liehr et al. 2002, chapters 1 and 6). Nowadays chromosome painting is the most powerful painting technique in FISH (Ferguson-Smith et al. 2005). In contrast to mammals where chromosomes can be sorted by flow cytometry on the basis of chromosome size and GC/AT ratio, plant chromosomes can-

Table 1. Table 1. Overview of chromosome rearrangements in the *Solanum* and *Capsicum* genera according to molecular genetic studies. Table is constructed presenting tomato as reference. a= (Doganlar et al., 2002); b= (Tanksley et al., 1992); c= (Bonierbale et al., 1988); d= (Livingstone et al., 1999); e= (Seah et al., 2004); f=(Canady et al., 2006); g= (Stamova and Chetelat, 2000); h= (Pertuzé et al., 2002); i= (van der Knaap et al., 2004); j= (Albrecht and Chetelat, 2009). * description of the chromosome part is according to Doganlar et al. (2002)

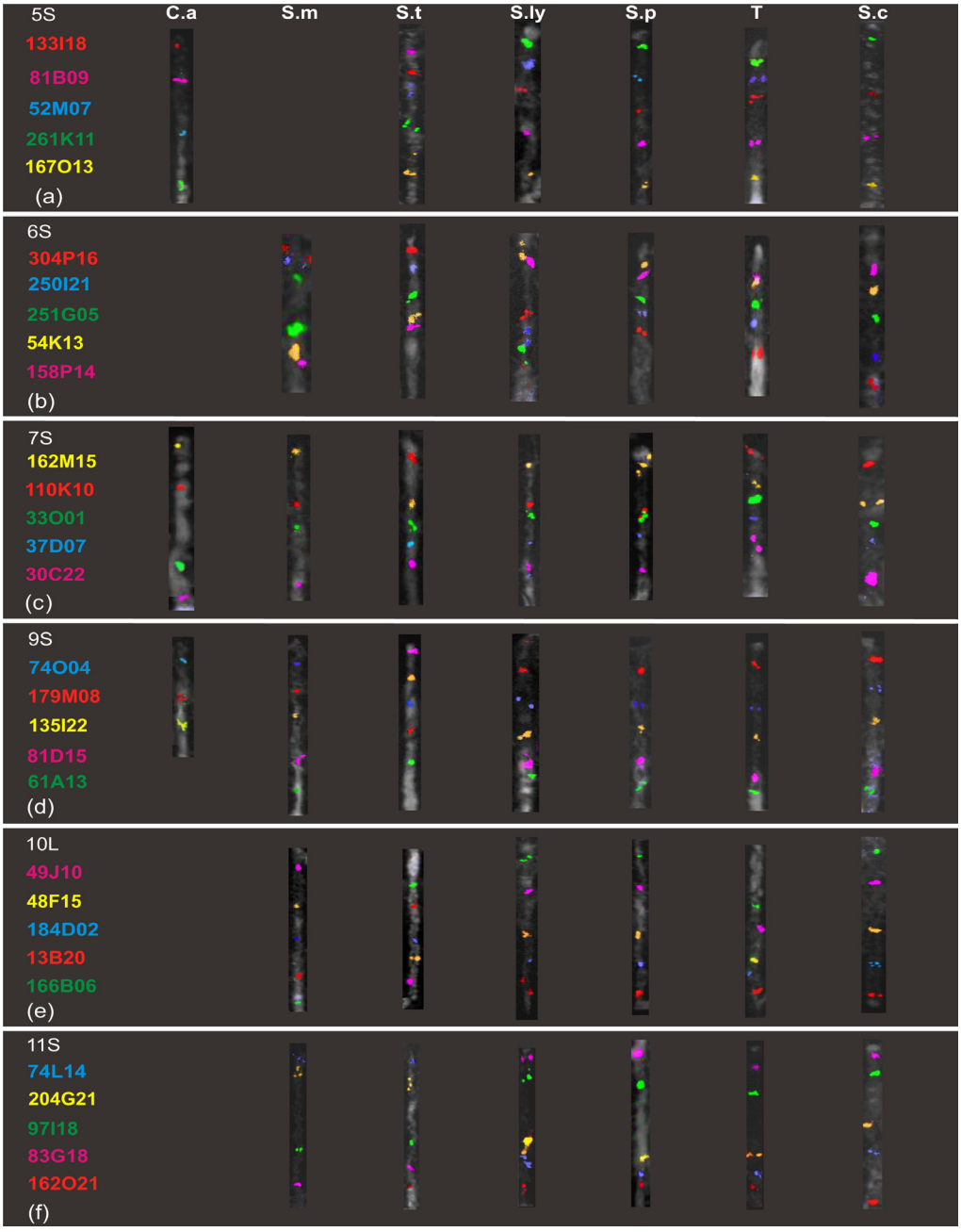
Chromosome Tomato (T)	*S. pennelli* LA716	*S. peruvianum* PI 128657	*S. chilense* LA0458	*S. sitiens* LA1974	*S. lycopersicoides* LA2951	*S. sect. jugandifolia* LA2788	*S. ochranthum*	Potato (P)	Eggplant (E)	Pepper (C)	References
5								5S inversion	5L+12L 5S and 12 L inversion	5S and 5L inversion	a, b, c, d
6		Mi-homologs inverted				6S inversion			Upper* 6 inversion	Collinear	a, f, e, j
7	7S inversion								Two inverted segments	Upper* 7 scattered	a, d, i
9								9S inversion	9 inversion	9 inversion Additional rearrangements	a, b, c
10				10L inversion	10L inversion	10S inversion		10L inversion	5S + 12S + 10L Lower* 10 inversion	Lower* 10 inversion	a, b, c, f, h
11								11S inversion	11S inversion Lower* 11 inversion 4S + 11 S	T11S = C12L 11S inversion (indication)	a, b
12			12S proximal inversion			Reciprocal translocation with 8S	Reciprocal translocation with 8S (cross with S. juglandifolium)	12S inversion	Upper T = E = C 12S + 11L Lower* 12 inversions with T and P	Upper* T = E = C Translocation 12S 11S	a, b, g, j

Table 2. Plant material used in this study.

Species / introgression lines	Accession number	Introgression source	Genome size (MB)	Taxonomy	Group
S. peruvianum	LA2172			subsection *Lycopersicon*	'tomato group'
S. peruvianum	LA2157				
S. habrochaites	G1.1290				
S. habrochaites	G1.1560				
S. pimpinellifolium	G1.1554				
S. glabratum	G1.1561				
S. chilense	LA1969				
S. pennellii LA716	LA716				
S. lycopersicoides				subsection *Lycopersicoides*	
Moneymaker (Mm)		None			
Motelle		*S. peruvianum*			
Vetomold		*S. pimpinellifolium*			
Moneyberg					
Zamir 7-4		*S. pennellii* LA716			
M82				Cultivated tomato material	
F1(Mm x Zamir 7-4)					
Heinz 1706			950		
S. tuberosum RH-89-039-16 (van der Voort et al., 1997)			850		
S. melongena (Half Lange Violette)			956		
C. annuum (Cayenne Pepper)			3000		

not be sorted or isolated, and even the isolation of specific chromosomes by microdissection will produce DNA libraries that hybridize all chromosomes. Hence, except in few cases, flow sorting in plants is not useful (Arumuganathan et al. 1994; Vrana et al. 2000). Even if unique chromosome selection would be possible, plant genomes carry large amounts of repetitive sequences, which are homogenized within a genome. This makes it impossible to use unique chromosome as a probe to identify rearrangements.

In plants, especially with small genomes, a better alternative for chromosome painting is the selection of single copy sequences by identifying BACs or other vectors containing DNA that are poor in repetitive sequences. *Arabidopsis thaliana* was the first plant for developing a chromosome painting technology accordingly. Pools of BACs were labeled with different colors and used as probes for hybridization on cell complements (Lysak et al. 2001; Schubert et al. 2001). Later, with the whole sequence of *A. thaliana*, FISH with BACs was applied in Arabidopsis and related Brassicaceae species to demonstrate chromosomal evolutionary processes and rearrangements (Fransz et al. 2000; Jackson et al. 2000; Lysak et al. 2006; Lysak et al. 2007; Lysak et al. 2005).


C.a
S.m
S.t
S.ly
S.p
T
S.c
5S
133I18
81B09
52M07
261K11
167O13
(a)
6S
304P16
250I21
251G05
54K13
158P14
(b)
7S
162M15
110K10
33O01
37D07
30C22
(c)
9S
74O04
179M08
135I22
81D15
61A13
(d)
10L
49J10
48F15
184D02
13B20
166B06
(e)
11S
74L14
204G21
97I18
83G18
162O21
(f)

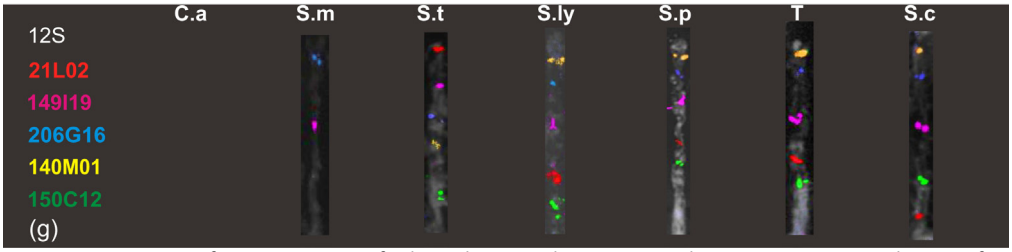


Figure 1. An overview of rearrangements of colours depicting their corresponding BAC positions as indicators of paracentric incersions. (a) 5S; (b) 6S; (c) 7S; (d) 9S; (e) 10L; (f) 11S; (g) 12S; C.a = *Capsicum annuum*, S.m = *S. melongena*, P = *S. tuberosum*, T = Tomato group, straightened chromosomes are: Heinz 1706 (image a, b, c, e, g), Moneymaker (image d, f), S.p = *S. pennellii* LA716, S.ly = *S. lycopersicoides*, S.ch = *S. chilense* LA1969.

Our research used essentially the same technology as for *Arabidopsis* for tomato and potato and other Solanaceae species, and focus on chromosomal rearrangements in wild Solanum species and crops. The Solanaceae crops represent one of the best-studied and attractive plant systems for genetics (Bonierbale et al. 1988; Doganlar et al. 2002; Fulton et al. 2002; Grube et al. 2000; Tanksley et al. 1992). Within this family, the largest and most diverse genus is Solanum containing 1400 species (Bohs, 2005). Gottschalk (1953) was the first who described the chromosome portraits of various Solanaceae using pachytene chromosome morphology. The basic karyotypes and major heterochromatin patterns in the *Solanum* genus are similar, and all have x=12 and their chromosomes exhibit large pericentromere regions (Khush and Rick, 1963; Ramanna and Prakken, 1967; Ramanna and Wagenvoort, 1976; Yeh and Peloquin, 1965). In contrast, telomere knobs were clearly observed in some species and absent in oth-

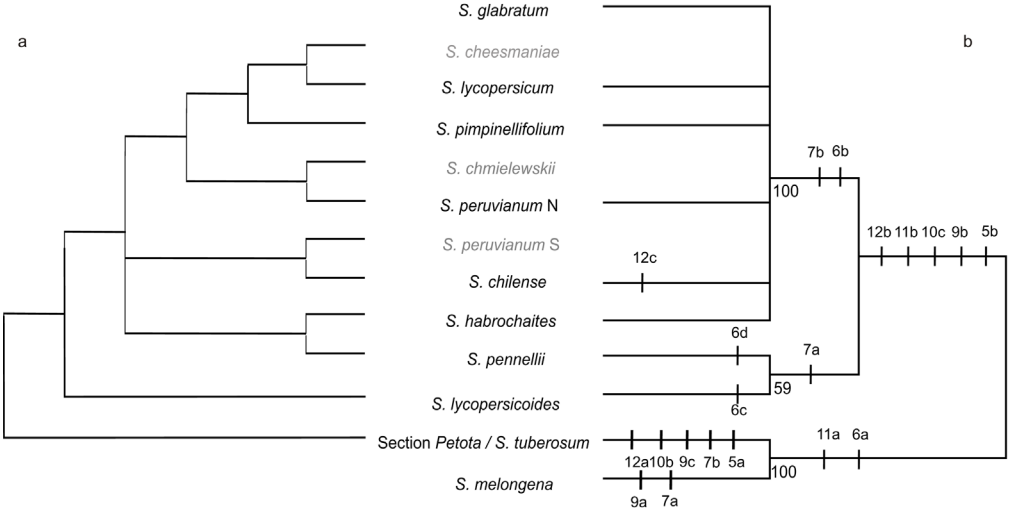


Figure 2. Comparison of a Solanum tree by Spooner et al., 2005 (a, left) with a cladogram based on chromosome inversion data (b, right). Character states in the cladogram (b) refer to characters presented in Table 4.

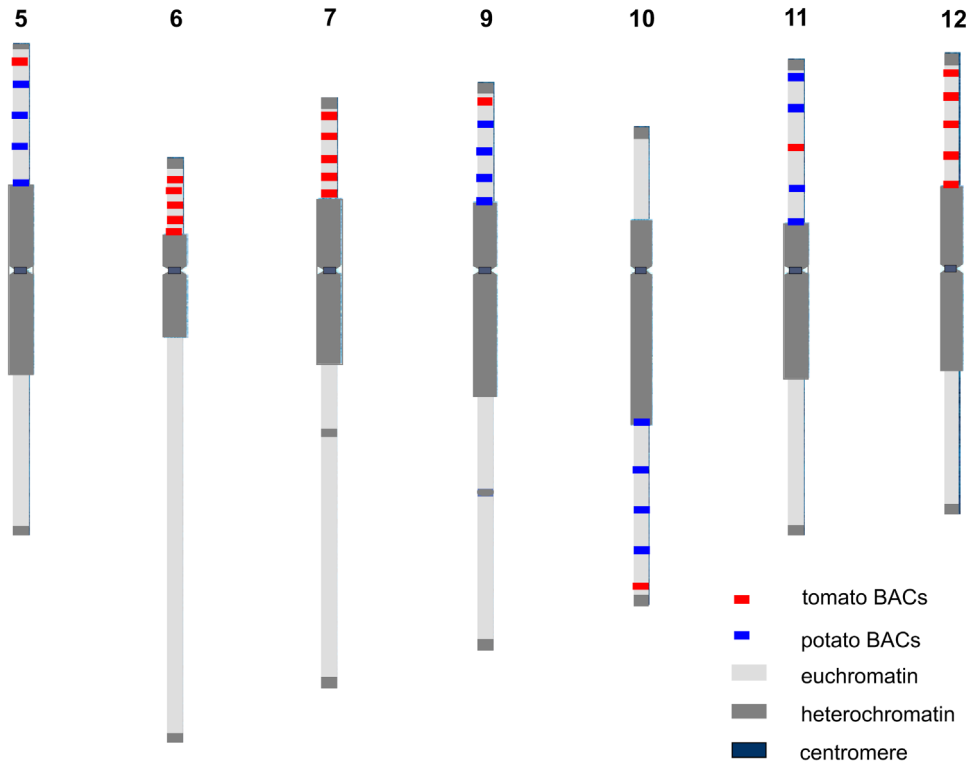


Figure 3. Chromosomal location of tomato (red) and potato (blue) BACs on the seven studied chromosome arms.

ers. Next to the *Solanum* genus in the Solanaceae family there are genera with chromosome numbers varying between x=7 (*Petunia*) and x=36 (*Atropa belladonna*) and there are huge differences in genome size and heterochromatin content as well.

Inversions are the most reported chromosomal rearrangements in the genera *Solanum* and *Capsicum* (see Table 1 for a summary and references for rearrangements). In *Solanum* most data on chromosomal rearrangements come from studies that used linkage maps based on molecular markers. Such linkage analyses are limited to areas where meiotic recombination takes place. If recombination is suppressed (*e.g.*, in heterochromatic regions or when rearrangements are present), rearrangements may remain unnoticed. Chromosomal rearrangements can also be directly studied in F1 hybrids using electron microscopic analyses of spread synaptonemal complexes (SCs). This technique is complex but allows a much higher resolution than light microscopy, giving superb details in aberrant chromosome pairing involving minor rearrangements between homoeologues. In somatic interspecific tomato (+) potato hybrids, for example, de Jong et al. (1993) described various aberrant pairing configurations strongly suggesting various translocations and inversions between the tomato and potato homoeologues. Recently, Anderson et al. (2010) used the same approach to show chromosomal

rearrangements between wild *Solanum* species by studying the pairing configurations of homoeologous in F1 hybrids between tomato and related wild relatives. They showed substantial changes of chromosome organization such as translocations, inversions and centromere mismatches.

Recently, it was shown that cross-species BAC FISH may provide a unique tool for synteny studies (Iovene et al. 2008; Tang et al. 2008). The study of Tang et al. (2008) used a five-colour FISH approach where different BACs of tomato and/or potato were labeled with different fluorescence dyes and then hybridized to cell complements of the other species. The sequences of the colours can then be used as markers to reveal chromosome colinearity. In the present study, we applied this technique to visualize chromosomal rearrangements within a range of *Solanum* species. We sampled wild representatives of section *Lycopersicon* subsection *Lycopersicon* as described by Spooner et al. (2005) plus cultivated tomato material and representatives of the crops potato, eggplant and pepper (Table 2). We focused on known inversions (Table 1) determining chromosomal synteny for a range of species.

Our main question was whether a multicolour BAC FISH can provide a valuable addition to existing tools for the study of rearrangements in crops and related species of the Solanaceae family in addition to comparative genetic mapping and chromosome pairing analysis of F1 hybrids. We also will explore and describe how the observed rearrangements can be used to describe evolutionary relationships in *Solanum*. Finally, we will discuss to what extent the used approach can be a tool in breeding of *Solanum* crops by supporting introgressive hybridization programs.

Materials and methods

Plant material

Young flowerbuds of eight accessions of seven wild tomato species, eight tomato breeding lines, one potato introgression line, one eggplant cultivar and one pepper cultivar (Table 2) were collected in the greenhouse in the morning. After fixing them for 1 day in fresh Carnoy solution (1:3=acetic acid:ethanol), the flower buds were transferred to 70 % ethanol for storage at + 4 °C. In total, 18 tomato and 17 potato BACs were selected covering seven chromosome arms (five BACs per chromosome arm). Their chromosomal locations are shown in Figure 3. The BACs were repeat poor except one (H146I19) on chromosome 12, which hybridized in the heterochromatin of several chromosomes. For chromosome 12 no repeat-poor BAC was available in the middle of the short arm euchromatin.

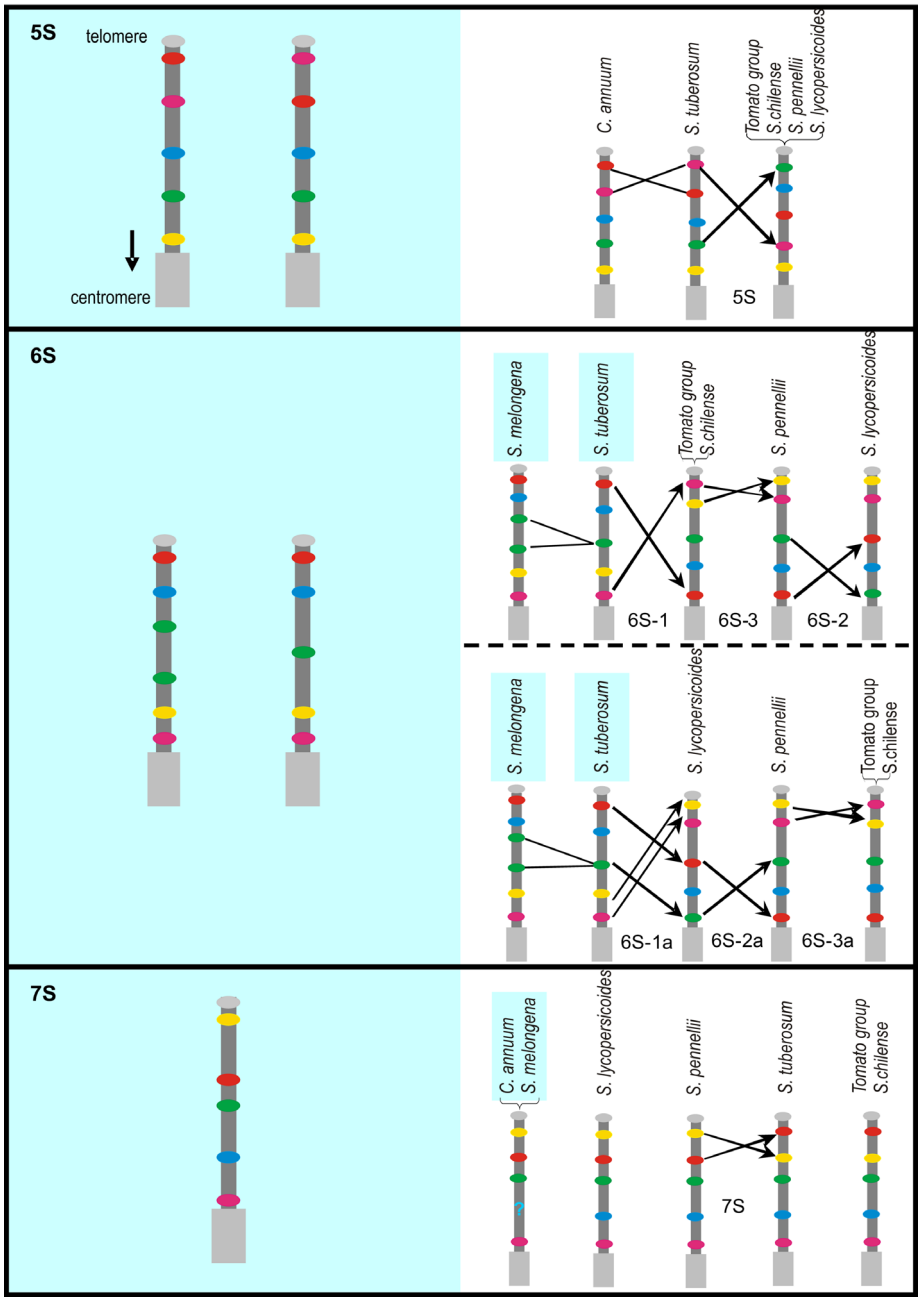


Figure 4. Schematic overview of rearrangements with reconstruction of possible origin. The left side of the figure shows the ancestral karyotype. The right side of the figure shows events of chromosomal rearrangements and their likely origin. Arrows indicate likely directions of inversions and a translocation.

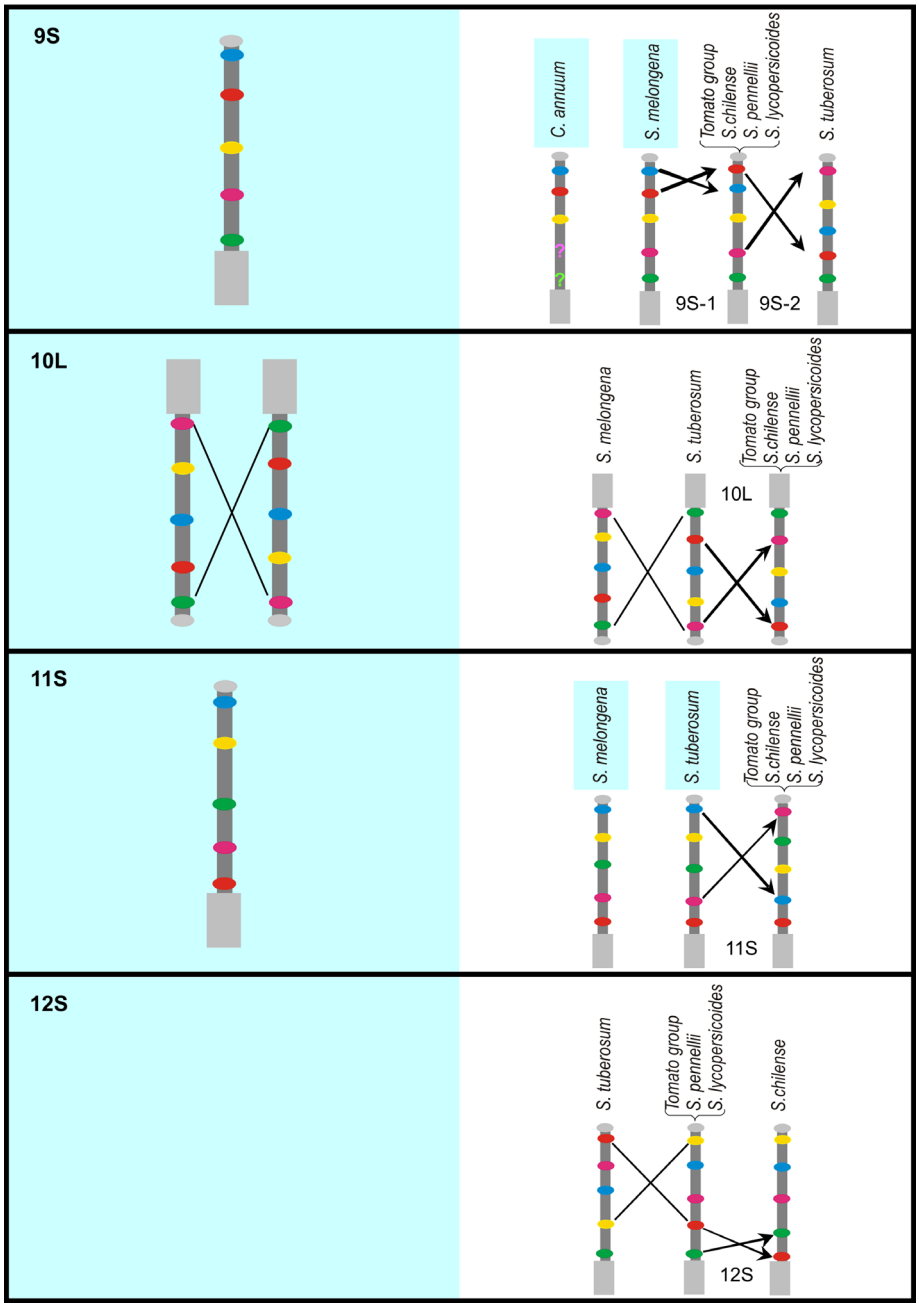


Fig. 4 (cont.). Below the arrows the chromosomes and the events are represented where more than one rearrangement occurred. Lines indicate inversion with unknown direction. Question marks are weak FISH results.

Slide preparation

Slides were prepared according to Szinay et al. (2008) with the following minor modifications. We adjusted the enzyme mix containing 1 % pectolyase Y23 (Sigma P-3026), 1 % cellulose RS (Yakult 203033, Yakult Pharmaceutica, Tokyo, Japan) and 1 % cytohelicase (Bio Sherpa 24970-014) to a five times higher concentration when anthers had thicker cell walls, such as potato, *S. pennellii*, and *S. lycopersicoides*, and three times higher concentration for eggplant compared to the tomato material. One anther was sufficient for two slide preparations for the wild species and the cultivated material of tomato and pepper, one anther of potato resulted in three slides and one anther of eggplant was sufficient for six slides. All slides were checked under the phase contrast microscope to check chromosome spreading and presence of cytoplasm on the nuclear material. A post-fixation step of 1% formaldehyde fixation was carried out except for the slides in which chromosomes were covered with cytoplasm.

BAC and Cot 100 isolation, and BAC DNA labeling

BACs were isolated as described in Szinay et al. (2008) with few modifications. In some of the cases we used High Pure Plasmid Isolation Kit (Roche 11754785001), because it was faster and cleaner compared to the classical alkaline isolation [see in Szinay et al. (2008)]. BAC DNA was labeled by nick translation following the manufacturer protocol by Roche (http://www.roche.com). We used the direct labeled dyes Cy3-dUTP (Amersham, http://www5.amershambiosciences.com/), Cy3.5-dCTP (Amersham) and Diethylaminocoumarin-5-dUTP (DEAC) (Perkin Elmer, http://www.perkinelmer.com), and two indirect labelling systems of using biotin - streptavidin conjugated with Cy5 and digoxigenin - antidig, conjugated to FITC. For combinatorial labelling we used equal amounts of biotin and digoxigenin haptens as probes. For painting BACs on chromosome 12, we used Cot 100 to block the repetitive signals that were present in the BAC (H146I19). Cot 100 was isolated according to Szinay et al. (2008).

FISH procedure, microscopy and data analysis

FISH was performed according to Szinay et al. (2008) with the following modifications. Hybridization was carried out for 2 or 3 overnight (Rens et al. 2006a, followed by a post-hybridization wash from 82 % to 64 % formamide at 42 °C (Schwarzacher and Heslop-Harrison, 2000) for 3x5 minutes. The biotin labeled probes were amplified three times for 45 minutes with Streptavidine-Cy5 and biotinylated-anti-streptavidin and the digoxigenin labeled probes were amplified twice with anti-digoxigenin-FITC and anti-sheep-FITC. Microscopy and FISH data interpretation were carried out as described by Szinay et al. (2008).

Table 4. Data matrix. First column: character numbers are indications of which chromosome is involved, letters are distinguishing different characters. Second column: order of the five fluorescent dyes on the chromosomes as independent characters

	plant material characters	*S. melongena*	*S. tuberosum* RH8903918	*S. lycopersicoides*	*S. pennelli* LA716	*S. peruvianum* LA2172	*S. peruvianum* LA2157	*S. habrochaites* G1.1590	*S. habrochaites* G1.1560	*S. pimpinellifolium* G1.1554	*S. glabratum* G1.1561	*S. chilense* LA1969	Heinz 1706 = *S. lycopersicum*
5b	green-blue-red-purple-orange	x	0	1	1	1	1	1	1	1	1	1	1
9b	red-blue-orange-purple-green	0	0	1	1	1	1	1	1	1	1	1	1
10c	red-blue-orange-purple-green	0	0	1	1	1	1	1	1	1	1	1	1
11b	purple-green-orange-blue-red	0	0	1	1	1	1	1	1	1	1	1	1
12b	orange-blue-purple-red-green	x	0	1	1	1	1	1	1	1	1	0	1
7b	red-orange-green-blue-purple	0	1	0	0	1	1	1	1	1	1	1	1
6b	purple-orange-green-blue-red	0	0	0	0	1	1	1	1	1	1	1	1
6a	red-blue-green-orange-purple	1	1	0	0	0	0	0	0	0	0	0	0
11a	blue-orange-green-purple-red	1	1	0	0	0	0	0	0	0	0	0	0
7a	orange-red-green-blue-purple	1	0	1	1	0	0	0	0	0	0	0	0
9a	blue-red-orange-purple-green	1	0	0	0	0	0	0	0	0	0	0	0
10a	green-red-blue-orange-purple	1	0	0	0	0	0	0	0	0	0	0	0
5a	purple-red-blue-green-orange	x	1	0	0	0	0	0	0	0	0	0	0
9c	purple-orange-blue-red-green	0	1	0	0	0	0	0	0	0	0	0	0
10c	purple-orange-blue-red-green	0	1	0	0	0	0	0	0	0	0	0	0
12a	red-purple-blue-orange-green	x	1	0	0	0	0	0	0	0	0	0	0
6d	orange-purple-red-blue-green	0	0	1	0	0	0	0	0	0	0	0	0
6c	orange-purple-green-blue-red	0	0	0	1	0	0	0	0	0	0	0	0
12c	orange-blue-purple-green-red	x	0	0	0	0	0	0	0	0	0	1	0

Phylogenetic analysis

A phylogenetic analysis was performed using the order of BAC FISH signals as presence/absence characters. Presence of a specific order was scored as 1 in the data matrix and absence 0. Each unique order of five BACs on a chromosome arm was considered as independent character (Table 4). MEGA4 (Tamura et al. 2007) was used to perform a parsimony analysis with 1000 bootstrap replicates (Figure 2b). The tomato introgression lines were excluded from our phylogenetic analysis, since these may possess non-natural character associations.

Interpreting chromosome evolution and inference of ancestral karyotypes

We determined ancestral karyotypes and the evolutionary history of the investigated chromosome arms by comparing the cladogram based on BAC FISH patterns with a consensus tree based on genetic markers from Bohs (2005) and Spooner et al. (2005). Ancestral karyotypes were determined by comparison of the most basal lineages (*C. annuum*, *S. melongena* and *S.*

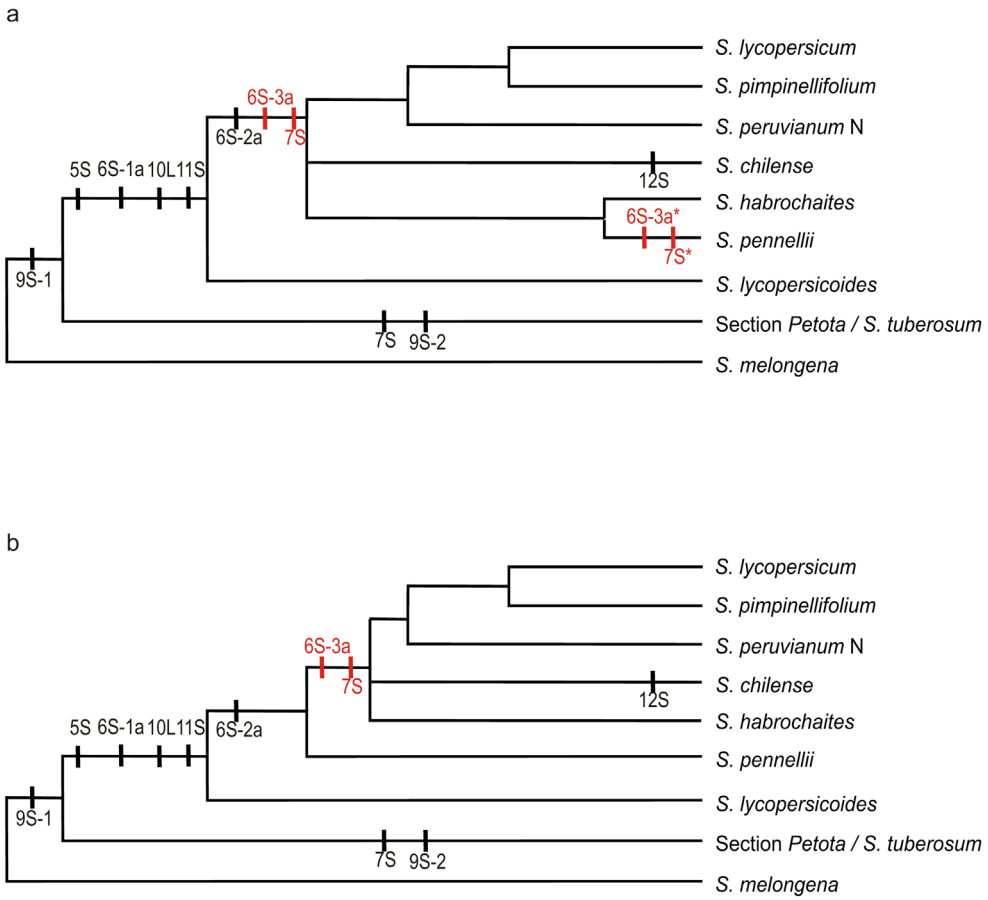


Figure 5. Simplified tree of Spooner et al. (2005, Fig. 9) and Bohs (2005), and a possible new tree topology with rearrangements indicated as presented in Figure 4.

a = Simplified tree after Spooner et al. (2005, Fig. 9; or see Figure 2 in this Chapter) and Bohs (2005). Species that were not sampled were omitted from the tree. "Section Petote" has been replaced with the taxon of that section we sampled: *S. tuberosum*, and *S. melongena* has been added as outgroup following Bohs (2005). The two rearrangements depicted in red (6S-3a and 7S) are shared by all species in the clade that branch leads to, except for *S. pennellii*. These characters therefore reversed (indicated by a *) in the branch of *S. pennellii*.

b = Alternative tree topology that is in better concordance with our chromosome data. The reversals in 6S-3a and 7S are resolved by placing *S. pennellii* as a sister to the clade including *S. habrochaites, S. chilense, S. peruvianum, S. pimpinellifolium* and *S. lycopersicum*.

Table 3. Chromosomal rearrangements that are not described in the literature. The dash (—) indicates comparison between species.

chromosome	rearrangement	between species
5S	distal inversion	*C. annuum* — *S. tuberosum*
6S	same BAC order, middle BAC split	*S. melongena* — *S. tuberosum*
	distal inversion	*S. pennellii, S. lycopersicoides* — tomato group, *S. chilense*
	proximal inversion	*S. lycopersicoides* — tomato group, *S. pennellii, S. chilense*
7S	distal inversion	tomato group, *S. chilense, S. tuberosum,* — *C. annuum, S. melongena, (S. pennellii), S. lycopersicoides*
9S	distal inversion	*S. melongena* — tomato group, *S. chilense, S. pennellii, S. lycopersicoides*

tuberosum). Later events could then be inferred by plotting the BAC order changes onto that tree (Figure 5a). We did this in such a way that the number of character state changes (*i.e.*, we minimized the number of inversion events/rearrangements) was minimized.

Results

New chromosomal rearrangements discovered by FISH

We investigated chromosomal rearrangements using BAC FISH on seven selected chromosome arms (Figure 1). Surprisingly, no differences were found between the cultivated tomato *S. lycopersicum* cv. Heinz 1706, and the accessions of *S. peruvianum, S. habrochaites, S. pimpinellifolium* and *S. glabratum* (i.e., the BAC order on the investigated chromosomes was identical, see Figure 1). In order to describe chromosomal rearrangements in a simple and clear way, this set of collinear *Solanum* species is therefore further collectively referred to as the 'tomato group'. Compared to this tomato group, a proximal inversion was detected on 12S of *S. chilense* (Figure 1g) and a distal inversion on 6S and 7S of *S. lycopersicoides* and *S. pennellii* (Figure 1b, c). Additionally, a proximal inversion is present in *S. lycopersicoides* (Figure 1b). *S. tuberosum* is similar to the tomato group and *S. chilense*, but different from all the other species by a proximal inversion on chromosome 7S (Figure 1c). All the tomato cultivars and introgression lines were similar to the tomato group on all chromosomes, except Zamir 7-4, which has a distal inversion on 7S identical to the one in *S. pennellii* (Zamir 7-4 is an introgression line derived from *S. pennellii*). Moreover, the F1 hybrid between Zamir 7-4 and Moneymaker revealed an inversion between the homoeologous chromosomes.

We discovered previously unknown chromosomal rearrangements comparing *Solanum* species and *Capsicum annuum* on the chromosome arms 5S, 6S, 7S and 9S (Table 3). In addition in one rearrangement, where the order of the BACs is the same on *S. tuberosum* and *S. melongena*, one BAC appeared as two signals on *S. melongena* (Figure 1b). This could be either

due to a deletion on S. tuberosum, for example with a breakpoint within this BAC, or an insertion on *S. melongena*, but it is also possible that there is a duplicated segment (of this BAC) in *S. melongena*.

Phylogenetic analysis

Three clades can be distinguished in the cladogram (Figure 2b). The first on groups *Solanum melongena* and *S. tuberosum* together with 100 % bootstrap support. A second clade, only weakly supported with 59 % bootstrap, joins *S. lycopersicoides* and *S. pennellii*. The third clade, also well supported (100 % bootstrap), comprises the members of the tomato group with *S. chilense*, forming a polytomy. Apparently, the utilized approach is unable to resolve relationships within the tomato group. The found rearrangements discriminate between clades rather than between species. Many character states are placed on the branch leading to the combination of the second and third clade described above. The exception to this is *S. tuberosum*, which has five character states on its branch. In Figure 2 our cladogram is compared to a tree by Spooner et al. (2005) based on AFLP / GBSSI / cpDNA / ITS data. Our tree results in a topology that is comparable but not identical to that of the tree by Spooner et al. (2005). Spooner et al. (2005) recovered a clade (the subsection *Lycopersicon*¸ also see Table 2) that is a sister to *S. lycopersicoides* (subsection *Lycopersicoides*). *S. pennellii* placed in the subsection *Lycopersicon* by Spooner et al. (2005) is not corroborated by our study, whereas *S. pennellii* is not included in the polytomy that we recovered in our study due to character states of chromosomes 6 and 7.

Determination of ancestral karyotypes and reconstruction the order of rearrangement events

Using the chromosomal rearrangements identified in this study (Figure 1) we endeavored to explain the course of chromosome evolution among the studied species. We first attempted to determine ancestral karyotypes (*i.e.*, ancestral BAC orders) for the different chromosomes. This was done by comparing the karyotype of the most basal lineages in the set of studied species in Figure 4. We used *S. melongena* as the ancestral karyotype based on Bohs (2005) and in some cases for conformation C. annuum was used that is outside of the Solanum.

An example of the derivation of the ancestral karyotype is the rearrangement observed in chromosome 11S (Figure 4). *S. melongena* and *S. tuberosum* share the same BAC order, and the BAC order of the other species can thus reliably be considered derived. Similarly, the ancestral karyotypes for 7S and 9S could be determined. A bit more complex is the situation in 5S in which the ancestral position of only the three proximal BACs can be determined. For 6S, 10L and 12S the ancestral karyotype is not equivocal.

Apart from ancestral karyotypes, we attempted to reconstruct the order in which rearrangements happened during evolution. Also here we made use of the *Solanum* tree published by Spooner et al. (2005) (Figure 5). Starting from the ancestral karyotype, we could indicate the most likely position of inversions on the branches of this tree. For example, one

inversion event of 5S is shared by all species in the tomato clade and by *S. lycopersicoides.* Since we may assume the ancestral karyotype is shared between *S. tuberosum*, *C. annuum* (and presumably *S. melongena*), the inversion event must be placed on the branch leading to the sister group of *S. tuberosum*.

The placement of inversion events of most chromosomes is straightforward (as presented in Figure 5a), but a few chromosomes showed surprises. Chromosome 6 was found to be the most variable chromosome among the sampled species. Reconstruction of inversion events for this chromosome can be done in various ways (two ways are shown in Figure 4; 6S). When these events are placed on the tree modified from Spooner et al. (2005), the second way (Figure 4, 6S, lower pane) that involves one complex rearrangement involving proximal-distal translocations (6S-1a/6S-2a/6S-3a), appears more plausible. Another surprise is that the one inversion (7S) apparently occurred twice independently in two lineages: in the lineage leading to *S. tuberosum* and after the split between the tomato group and *S. lycopersicoides*. Due to the surprising events on 6S and 7S we proposed a solution placing our inversion events on the tree based on Spooner et al. (2005) for a more parsimonious way.

Discussion

The fixation of a chromosomal rearrangement is a rare event in evolution, because it usually suppresses fertility or results in lethal progeny. In the *Solanum* genus paracentric inversions have been reported as dominant events for chromosomal rearrangements (summarized in Table 1). Moreover, due to difficulties of detecting especially pericentric inversions many other inversions might have remained undetected. By using multicolor BAC FISH, we could specify that nine inversions out of thirteen are paracentric and the putative break points of both sides of the nine inversions are specified between two BACs. Most likely all of the found inversions are paracentric although the proximal breakpoints are not definable. We not only confirmed previously described chromosomal rearrangements on the selected chromosomes and *Solanum* accessions but also identified five new, unknown inversions. Overall, our results demonstrate that cross-species BAC FISH is a powerful and reliable tool to study chromosomal colinearity in *Solanum*.

Chromosomal rearrangements among accessions within one species

Discrepancies can be observed between the results obtained in this and previous studies. For example, Doganlar et al. (2002) reported colinearity between *S. tuberosum* and *S. melongena* on chromosome 9S and 10L by using genetic markers, whereas our study showed multiple rearrangements on these arms between the two species (Figure 1 and 4). We identified the presence of a large inversion comprising most of the euchromatic part of the long arm on 10L between *S. lycopersicoides* and *S. tuberosum* (Figure 1). However Pertuzé et al. (2002) described

marker synteny on 10L for these two species. Moreover, Chetelat et al. (2000) showed the highest suppression of recombination on 10L between *S. lycopersicum* and *S. lycopersicoides*, suggesting chromosomal rearrangements, while we found that *S. lycopersicum* and *S. lycopersicoides* are colinear. A possible explanation for these differences is that we used different accessions (Canady et al. 2006) than in the other studies. Chromosomal rearrangements may be more readily fixed in cultivated species where selection pressure may differ from selection favored in the wild. This could explain the presence of karyotype differences between accessions of cultivated potato. Another possibility is that the genetic studies misinterpreted the inversions due to low marker density or suppression of recombination.

Phylogenetic inference using chromosomal rearrangements

Our phylogenetic analysis using MEGA4 produced a topology comparable to the phylogeny recovered by Spooner et al. (2005), but chromosomal rearrangements are best used to explain higher order phylogenies, and have little power in discriminating species. This observation poses a nice contrast to most published phylogenies based on genetic markers that have difficulties in resolving higher order relationships (Asamizu and Ezura, 2009; Bohs, 2005; Bohs and Olmstead, 1997; Levin et al. 2009; Tam et al. 2005; Wang et al. 2008).

Plotting our data on the tree published by Bohs (2005) and Spooner et al. (2005) showed that the inversions we recovered are most likely excellent witnesses of evolutionary history. All inversions we recovered could be placed on the earlier phylogenetic tree, and this brought no large inconsistencies to light. Our data strongly suggest that a complex rearrangement occurred on chromosome 6 after the split between S. tuberosum and the clade including S. lycopersicoides. If our interpretations are correct, then rearrangements are spread throughout the evolutionary tree, but they occurred mainly along the branch leading to the sister group of S. tuberosum, the clade including *S. lycopersicoides*, *S. pennellii* and all members of our tomato group. If the observed chromosomal rearrangements are all independent events, then the evolutionary time between the moment that the S. tuberosum lineage split from the *S. lycopersicoides* / *S. pennellii* / tomato group clade is much longer than the split between *S. lycopersicoides* and the other species in this clade. Alternatively, rearrangements may have accumulated quickly on this branch. This would then also include the complex rearrangement that allegedly occurred on chromosome 6 (see Figure 4 and 5). Our analysis also pointed to the probability that *S. pennellii* should be regarded the sister of the clade including the tomato group, *S. chilense* and *S. habrochaites*. Evidence for placing *S. pennellii* as sister to a clade including *S. habrochaites* and tomato comes from chromosomes 6 and 7 (see Figure 5b).

Implication of BAC FISH in breeding hybridization

Chromosomal rearrangements between species and between accessions of a species have great implications on the success of introgression breeding. In breeding programs, interspe-

cific crosses are very important to introduce valuable traits from wild relatives to crops. For example, almost all resistance genes in cultivated tomato have been introgressed from its 12 wild relatives (Bai and Lindhout, 2007). In interspecific crosses, suppression in recombination frequency is often encountered which results in linkage drag and may have resulted from chromosomal rearrangements. Assuming existence of chromosomal rearrangements in related wild tomato species, meiosis in F1 hybrids between cultivated tomato and such a related species may give rise to numerous smaller distortions in the distributions of crossovers along the chromosomes, and therefore, genetic maps are unreliable in regions that cause meiotic abnormalities or segregation distortions and/or recombination suppression. For example, the distal inversion between the tomato group and *S. pennellii* could explain the fact that, on the genetic map between *S. lycopersicum* and *S. pennellii*, hardly any molecular markers could be identified for the most distal part in contrast to the clustering of many markers at the genetic distance of 5 cM. Other evidence has been accumulated that chromosomal rearrangements may exist between cultivated tomato and some of its wild relatives (Liharska et al. 1996). Moreover, discrepancies have been observed between the actual chromosomal positions of BACs and their associated markers on the linkage group, which were most notable in the pericentromere heterochromatin where crossovers were almost absent (Sherman and Stack, 1995; Szinay et al. 2008).

The tomato breeding line Zamir 7-4 was included in our analysis, which was known to carry an introgressed segment of *S. pennellii* on chromosome 7. The presence of this introgressed segment was confirmed by our research by the presence of an inversion on 7S that is not present in other investigated tomato breeding lines. Our approach provides good evidence that cross-species multicolor BAC FISH is a reliable tool to study introgressed inverted regions. BAC FISH could therefore provide a powerful tool in verification whether chromosomal rearrangements play a role in suppression in recombination, and in explaining disagreements between the genetic and physical location of molecular markers and genes. Even though the FISH technique is undoubtedly useful to show chromosomal rearrangements between closely related species, the technique needs some adaptation to increase the resolution, for example more BACs should be included. An other powerful approach would be using electron microscope to investigate SC of F1 hybrids as Anderson et al. (2010) described, and then confirm the found rearrangements by BAC FISH.

Future perspectives

We have shown that chromosomal rearrangements can be used in phylogenetic studies in *Solanum*. The phylogenetic analysis in which we coded the different BAC orders as different characters provided some basic insights into the phylogenetic relations. Our method was not suitable to resolve phylogenetic relationship between closely related species as it is shown by the polytomy within the tomato group (Figure 2b). Divergence of the tomato group, *S. chilense*, *S. pennellii* and *S. lycopersicoides* is recent (~ few million years), because the chance

to find large chromosome rearrangements that are visible by FISH is low. Phylogenetic relationship between more distantly related species can be resolved by FISH, as was indicated by our analysis.

The weakness of studying phylogenetic relationships by FISH is that the polarity of character states is not always obvious or easily interpretable. Designating which state is plesiomorphic we believe will be more straightforward when we investigate more distantly related species. Therefore, it is important to include more species in the analysis. Most of the chromosomal rearrangements were found between *S. tuberosum* and the clade including *S. lycopersicoides*. Including more species between the mentioned species would of course provide better understanding of whether all the rearrangements happened once, or separately. Unfortunately, there are no species known that are – phylogenetically – in between *S. tuberosum* and *S. lycopersicoides*. We are therefore extending the research now into species that are phylogenetically in between *S. lycopersicoides* and the tomato group (representatives of section *Juglandifolium*). This should give us better understanding and confidence in interpreting the order of inversions in *Solanum* evolution, and could perhaps shed more light on the affinities of *S. pennellii* with its close relatives.

We will also extend our study to include more representatives of section *Petota* (the wild relatives of the potato) and a number of more distantly related *Solanum* species (section *Etuberosum*) that are phylogenetically in between *S. melongena* and *S. tuberosum*. This should – together with additional data on Capsicum annuum – better clarify the ancestral karyotypes for the different chromosomes. We aim to produce BAC-synteny-maps which will provide the basis for genome-wide microsynteny studies in order to transfer the tomato/potato sequences to other crops in the same family, as well as the basis for breeding strategy to introgress genes from wild species of Solanaceae to cultivated species.

Chapter 6

General Discussions: FISH applications for genomics and plant breeding strategies in tomato and other Solanaceous crops

Dóra Szinay[1,2], Yuling Bai[2], Richard Visser[2] and Hans de Jong[1,*]

1. Wageningen University and Research Centre (WUR), Laboratory of Genetics
2. WUR, Laboratory of Plant Breeding

* Author for correspondence:
WUR, Laboratory of Genetics,
P.O. Box 309
6700 AH Wageningen
The Netherlands
e-mail: hans.dejong@wur.nl

Accepted for publication in Cytogenetics and Genome Research 2010 (Themed Issue on Plant Cytogenetics)

Abstract

This paper describes the use of advanced Fluorescence *in situ* Hybridization (FISH) technologies for genomics and breeding of tomato and related *Solanum* species. The first part deals with the major determinants of FISH technology: 1) spatial resolution, which depends on the diffraction limit of the microscope and the type of chromosome, chromatin or isolated DNA fibres as target for the hybridisation; 2) the detection sensitivity, which is limited by the sensitivity and dynamic range of the CCD camera and the quality of the microscope, and the amplification system of the weak and tiny probe molecules; 3) simultaneous detection of multiple probes labelled directly or indirectly with up to five different fluorophores, whether or not in different combinations and/or mixed at different ratio. The power and usability of such multicolour FISH is indispensable when large numbers of bacterial artificial chromosomes (BACs) or other vectors with genomic DNA are available. Mapping of multiple BACs on chromosomes are powerful instruments confirming their assumed genetic position, whereas pooled BACs for a given chromosome arm will reveal the gaps between the BACs or derived contigs of their physical maps. Tandem and dispersed repeats, which are abundant in the genomes of most species, can be analysed in repeat bar coding FISH showing the major blocks of repeats in heterochromatin and euchromatin areas. Repeat-rich areas of the chromosomes can also be demonstrated by hybridisation of probed Cot fractions of sheared genomic DNA, a powerful method to elucidate the heterochromatin domains for genomic studies. In addition, unlabelled Cot DNA as blocking in BAC FISH painting suppresses repetitive sequences from the BACs to hybridise on the chromosomes. Cross-species BAC FISH painting with labelled probes from tomato and potato BACs and hybridised on the chromosomes of related species under appropriate conditions is a powerful instrument to demonstrate chromosomal rearrangements including inversions and translocations. The technology not only supports phylogenetic studies between the taxa under study, but can also be helpful in breeding programs with crops containing introgressed regions from related species when linkage drag or meiotic pairing disturbances between the homoeologues are assumed. In the next steps in comparative genomics, we now can detect smaller chromosomal and DNA rearrangements, diminutions and amplifications of repeats and changes of the epigenetic status of introgressed regions.

Tomato and potato, the two major crops of the Solanaceae family, have highly related genetic maps, chromosome portraits and genomes, which make them outstanding models for a comparative analysis of their genomes. International consortia for both crops were established for joining efforts in generating high-quality physical maps by building ordered Bacterial Artificial Chromosome (BAC) contigs of the euchromatin regions of all chromosome arms as well as whole-genome shotgun sequencing techniques (Mueller et al. 2009; Visser et al. 2009, and references therein). The genomes of tomato (980 MB) and potato (840 MB) are much larger than that of *Arabidopsis thaliana* (157 MB), due to the much larger proportion of repetitive sequences. Their sequencing in spite of the plethora of advanced molecular tools, has become a grand force with numerous technical and bioinformatical challenges. Fluorescence *in situ* Hybridisation (FISH) mapping of single copy and repetitive sequences are complementary tools to understand the complexity of these genomes. Most of this chromosome mapping is based on high-resolution FISH on pachytene complements as a target, which not only demonstrates positions of repeats and single copy sequences at high spatial resolution, without loosing the structural integrity of an intact chromosome, but also shows clear differentiation of euchromatin and heterochromatin blocks, which are most obvious in tomato pachytene chromosomes (Ramanna and Prakken, 1967). Using Cot-painting on pachytene chromosomes, Peterson et al. (1998) and Chang et al. (2008) showed that heterochromatin regions are rich in repetitive sequences. In addition, chromosomes at pachytene display homologous pairing and so can inform the researcher about any structural aberration, lack of homology or impairment in homologous pairing and recombination (Xu et al. 2000). FISH can generate more detailed information concerning chromosomal rearrangements or pairing failures, comparing to other cytogenetic methods such as chromosome banding techniques (Gill and Kimber, 1974; Appels et al. 1978; Berg and Greilhuber 1993; Schubert and Rieger, 1979) or synaptonemal complex analyses (e.g., Sherman and Stack, 1992; Albini and Jones, 1987) that have been used in previous years.

Quality factors of FISH

As discussed in various papers in chromosome painting technology, FISH has a couple of important technical aspects that determine the final quality of the microscopic images. The first is spatial resolution, which is defined as the distance between two adjacent fluorescent foci that a light microscope can resolve. If we leave more advanced and expensive microscope systems such as STED (Stimulated-emission-depletion fluorescence microscopy, Hell and Wichman, 1994), 4Pi (Fernández-Suárez and Ting, 2008) and SIM (Structured illumination microscopes, Schermelleh et al. 2008) aside, then spatial resolution in high-quality fluorescence microscopy is solely determined by the diffraction limit of the optics, which amounts to 0.2 µm for a 1.4 NA objective. With that in mind it follows that plant species have a spatial resolution along the linear condensed chromosomes in the range of tens of kilobases to several Mb depending on the stage of mitotic or meiotic division and the amount of heterochromatin (de Jong et al. 1999). Chromosomes in large genome species such as maize, wheat and lily are rich

in highly condensed heterochromatin containing long stretches of repetitive sequences, and hence they will have a proportionally lower spatial resolution, despite their huge lengths at pachytene, whereas in small genome species such as *Arabidopsis*, rice and cucumber, species with relatively clean euchromatin spatial resolution can be in the range of tens to hundreds of kilobases. Much higher resolution can be obtained with stretched chromatin (Valárik et al. 2004; Koo and Jiang 2009), extended fibres (Fransz et al. 1996) and spread DNA molecules (Jackson et al. 1999; Bensimon et al. 1994), claiming resolution values of about 1 kb, but all suffer from lacking the native chromosome structure.

The second important factor in FISH is detection sensitivity, which describes the minimum target size that a high sensitivity CCD camera in a fluorescence microscope can visualise. With optimised protocols it is nowadays possible to detect chromosomal targets of less than 110 kb (like small single copy sequences in pericentromere heterochromatic areas or cross-species FISH of tomato or potato sequences on distant Solanaceous species), whereas advanced probe amplification protocols like Tyramid- (Khrustaleva and Kik, 2001), PNA-, Padlock-based and Rolling Circle DNA Amplification protocols developed for human applications (Nilsson et al. 1997, reviewed in Tanke et al. 2005; Zhong et al. 2001), have not yet been implemented for tomatoes and other plants.

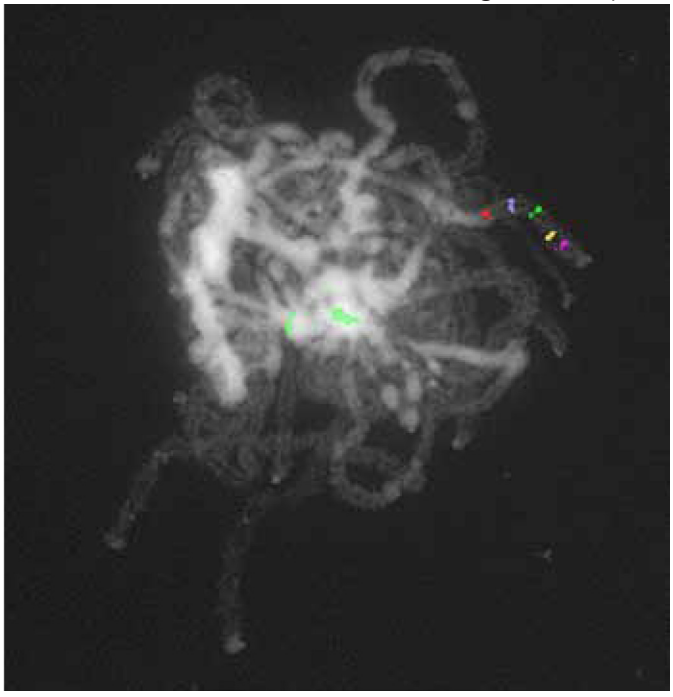


Fig. 1. Example of a 5-colour BAC FISH painting on tomato (*Solanum esculentum* cv. Money-maker). The BACs, labelled with DEAC (blue), FITC (green), Cy3 (orange), Cy3.5 (red) and Cy5 (purple) were hybridized on cell spread preparations containing pachytene complements. Details on probe labelling, FISH technique, microscopy and image processing are described in Szinay et al., 2008).

The third major hallmark of modern FISH is the simultaneous use of five different fluorophores in the same FISH experiment (5-colour FISH), either in a direct or indirect detection system. For tomato and other Solanaceae we optimised the detection of DNA sequences directly and indirectly labelled with nucleotides containing DEAC (blue excitation), FITC (green), Cy3 (orange), Cy3.5 (red) or Cy5 (far-red), together with DAPI as a DNA-specific counterstain. An example of the 5-colour FISH is shown in Fig. 1 and the technology is discussed in Birchler et al. (2008), Szinay et al. (2008) and Tang et al. (2009). In the case of small target sequences we used the indirect detection system of

biotin and digoxigenin haptens and with the streptavidin and anti-dig signal amplifications conjugated with different fluorochromes. Direct and indirect labelling each have their advantages and drawbacks. Direct labelling generally for chromosomal targets of 30 kb and more may give weaker signals, but background and precipitation of fluorescing particles is low, whereas indirect labelling can detect far targets as small as few kb, but may also easier show aspecific background fluorescence. The major advantage of this multicolour FISH is the simultaneous detection of five different probes, which means higher efficiency and less change for erratic interpretations, and hence more reliable chromosomal mapping. As the number of fluorescence filters is limited to the bandwidth of excitation and emission spectra of the filters never more than eight filters (the filters that we use + Cy5.5 and Cy7) can be used in the UV – visible light – infrared spectrum. For even more colours researchers have developed the so called combinatorial labelling, in which sequences are labelled with combinations of different fluorochromes, or with mixtures in different proportions (ratio labelling), or a combination of both (reviews in Fauth and Speicher, 2001; Geigl et al. 2006; Raap et al. 2006). All these applications that also require special advanced image processing software, have so far only be adapted to human and mammalian multi-colour FISH protocols (Rens et al. 2006b, Wiegant et al. 2000, Szuhai and Tanke, 2006).

Chromosome identification with BACs and repeats mapping is in most plant species difficult due to insufficient differentiation in morphology between the chromosomes. If the arm lengths and centromere positions and heterochromatin patterns are not enough for unequivocal identification, additional banding or FISH markers are required. The best method is using single copy sequences in BACs as FISH markers, as was shown for tomato chromosomes 1 (Chang et al. 2007), chromosome 2 (Koo et al. 2008), and the chromosomes 4, 6, 7, 9 and 12 (Table 1, see below). Other examples are from rice (Cheng et al. 2001), and potato (Dong et al 2000, Tang et al. 2009). Repeat bar coding with several repeats will work, but not so in tomato as most of repeats colocalise in the pericentromeres (Chang et al. 2008). More successful applications with combinations of microsatellites, rDNA, telomere sequences, NOR or centromere satellites as probes have been demonstrated for maize (Kato et al. 2004) and potato (Dong et al. 2000). In interspecific hybrids, the BAC- and repeat FISH chromosome identification can well be combined with genome painting, as was demonstrated for *Brassica napus* where labelled C genome and unlabelled A genome and 45S rDNA were used (Howell et al. 2008).

Table 1. Overview FISH mapping of tomato BACs for the chromosomes 4 (UK), 6 (NL), 7 (FR), 9 (ES) and 12 (IT). Data from Szinay et al. (2008); Peters et al. (2009); http://www.eu-sol.net/science.

chromosome	4	6	7	9	12
Total number of BACs used for FISH	78	102	60	22	51
FISH signal on expected chromosome	32	74	51	7	40
FISH signal on other chromosome(s)	46	25	8	14	11
Could not be mapped (repeat signals or empty)		3	1	1	-

Other FISH applications

The most common application of FISH in present-day genomics and plant genetics is the hybridisation of labelled DNA from large insert vectors like BACs and fosmids, with 50-150 kb of eukaryotic DNA, that produce in most cases clear single fluorescing foci on the chromosomes. But in plant species with moderate or large genomes, such vectors may contain large amounts of tandem or dispersed repeats, which hybridise on several or all chromosomes. Such repeat cross-hybridisation is especially noticeable in clones with DNA from distal and pericentromere heterochromatin areas. In our experiments we found that repeats could effectively be suppressed from hybridisation to the chromosomal target, when unlabelled Cot 100 DNA was added to the hybridisation mix (Peterson et al. 1998; 2002; Chang et al. 2008). In previous experiments it was shown that Cot 100 (10 – 100 x probe concentration) containing all pooled repetitive genomic tomato DNA was able to obtain single foci from a repeat-rich clone (Chang et al. 2008; Szinay et al. 2008). For the European partners of the European tomato genomics consortium a total of 313 BACs were mapped on the chromosomes 4 (Unites Kingdom), 6 (Netherlands, co-financed by the Dutch CBSG), 7 (France), 9 (Spain) and 12 (Italy). Table 1 gives an overview of the BAC FISH painting and shows how many of them could be mapped unequivocally according to their predicted genetic position, landed on different positions on other chromosomes, and which BACs demonstrate repeat signals, even in the presence of excessive Cot 100 blocking. The BACs that were confirmed for the expected genetic and chromosomal position served as seed BACs for subsequent contig building and DNA sequencing. Most of the BACs with deviating position or with multiple repeat signals were from chromosomal regions in the pericentromere heterochromatin, which is rich in repeats and devoid of crossover recombinations. On the tomato genome gaps of different sizes were found along the chromosomes, a phenomenon that will be discussed later.

Cot DNA has a second attractive application for genomics studies. As was shown in Chang et al. (2008), this pooled set of repeats with similar complexity can be labelled and hybridised on chromosomes to reveal the regions that are rich in repetitive sequences. Cot fractions with shorter reannealing time contain only the large tandem arrays, such as Cot 10 that hybridise to the distal heterochromatin blocks, whereas fractions obtained after longer reannealing (Cot 100) hybridised to the same distal and pericentromere heterochromatin segments (Chang et al. 2008). In an effort to demonstrate repeats of lower complexity, such as short stretches of transposable elements, Ty1/*Copia* and microsatellites, we also painted chromosomes with Cot fractions of even longer reannealing time (Cot 200, Cot 400 and Cot 1000). However, as shown in Figure 2, the overall distribution of repeats from different Cot fractions all hybridised to the distal blocks, the NORs and centromeres and pericentromeres. In an additional series of experiments, we also tested hybridisations of Cot DNA fractions at lower hybridisation stringency of 64% instead of the normal 82% (Schwarzacher and Heslop-Harrison, 2000), which showed brighter signals at chromosomal targets with small repeat segments (data not shown).

An additional strong feature of Cot DNA is a preselection of BACs by filter hybridisation. This was already pointed out by Chang et al. (2008) in their paper on tomato repeat organisation. The Cot 100 fraction that first was checked to paint all heterochromatin domains, was then used as probe in a hybridisation on BAC filters directly showing all clones that are rich, moderate or poor in repeat content. Their estimate of repeat content for the tomato genome based on analysing a set of 800 Heinz 1706 *Hind*III BACs revealed that tomato consists of 30% of repeats, a result that corresponds well with other estimates for highly repetitive DNA content of the tomato genome (Zamir and Tanksley, 1988; Ganal et al. 1988). Using one of the pericentromere or distal heterochromatin repeats such a BAC filter hybridization can even be more efficient to discover BACs from centromeres, pericentromeres, distal heterochromatin and NOR regions (Peterson et al. 2002; Yang et al. 2005).

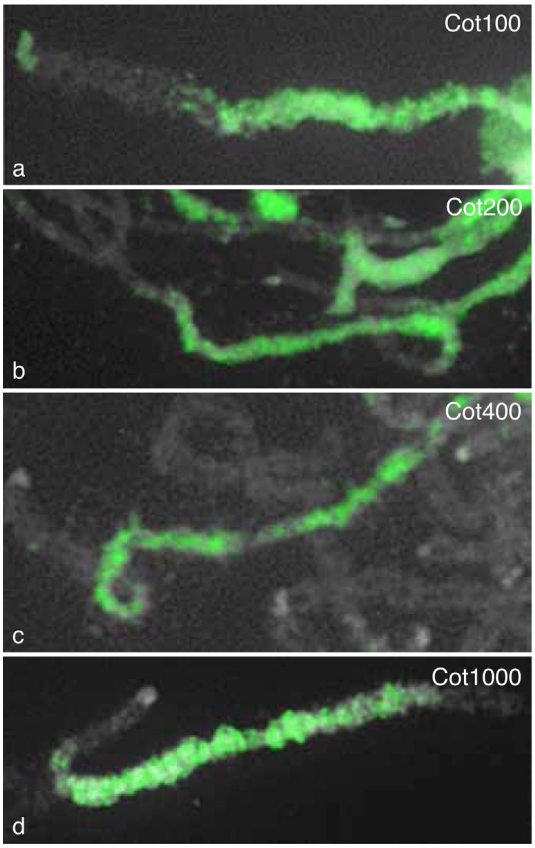


Fig. 2. Details of pachytene chromosomes of tomato (*S. lycopersicum* cv. Heinz 1706) painted with different Cot fractions of genomic tomato DNA. (a) Cot 100; (b) Cot 200; (c) Cot 400 and (d) Cot 1000. In the FISH with all Cot fractions the probe hybridizes to the heterochromatin segments of the chromosomes.

Finally, Cot technology can also be employed for genome complexity reduction. As Peterson et al. (2002) and Rabinowicz (2007) discussed, one of the ways to isolate single copy and low repetitive sequence genomic DNA is Hydroxyapatite (HAp) column chromatography, which is based on differential affinity of single and double strand DNA under specific phosphate buffer concentrations (Smith et al. 1975). Upon high temperature denaturation followed by long time reannealing the middle and high copy sequences that become double stranded can thus be separated from the single stranded low copy sequences. This method as a general strategy discussed in Paterson (2006) was successful in maize (Yuan et al. 2003) and *Sorghum* (Peterson et al. 2002), where the high Cot fraction could be applied to enrich low copy sequences, but in hexaploid wheat the method need further optimisation (Lamoureux et al. 2005).

Seed BAC distribution and contig gap size estimates

A major question in the tomato BAC-by-BAC sequencing projects was how much of the short and long arm euchromatin segments are covered with seed BACs and if large gaps still occur. In chromosome 6, the distribution of 15 BACs in the short arm and 40 BACs in the long arm euchromatin were analysed in a so called pooled BAC FISH experiment (Szinay et al.2008; Peters et al. 2009). Microscopic measurements of the FISH signals in the short euchromatin arm revealed a major gap of 0.70 µm close to the distal heterochromatin block. Using the average MB/µm ratio (Peters et al. 2009), gap was estimated at 1.59 MB. In the long arm three major gaps were found from centromere to telomere with MB/µm ratio amounted to 7.00, 1.17 and 1.44 Mb, respectively. These gaps together represent a total of 11.2 MB, which is 36 % of the short and long arm euchromatin.

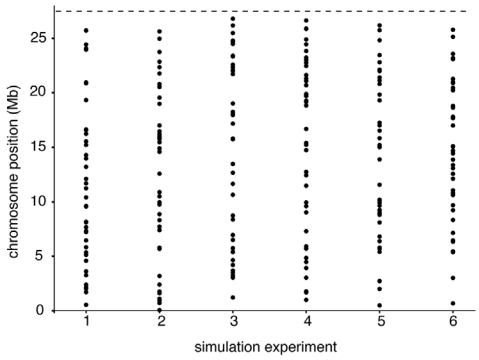


Fig. 3. Simulation experiment of BAC position along chromosomes by generating 40 randomly chosen BACs along a virtual euchromatic chromosome region. In this example the 26.9 MB tomato chromosome 6 long arm euchromatin was taken, where FISH experiment reveal gaps along the euchromatin.
The chart clearly show that big gaps between neighbour BACs can be found just by chance. Y-axis: chromosome 6 long arm euchromatin in Mb. X-axis: six sets of fourty random BACs.

Gap closure is one of the major challenges in physical mapping of eukaryotic genomes (Dolgin, 2009) and the challenge is proportionally bigger when genomes are larger by the presence of high amounts of repeats. Several reasons can be given for the gaps that we observed. Firstly, random positions of BACs may produce gaps as large as we found in chromosome 6 long arm euchromatin (Figure 3). In a simple spreadsheet simulation such a set of randomly distributed BACs can be easily tested by plotting 40 random points along a line, representing the 40 BACs on the 26.9 µm long arm euchromatin that we used for our FISH study (Figure 3). Sorting these dots and calculating their distances between neighbours in just six repetitions revealed gaps as large as 2.3 µm (8.6%), which is about the largest gap of 7 MB that were in our experimental data (Peters et al. 2009). In other words: large gaps can be explained if we only consider random distribution of the BACs. Secondly, genetic markers used to recruit the seed BACs may be excessive in some regions and absent in other, while crossover recombination hotspots and coldspots may also bias the distribution of markers, and so will be their seed BACs. The third major reason for large physical gaps is the lack of BACs in certain regions, due to the absence of restriction sites on which the BACs were obtained. The tomato sequencing consortium uses BAC libraries based on *Hind*III, *Mbo*I and *Eco*RI restriction enzymes (http://solgenomics.net/about/tomato_sequencing.pl), but is now complemented with

a new BAC library and a fosmid library, which both are based on sheared DNAs (Budiman et al. 2000). Adams et al. (2000), Eichler et al. (2004), Blakesley et al. (2004), Bovee et al. (2008) and Dolgin (2009) have discussed the significance of gap closures extensively. In *Drosophila* the gaps were closed by various sized cloned sequences (Adams et al. 2000) and in the human genome the fosmid library closed most of the gaps (Bovee et al. 2008).

Repeat BAR code painting in tomato

Complementary to the sequencing efforts is the mapping of major repeat classes on the chromosomal level. A concept of multicolour FISH mapping of various major classes of repeats was introduced by Schmidt and Heslop-Harrison (1998) as the so-called repeat bar code. Repeat bar codes can be complex and informative in large genome size species (wheat and rye), but are simple in small and medium sized plant genomes (such as those of *Arabidopsis*, *Brassica*, cucumber and tomato) where most of the retrotransposons, microsatellites occur in all pericentromeres and tandem repeats at the distal ends of the chromosomes. Homogenisation processes average the repeat content of most chromosomes, so that repeat bar codes are not always instrumental in chromosome identification. In tomato (Chang et al. 2008 and references therein) we described the global distribution of all repeats together (Cot painting) and several retrotransposons and tandem repeats apart. More details are now available on the chromosomal distribution and sequence data of the repeats on the tomato chromosomes 6 and 7. These data revealed that the tandem repeat TGRI that occurs on most of the chromosome ends (Zhong et al. 1996b) can be copied into interstitial segments of long arms of other chromosomes, likely by a mechanism of extrachromosomal circular DNA transposition. The other tomato repeats, TGRII, TGRIII and TGRIV are members of the Ty3/*Gypsy* family and occupy the pericentromere areas but has significantly different coverage (M. Philippot, personal communication). This amalgamation of molecular cytogenetics, sequencing and bioinformatics is mutually complementary, in the sense that detection and sensitivity by FISH is never as accurate as the sequencing itself, but it can give a rough indication of repeats along the chromosomes without a priori sequencing data. In addition, BACs with any of these diagnostic BACs, in combination or alone, and together with yet unknown repeat classes can now reveal chromosome positions with respect to telomeres, centromeres and NORs. Since the tomato BAC-for-BAC sequencing project focuses on the euchromatin part of the genome genetic and cytogenetic maps are indispensable for the physical map on tomato, otherwise misallocated BACs and even whole contigs would hamper the sequencing process.

The confusions of the heterochromatin concept

One of the most disputed terms in the sequencing of eukaryotic genomes is heterochromatin. Historically it was defined as the chromosomal regions that retain their condensed state throughout all cell cycle phases, in contrast to euchromatin that is decondensed in interphase

and condensed at metaphase (Heitz, 1932). Later on the concept of heterochromatin was further refined with Brown's distinction between constitutive (structural) and facultative (functional) heterochromatin (Brown, 1966), but with the advent of various chromosome banding technologies in the seventies it became clear that its definition was even far more complex and involved a mix of biochemical, physical and staining properties in different classes of chromatin. Nowadays with our acquired knowledge on molecular composition of chromatin, genetics and genomics, the term heterochromatin has become even fuzzier and controversial. Good overviews about the different types of heterochromatin and the oxymoron of genes and heterochromatin are given in Bennetzen (2000) and Yasuhara and Wakimoto (2006).

As to the interest of most researchers it is practical to define heterochromatin as regions that are relatively rich in repetitive sequences, either in tandem arrays or dispersed repeats such as transposable elements, and retrotransposons, SINES, LINES and microsatellites. However, in heterochromatin segments of several model species, it has been showed that heterochromatin does contain expressed hypothetical genes (Yasuhara et al. 2005; 2008). In tomato, Peters et al. (2009) showed that pericentromere BACs in large heterochromatin region of tomato chromosome 6 harbor an unexpectedly high gene content of one gene per 36.7 kb, whereas the short arm and long arm euchromatin gene content amounted one gene per 15.3 kb and 8.8 kb, respectively. In addition, the short arm euchromatin had an average repeat content of 13.4%, which is only 40% less than the clearly distinguishable pericentromere heterochromatin. These relatively small differences in genomic constitution between euchromatin and heterochromatin are not reflected by the conspicuous differences in DAPI fluorescence between the euchromatin and heterochromatic parts. But there are more interesting observations that point at more complex properties of the tomato heterochromatin. Firstly, for certain regions in the chromosomes, no extension BACs could be found in any of the available BAC libraries, such as in the very distal end of the short arm of chromosome 6 where extension BACs could not be found for BAC H016K14 that contain traces of the telomere associated repeat TGRI. Such distal gaps have also been reported for human (Eichler et al. 2004). A few plausible explanations can be given. Genetically, the very distal part of the chromosome is erratic by the lack of crossovers and possible minor rearrangements between the crossing parents of the linkage map (*S. lycopersicum* LA925 and *S. pennellii* LA716). Alternatively, gaps may be explained by the absence of restriction sites of enzymes used for the three tomato BAC libraries (*Hind*III, *Mbo*I and *Eco*RI) or that the small heterochromatic knob at the end of the chromosome that contains large copy numbers of the TGRI tandem repeat is too difficult for physical mapping.

Secondly, we postulate that small islands of heterochromatin, the so-called chromomeres are regions rich in middle repetitive sequences that may hamper tiling extension BACs in contig building. In contrast to *Arabidopsis*, where euchromatin is almost devoid of chromomeres, tomato has a large number of such chromomeres, which can easily be discerned in DAPI stained cell spreads without FISH. In Figure 4, the chromomeres of a straightened chromosome 6 are displayed using a pseudocolorization imageprocessing tool to enhance the contrast of

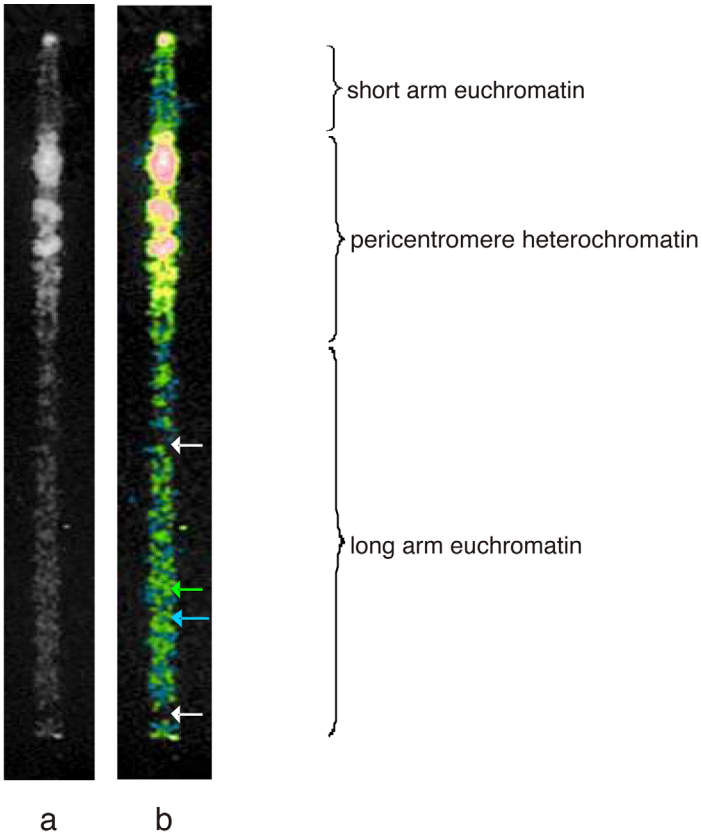


Fig. 4. Pseudocolouring of weakly stained chromomeres. (a) Straightened chromosome 6, DAPI stained; (b) Pseudocoloured chromosome 6. Green and blue colour show examples of weak DAPI fluorescence (chromomeres). White and orange areas are the relatively bright fluorescing pericentromere regions, the unstained regions areas the euchromatin gaps (white arrows).

these small local regions of enhanced chromatin density. A large number of these chromatin structures are obvious all along the long and short arms euchromatin. Although FISH probed with higher Cot fractions did not reveal any support for these chromomeres (Figure 2b, c, d), it is still not clear whether these local chromatin packages are domains of specific repetitive sequences.

Thirdly, the long arm interstitial heterochromatic knobs in various chromosomes are special cases of hyperdynamic repeat domains. As discussed in Zhong et al. (1996b) and Chang et al. (2008), such long arm interstitial heterochromatic knobs are island containing TGRI tandem arrays originating from one or more of the distal short arm heterochromatic knobs. We assume that such footprints arise from short distance transfer of TGRI sequences through a transposon directed rolling circles amplification producing extrachromosomal circular DNA molecules from the distal knob to the interstitial long arm target (Navrátilova et al. 2008;

Cohen et al. 2008). As short arm distal ends are aligned to long arm interstitial regions of the same chromosome and other chromosomes in what is known as Rabl or anaphase orientation (Rabl, 1885), we can explain formation of interstitial TGRI footprints by such short distance transpositions. We further postulate that such highly unstable sites of newly formed and rapidly disappearing TGRI footprints can explain the genetic problems correlated with mapping markers in these (40-45 cM) interstitial regions.

Heterochromatin is also unique in the sense that these regions are (almost entirely) devoid of meiotic crossovers. This was convincingly demonstrated for *Arabidopsis* (Copenhaver et al. 1998; Singer et al. 2006), tomato (Stack and Anderson 1986, Sherman and Stack 1992, Peterson et al. 1999), maize (Anderson et al. 2003; Stack et al. 2002) and other plants. For the construction of linkage maps it means that markers at either border site of heterochromatin will have no or very few recombinants, and hence all markers and genes in between these borders will occupy the same genetic position. Thus, introgression of economically important genes inside pericentromeres will always contain all other genes from the introgressed segment giving rise to the problem what breeders call linkage drag.

The final definition of heterochromatin deals with its epigenetic status and is probably the most universal for this type of chromatin. It describes that heterochromatin can be distinguished from its remaining euchromatin by methylation of cytosine in the DNA (5mC) and specific histone modification like H3K9met2 and H3K4me2 and absence of H3 acetylation. In an increasing number of studies it is indicated that the epigenetic status of chromatin is directly involved in the silencing and activation of transposable elements and genes and so plays a crucial role in the stabilization of the genome and transcription of specific genes. Once the larger part of the euchromatin of the tomato genome is sequenced, the next step will be the study of its epigenome, that allows understanding the mechanisms of alien chromosome segments containing important genes and introgressed into one of the tomato homoeologues become silent due to epigenetic reprogramming. One intriguing example is the silencing of the nematode resistance gene *Mi* from *S. peruvianum* on chromosome 6S, that became silenced after a few generation due to DNA methylation of that region (P. Zabel, personal communication).

SWOT analysis of molecular cytogenetics research

In this overview we have given various examples on the benefits and drawbacks of molecular cytogenetics for tomato. In the context of the huge genomics initiatives the question may arise: what significance does this research field still have. The following analysis gives an overview of the strengths, weaknesses, opportunities and threats, the so-called SWOT analysis.

The strengths include 1) BAC-FISH maps bridge the genetic and physical maps and so is intermediate in genomics to explains aberrant BAC positions, gaps between contigs etc.; 2) FISH technology reveals repeat rich regions on the chromosomal level, and makes size estimates of the repeat domains and distribution without sequencing; 3) BAC FISH can position single copy sequences in (pericentromere) heterochromatin regions where crossovers are lacking; 4) Can

elucidate aberrations in meiotic pairing recombination and deviations in synteny between species in region of interest without a priori laborious genetic mapping and 5) Combining chromosome data with chromatin immuno precipitation analyses (CHiP) can reveal the epigenetic status of chromatin regions. Weaknesses of FISH technology are 1) Molecular cytogenetics requires a long learning curve (genetics, microscopy, molecular technologies, image analysis and interpretation); 2) There are only a limited number of specialised laboratories world-wide; 3) Due to time consuming microscopic observations and interpretation, the technology is not suitable for high throughput analysis; 4) No information at sequence level and 5) High costs for advanced microscopes and fluorophores. Opportunities are 1) Unique contribution in plant genomics, plant genetics, taxonomy and breeding; 2) Crosslinks to epigenetics; 3) Increasing interest from commercial (breeding) companies for its contribution to gene (BAC) mapping in relation to suppression of crossover recombination, and finally, the threats are that microscopy and genetics 1) are not fashionable in modern biology, which results in difficulty to raise funds for this type of research; 2) depends on well trained personnel and 3) requires new optimization of techniques and protocols for every model and crop species.

Future perspectives

The final section deals with an example of applied molecular cytogenetics in support of an introgressive hybridisation program for geneticists and plant breeders in tomato and other crop species for which genomics tools and BAC libraries are available. The overview of the different steps in the procedure is explained in Figure 5. Introgressive hybridisation is an important strategy to broaden the genetic base of highly inbred crops like tomato and mostly deals with transferring economically important traits such as drought and salinity tolerance and disease resistance transferred from a related species or genus to the crop. Such transfers are often longstanding procedures starting with an interspecific or intergeneric hybridisation, followed by backcrossings with the recipient crop. When crossing barriers exist between the donor and recipient species, cross bridges or in vitro technologies including somatic hybridization and embryo rescue can be helpful in transferring genes between taxa which otherwise cannot hybridise sexually. In most cases backcross derivatives are selected on the basis of the selected trait, or with molecular markers genetically tightly linked to the gene(s) of interest. When the size and position of the introgressed region are under debate, and the species sufficiently different in terms of dispersed repeats, a genome painting can be used to establish the chromosome segments that were integrated into the chromosomes of the crop. Several successful examples of introgressive hybridisations are known for various interspecific and intergeneric hybrids, including wheat x rye (Sánchez-Morán et al. 1999), tomato (+) potato (Jacobsen et al. 2002; de Jong et al. 1993), sugarbeet x *B. corolliflora* (Desel et al. 2002).

A second important cytogenetic tool based on detection of single copy sequences in the region of interest and is possible when BACs are available for the region of interest. Thus, a FISH map can be constructed using BACs between related species, which provides a good

tool to study rearrangements and linearity between related species and to reveal pairing disturbances in plant breeding. Between tomato and potato seven paracentric inversions were described, either genetically or cytogenetically. Such inversions can arise with two breaking points within a chromosome arm, or with one breakage at the border of the eu- and heterochromatin, followed by a *de novo* development of a telomere. Within the *Solanum* genus mostly paracentric inversions were achieved through comparative genetic mapping between tomato and potato or between the crops and related wild species. Electron microscope analyses of spread synaptonemal complexes at late pachytene showed numerous unknown chromosome rearrangements in several *Solanum* F1 hybrids in euchromatin and heterochromatin that could not be detected in genetic mapping due to suppression of recombination (Anderson et al. 2010 in press.). The more distantly related eggplant and tomato showed few translocated segments (Doganlar et al. 2002). In the Brassicaceae family, which is an older plant family than the Solanaceae there are mostly translocated segments. Similar rearrangements within the Brassicaceae occurred independently, which suggested its existence in the ancestor (Lysak et al. 2005).

A major aspect of BAC FISH of introgression derivatives is the analysis of chromosome pairing of the introgressed region. This study will focus on the chromosome region that contains the gene of interest. BAC FISH painting will now directly reveal the synteny between the donor region and the homoeologous recipient part. If the study on meiosis of this hybrid reveals pairing failure, i.e., the two homoeologous regions are inverted, translocated or hemizygous for a duplication or deletion, crossovers cannot be formed in the chromosome region, and so cannot produce recombinant chromosomes containing the gene of interest, but with a minimum of “wild” donor sequences. Alternatively, if pairing occurs while forming a inversion loop the consequences of a single crossover will lead to a dicentric chromosome and acentric bridge in the case of a paracentric inversion, and hence leads to lethality of the formed gametes. In the case of a pericentric inversion crossovers in the inversion loop will result in deletion duplication chromatids and two normal chromatids, without elimination of the inverted region. Yet a paracentric inversion – and most inversions between tomato and other Solanum species are known paracentric inversions, may in the rare cases of specific gene conversion, or double crossovers and even somatic crossovers give rise to the transfer of the desired region from the donor species to the recipient crop. However, no records are known in which these events have been described. Still the research of BAC FISH painting across the introgression region can be of utmost importance as it can warn the breeder about absence of chromosomal collinearity between the homoeologous regions, and hence will explain the observed linkage drag. There are more chromosomal disturbances that are informative for the breeder. Chromosomal regions may accompany with losses of DNA sequences due to transposition, or may change its epigenetic status leading to loss and duplication of smaller chromosome parts. An example was found in the short arm chromosome 6 inversion between tomato and potato in which DNA sequences around the inversion breakpoint revealed loss of DNA (Tang et al. 2008).

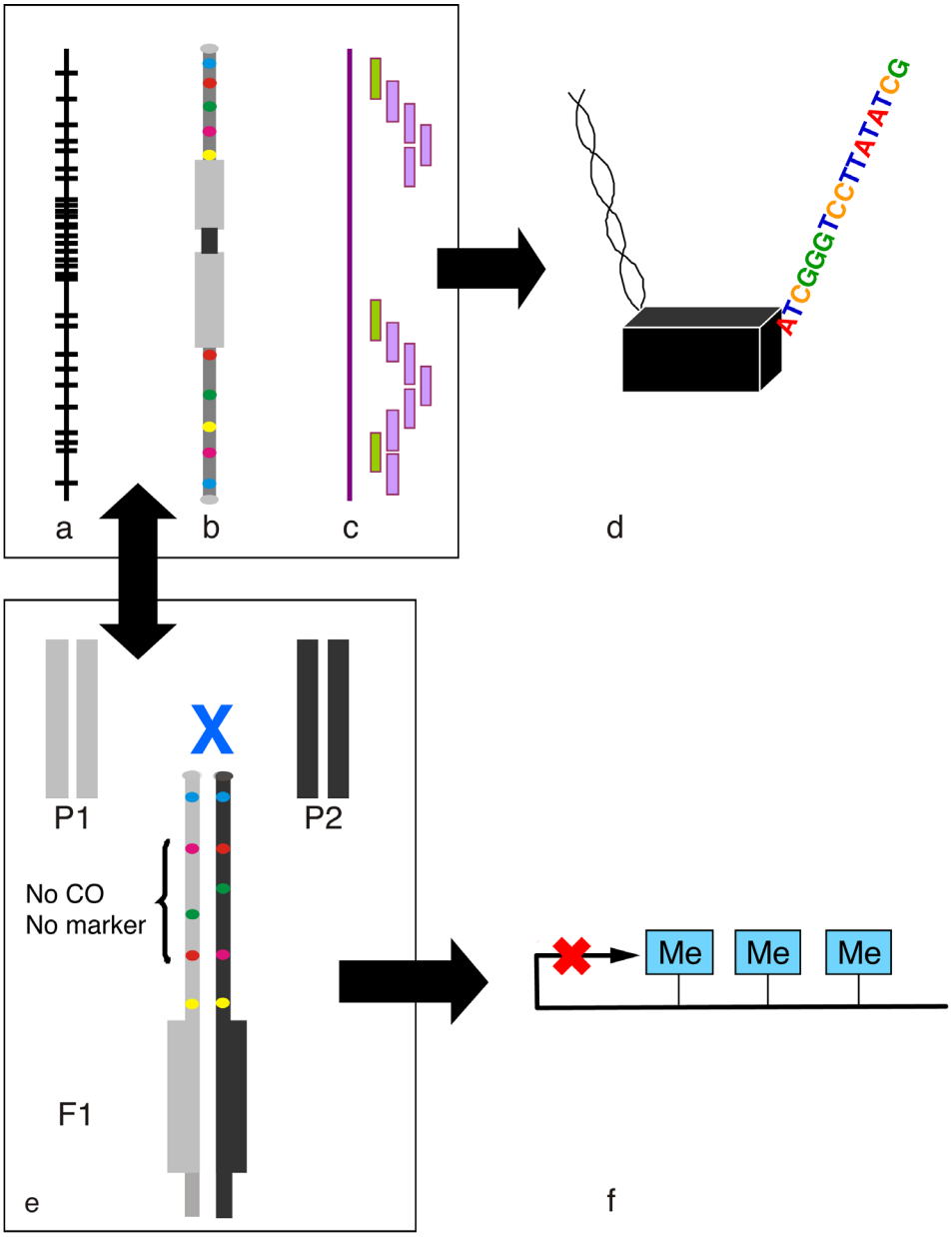


Fig. 5. Schematic representation of introgressive hybridization related to genome sequencing and epigenetic state. (a) genetic map; (b) cytogenetic map; (c) physical map; (d) sequencing; (e) introgressive hybridization; (f) epigenetic changes. Discovering a genome requires genetic, physical and cytogenetic maps, then sequencing. Inserting desired traits from wild relatives can be difficult especially when chromosome rearrangement takes place and there is no homoeologous pairing. Changing in the parental genomes can result from epigenetic modifications and in this case even if the pairing is successful the important region can be silenced.

It is expected that introgressive hybridisation programs will focus more and more on the effect of epigenetic reprogramming of parts of the genomes of the hybrid and its backcross derivatives. Studies of Shaked et al. (2001) have shown that newly formed allopolyploid wheat hybrids may undergo non-mendelian genetic and epigenetic changes. A comparable phenomenon was observed in polyploid *Arabidopsis* where ploidy differences between the parental plants do influence expression of a transgenic resistance gene (Mittelsten Scheid et al. 1996). Also Comai (2000), Comai et al. (2000) and Chen (2007) reviewed the effect of hybridisation and ploidy changes on the epigenetic status of the hybrids, whereas Liu et al. (2004) demonstrated that alien DNA introgression of *Zizania latifolia* into rice can induce extensive alterations in DNA methylation and transcription of both cellular genes and TE-related DNA segments in a genotype-independent manner. Such studies encourage further research in the many other undocumented situations where introgressed genes became silenced during the backcrossings of the introgressive hybridisation program.

References

References

Adams MD, Celniker SE, Holt RA, Evans CA, Gocayne JD, Amanatides PG, Scherer SE, Li PW, Hoskins RA, Galle RF, George RA, Lewis SE, Richards S, Ashburner M, Hender-son SN, Sutton GG, Wortman JR, Yandell MD, Zhang Q, Chen LX, Brandon RC, Rogers YC, Blazej RG, Champe M, Pfeiffer BD, Wan KH, Doyle C, Baxter EG, Helt G, Nelson CR, Gabor Miklos GL, Abril JF, Agbayani A, An HJ, Andrews-Pfannkoch C, Baldwin D, Ballew RM, Basu A, Baxendale J, Bayraktaroglu L, Beasley EM, Beeson KY, Benos PV, Berman BP, Bhandari D, Bolshakov S, Borkova D, Botchan MR, Bouck J, Brokstein P, Brottier P, Burtis KC, Busam DA, Butler H, Cadieu E, Center A, Chandra I, Cherry JM, Cawley S, Dahlke C, Davenport LB, Davies P, De Pablos B, Delcher A, Deng Z, Mays AD, Dew I, Dietz SM, Dodson K, Doup LE, Downes M, Dugan-Rocha S, Dunkov BC, Dunn P, Durbin KJ, Evangelista CC, Ferraz C, Ferriera S, Fleischmann W, Fosler C, Gab-rielian AE, Garg NS, Gelbart WM, Glasser K, Glodek A, Gong F, Gorrell JH, Gu Z, Guan P, Harris M, Harris NL, Harvey D, Heiman TJ, Hernandez JR, Houck J, Hostin D, Hous-ton KA, Howland TJ, Wei MH: The genome sequence of *Drosophila melanogaster*. Science 287:2185-2195 (2000).

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

Albini SM, Jones GH: Synaptonemal complex spreading in *Allium cepa* and *A. fistulosum*. The initiation and sequence of pairing. Chromosoma 95:324-338 (1987).

Albrecht E, Chetelat RT: Comparative genetic linkage map of *Solanum* sect. *Juglandifolia*: Evidence of chromosomal rearrangements and overall synteny with the tomatoes and related nightshades. Theoretical and Applied Genetics 118:831-847 (2009).

Alkhimova OG, Mazurok NA, Potapova TA, Zakian SM, Heslop-Harrison JS, Vershinin AV: Diverse patterns of the tandem repeats organization in rye chromosomes. Chromosoma 113: 42–52 (2004).

Alvarez AE, van de Wiel CCM, Smulders MJM, Vosman B: Use of microsatellites to evaluate genetic diversity and species relationships in the genus *Lycopersicon*. Theoretical and Applied Genetics 103:1283-1292 (2001).

Ammiraju, J.S.S., Veremis, J.C., Huang, X., Roberts, P.A. and Kaloshian, I.: 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 (2003).

Anamthawat-Jónsson K: Molecular cytogenetics of introgressive hybridization in plants. Methods in Cell Science 23:139-148 (2001).

Ananiev EV, Phillips RL, Rines HW: Complex Structure of Knob DNA on Maize Chromosome 9: Retrotransposon Invasion into Heterochromatin. Genetics 149: 2025–2037 (1998).

Anderson LK, Doyle GG, Brigham B, Carter J, Hooker KD, Lai A, Rice M, Stack SM: High-resolution crossover maps for each bivalent of *Zea mays* using recombination nodules. Genetics 165:849-865 (2003).

Anderson L, Covey P, Larsen L, Bedinger P, Stack S: Structural differences in chromosomes distinguish species in the tomato clade. Cytogenetics and Genome Research in press (2010).

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

Appels R, Peacock WJ: The arrangement and evolution of highly repeated (satellite) DNA sequences with special reference to *Drosophila*. International review of cytology Supplement Suppl 8:69-126 (1978).

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

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

Arumuganathan K, Martin GB, Telenius H, Tanksley SD, Earle ED: Chromosome 2-specific DNA clones from flow-sorted chromosomes of tomato. Molecular and General Genetics 242:551-558 (1994).

Asamizu E, Ezura H: Inclusion of tomato in the genus *solanum* as "*Solanum lycopersicum*" Is evident from phylogenetic studies. Journal of the Japanese Society for Horticultural Science 78:3-5 (2009).

Bai Y, van der Hulst R, Huang CC, Wei L, Stam P and Lindhout P: Mapping *Ol*-4, a gene conferring resistance to *Oidium neolycopersici* and originating from *Lycopersicon peruvianum* LA2172, requires multi-allelic, single-locus markers. Theor Appl Genet 109: 1215-1223 (2004).

Bai Y, Lindhout P: Domestication and breeding of tomatoes: What have we gained and what can we gain in the future? Annals of Botany 100:1085-1094 (2007).

Barba-Gonzalez R, Lim KB, Ramanna MS, Visser RGF, Van Tuyl JM: Occurrence of 2n gametes in the F1 hybrids of oriental x asiatic lilies (*Lilium*): Relevance to intergenomic recombination and backcrossing. Euphytica 143:67-73 (2005).

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

Benavente E, Cifuentes M, Dusautoir JC, David J: The use of cytogenetic tools for studies in the crop-to-wild gene transfer scenario. Cytogenetic and Genome Research 120:384-395 (2008).

Bennett MD: Nucleotypic basis of the spatial ordering of chromosomes in eukaryotes and the implications of the order for genome evolution and phenotypic variation. In Genome evolution, pp. 239–261 (1982).

Bennetzen JL: The many hues of plant heterochromatin. Genome biology 1 (2000).

Bensimon A, Simon A, Chiffaudel A, Croquette V, Heslot F, Bensimon D: Alignment and sensitive detection of DNA by a moving interface. Science 265:2096-2098 (1994).

Berg C, Greilhuber J: Cold-sensitive chromosome regions and heterochromatin in *Cestrum aurantiacum* (Solanaceae). Plant Systematics and Evolution 185:259-273 (1993).

Birchler JA, Albert PS, Gao Z: Stability of repeated sequence clusters in hybrids of maize as revealed by FISH. Tropical Plant Biology 1: 34-39 (2008).

Blackburn EH: Structure and function of telomeres. Nature 350: 569-573 (1991).

Blakesley RW, Hansen NF, Mullikin JC, Thomas PJ, McDowell JC, Maskeri B, Young AC, Benjamin B, Brooks SY, Coleman BI, Gupta J, Ho SL, Karlins EM, Maduro QL, Stantripop S, Tsurgeon C, Vogt JL, Walker MA, Masiello CA, Guan X, Bouffard GG, Green ED: An intermediate grade of finished genomic sequence suitable for comparative analyses. Genome Research 14:2235-2244 (2004).

Bohs L: Major clades in *solanum* based in ndhf sequences, in R. C. Keating VCH, and T. B. Croat (ed): A festschrift for William G d'Arcy: The legacy of a taxonomist monographs in systematic botany from the Missouri botanical garden (Missouri Botanical Garden Press, St. Louis 2005).

Bohs L, Olmstead RG: Phylogenetic relationships in *Solanum* (Solanaceae) based on ndhf sequences. Systematic Botany 22:5-17 (1997).

Bonierbale MW, Plaisted RL, Tanksley SD: RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics 120:1095-1103 (1988).

Bovee D, Zhou Y, Haugen E, Wu Z, Hayden HS, Gillett W, Tuzun E, Cooper GM, Sampas N, Phelps K, Levy R, Morrison VA, Sprague J, Jewett D, Buckley D, Subramaniam S, Chang J, Smith DR, Olson MV, Eichler EE, Kaul R: Closing gaps in the human genome with fosmid resources generated from multiple individuals. Nature Genetics 40:96-101 (2008).

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

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

Brown SW: Heterochromatin. Science 151:417-425 (1966).

Buckler ES, Phelps-Durr TL, Buckler CSK, Dawe RK, Doebley JF, Holtsford TP: Meiotic Drive of Chromosomal Knobs Reshaped the Maize Genome. Genetics 153: 415–426 (1999).

Budiman MA, Mao L, Wood TC, Wing RA: A deep-coverage tomato BAC library and prospects toward development of an STC framework for genome sequencing. Genome Research 10:129-136 (2000).

Budiman MA., Chang SB., Lee S, Yang TJ, Zhang HB, de Jong J.H. and Wing RA: 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 (2004).

Bushman FD: Targeting Survival: Integration Site Selection by Retroviruses and LTR-Retrotransposons. Cell 115, 135–138 (2003).

Canady MA, Ji Y, Chetelat RT: Homeologous recombination in *Solanum lycopersicoides* introgression lines of cultivated tomato. Genetics 174:1775-1788 (2006).

Caspersson T, Farber S, Foley GE, Kudynowski J, Modest EJ, Simonsson E, Wagh U, Zech L: Chemical differentiation along metaphase chromosomes. Experimental Cell Research 49:219-222 (1968).

Chalker DL, Sandmeyer SB: Ty3 integrates within the region of RNA polymerase III transcription initiation. Genes & Dev. 6: 117–128 (1992).

Chang SB, de Jong H: Production of alien chromosome additions and their utility in plant genetics. Cytogenetic and Genome Research 109:335-343 (2005).

Chang SB, Anderson LK, Sherman JD, Royer SM and Stack SM: Predicting and testing physical locations of genetically mapped loci on tomato pachytene chromosome 1. Genetics, 176, 2131–2138 (2007).

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

Chase MW, Knapp S, Cox AV, Clarkson JJ, Butsko Y, Joseph J, Savolainen V, Parokonny AS: Molecular systematics, GISH and the origin of hybrid taxa in *Nicotiana* (Solanaceae). Annals of Botany 92:107-127 (2003).

Chen ZJ: Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu. Rev. Plant Biol. 2007. 58:377–406 (2007).

Cheng Z, Buell CR, Wing RA, Gu M, Jiang J: Toward a cytological characterization of the rice genome. Genome Research 11:2133-2141 (2001).

Chetelat RT, Meglic V, Cisneros P: A genetic map of tomato based on BC1 *Lycopersicon esculentum* x *Solanum lycopersicoides* reveals overall synteny but suppressed recombination between these homeologous genomes. Genetics 154:857-867 (2000).

Chojecki AJS, Bayliss MW, Gale MD: Cell production and DNA accumulation in the wheat endosperm, and their association with grain weight. Annals of Botany 58:809-817 (1986).

Cionini PG, Bassi P, Cremonini R, Cavallini A: Cytological localization of fast renaturing and satellite DNA sequences in Vicia faba. Protoplasma 124:106-111 (1985).

Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR: Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (Solanaceae). New Phytologist 168:241-252 (2005).

Coghlan A, Eichler EE, Oliver SG, Paterson AH, Stein L: Chromosome evolution in eukaryotes: A multi-kingdom perspective. Trends in Genetics 21:673-682 (2005).

Cohen S, Houben A, Segal D: Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. The Plant Journal 53, 1027–1034 (2008).

Comai L: Genetic and epigenetic interactions in allopolyploid plants. Plant Mol Biol 43(2):387-399 (2000).

Comai L, Tyagi AP, Winter K, Holmes-Davis R, Reynolds SH, Stevens Y, Byers B: Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. The Plant Cell Online 12(9):1551-1568 (2000).

Copenhaver GP, Browne WE, Preuss D: Assaying genome-wide recombination and centromere functions with *Arabidopsis* tetrads. Proc. Natl. Acad. Sci. USA 95: 247–252 (1998).

Coriton O, Barloy D, Huteau V, Lemoine J, Tanguy AM, Jahier J: Assignment of *Aegilops variabilis* Eig chromosomes and translocations carrying resistance to nematodes in wheat. Genome 52:338-346 (2009).

Cuadrado A, Schwarzacher T: The chromosomal organization of simple sequence repeats in wheat and rye genomes. Chromosoma 107:587– 594 (1998).

Cuadrado A, Jouve N: The non-random distribution of long clusters of all possible classes of tri- nucleotide repeats in barley chromosomes. Chromosome Res 15:711–720 (2007a).

Cuadrado A, Jouve N: Similarities in the chromo- somal distribution of AG and AC repeats with- in and between *Drosophila*, human and barley chromosomes. Cytogenet Genome Res 119:91– 99 (2007b).

Cuzzoni E, Ferretti L, Giordani C, Castiglione S, Sala F: A repeated chromosomal DNA sequence is amplified as a circular extrachromosomal molecule in rice (*Oryza sativa* L.). Mol Gen Genet 222: 58-64 (1990).

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

de Jong JH, Wolters AMA, Kok JM, Verhaar H, van Eden J: Chromosome pairing and potential for intergeneric recombination in some hypotetraploid somatic hybrids of *Lycopersicon esculentum* (+) *Solanum tuberosum*. Genome 36:1032-1041 (1993).

de Jong JH, Fransz P, Zabel P: High resolution FISH in plants - techniques and applications. Trends in Plant Science 4:258-263 (1999).

de Jong JH. Zhong XB, Fransz PF, Wennekes-van Eden J, Jacobsen E, Zabel P: 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: Birkha¨ user Verlag, pp. 267–275 (2000).

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

de Sá MM, Drouin G: Phylogeny and Substitution Rates of Angiosperm Actin Genes. Mol. Biol. Evol. 13(9):1198-1212 (1996).

Desel C, Jansen R, Dedong G, Schmidt T: Painting of parental chromatin in *Beta* hybrids by multi-colour fluorescent in situ hybridization. Annals of Botany 89:171-181 (2002).

Devine SE, Boeke JD: Regionally specific, targeted integration of the yeast retrotransposon Ty1 upstream of genes transcribed by RNA polymerase III. Genes & Dev. 10: 620–633 (1996).

Dietrich AJJ, de Jong JH, Mulder RJP: Location and variation of the constitutive heterochromatin in *Petunia hybrida*. Genetica 55:85-91 (1981).

Dobzhansky T: Genetics of the evolutionary process. Columbia University Press New York (1970).

Doganlar S, Frary A, Daunay MC, Lester RN, Tanksley SD: A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae. Genetics 161:1697-1711 (2002).

Dolezel J, Kubaláková M, Paux E, Bartos J, Feuillet C: Chromosome-based genomics in the cereals. Chromosome Research 15:51-66 (2007).

Dolgin E: Human genomics: The genome finishers. Nature 462:843-845 (2009).

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

Draye X, Lin RY, Qian YX, Bowers EJ, Burow BG, Morrell PL, Peterson DG, Presting GG, Ren SX, Wing RA, Paterson AH: Toward integration of comparative genetic, physical, diversity, and cytomolecular maps for grasses and grains, using the *Sorghum* genome as a foundation. Plant Physiol. 125: 1323-1341 (2001).

Drouaud J, Camilleri C, Bourguignon PY, Canaguier A, Bérard A, Vezon D, Giancola S, Brunel D, Colot V, Prum B, Quesneville H, Mézard C: Variation in crossing-over rates across chromosome 4 of *Arabidopsis thaliana* reveals the presence of meiotic recombination 'hot spots'. Genome Res. 16, 106–114 (2006).

Dutrillaux B, Lejeune J: A new technic of analysis of the human karyotype. Sur une nouvelle technique d'analyse du caryotype humain 272:2638-2640 (1971).

Eichler EE, Clark RA, She X: An assessment of the sequence gaps: Unfinished business in a finished human genome. Nature Reviews Genetics 5:345-354 (2004).

Escalante A, Imanishi S, Hossain M, Ohmido N, Fukui K: RFLP analysis and genomic *in situ* hybridization (GISH) in somatic hybrids and their progeny between *Lycopersicon esculentum* and *Solanum lycopersicoides*. Theoretical and Applied Genetics 96:719-726 (1998).

Eshed Y, Zamir D An introgression line population of *Lycopersicon pennellii* in the cultivated tomato enables the identification and fine mapping of yield-associated QTL. Genetics 141, 1147–1162 (1995).

Fauth C, Speicher MR: Classifying by colors: FISH-based genome analysis. Cytogenetics and Cell Genetics 93 (2001).

Ferguson-Smith M, Yang F, Rens W, O'Brien P: The impact of chromosome sorting and painting on the comparative analysis of primate genomes. Cytogenetic and Genome Research 108:112-121 (2005).

Fernandez Calvin B, Orellana J: Metaphase I-bound arms frequency and genome analysis in wheat *Aegilops* hybrids. 3. Similar relationships between the B genome of wheat and S or S' genomes of *Ae. speltoides, Ae. longissima* and *Ae. sharonensis*. Theoretical and Applied Genetics 88:1043-1049 (1994).

Fernández-Suárez M, Ting AY: Fluorescent probes for super-resolution imaging in living cells. Nature Reviews Molecular Cell Biology 9:929-943 (2008).

Fransz PF, Alonso-Blanco C, Liharska TB, Peeters AJM, Zabel P, Jong JH: High-resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence *in situ* hybridization to extended DNA fibres. The Plant Journal 9:421-430 (1996).

Fransz F., Armstrong S, Alonso-Blanco C, Fischer TC, Torres-Ruiz RA, Jones G: Cytogenetics for the model system *Arabidopsis thaliana*. The Plant Journal 13(6), 867–876 (1998).

Fransz PF, Armstrong S, De Jong JH, Parnell LD, Van Drunen C, Dean C, Zabel P, Bisseling T, Jones GH: Integrated cytogenetic map of chromosome arm 4s of *A. thaliana*: Structural organization of heterochromatic knob and centromere region 100 (3):367-376 (2000).

Friesen N, Klaas M: Origin of some minor vegetatively propagated *Allium* crops studied with RAPD and GISH. Genetic Resources and Crop Evolution 45:511-523 (1998).

Fuchs J, Kuhfittig S, Reuter G: Chromosome painting in *Drosophila*. Chromosome Research 6:335-336 (1998).

Fulton TM, Van der Hoeven R, Eannetta NT, Tanksley SD: Identification, analysis, and utilization of conserved ortholog set markers for comparative genomics in higher plants. Plant Cell 14:1457-1467 (2002).

Ganal MW, Lapitan NLV, Tanksley SD: A molecular and cytogenetic survey of major repeated DNA sequences in tomato (*Lycopersicon esculentum*). Molecular & General Genetics 213:262-268 (1988).

Ganal MW, Young ND, Tanksley SD: 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 (1989).

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

Gao X, Hou Y, Ebina H, Levin HL,Voytas DF: Chromodomains direct integration of retrotransposons to heterochromatin. Genome Research 18:359–369 (2008).

Geigl JB, Uhrig S, Speicher MR: Multiplex-fluorescence in situ hybridization for chromosome karyotyping. Nature Protocols 1:1172-1184 (2006).

Gill BS, Kimber G: Giemsa C banding and the evolution of wheat. Proceedings of the National Academy of Sciences of the United States of America 71:4086-4090 (1974).

Gottschalk W: Die Chromosomenstruktur der Solanaceen unter Berücksichtigung phylogenetischer Phragestellungen. Chromosoma 6, 539-626 (1954).

Greilhuber J: Why Plant Chromosomes Do Not Show G-bands. Theor. Appl. Genet. 50, 121-124 (1977).

Greilhuber J, Loidl J: On regularities of C-banding patterns, and their possible cause. In Proceedings of the Second Kew Chromosome Conference (Brandham, PE. and Bennett, M.D., eds). London: George Allen & Unwin, pp. 344 (1983).

Grube RC, Radwanski ER, Jahn M: Comparative genetics of disease resistance within the solanaceae. Genetics 155:873-887 (2000).

Guerra M: Patterns of heterochromatin distribution in plant chromosomes. Genetics and Molecular Biology 23(4) 1029-1041 (2000).

Guerra M: Fluorescent *in situ* hybridization in plant polytene chromosomes. Methods in Cell Science 23:133-138 (2001).

Hall N: Advanced sequencing technologies and their wider impact in microbiology. Journal of Experimental Biology 210:1518-1525 (2007).

Han FP, Lamb JC, Birchler JA: High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. Proceedings of the National Academy of Sciences of the United States of America 103:3238-3243 (2006).

Heitz E: Die Herkunft der Chromocentren - dritter Beitrag zur Kenntnis der Beziehung zwischen Kernstruktur und qualitativer Verschiedenheit der Chromosomen in ihrer Längsrichtung. Planta 18:571-636 (1932).

Hell SW, Wichmann J: Breaking the diffraction resolution limit by stimulated emission: Stimulated-emission-depletion fluorescence microscopy. Optics Letters 19:780-782 (1994).

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

Hoffmann AA, Sgro MC,Weeks RA: Chromosomal inversion polymorphisms and adaptation. Trends Ecol Evol 19: 482-488 (2004).

Howell EC, Kearsey MJ, Jones GH, King GJ, Armstrong SJ: A and C genome distinction and chromosome identification in *Brassica napus* by sequential fluorescence *in situ* hybridization and genomic in situ hybridization. Genetics 180:1849-1857 (2008).

Huang CC, de Putte PMHV, van der Meer JGH, Dekens FM, Lindhout P: 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 (2000).

Hulbert SH, Webb AC, Smith MS, Sun Q: Resistance gene complexes: evolution and utilization. Annual Review of Phytopathology 39: 285-312 (2001).

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

Iovene M, Savarese S, Cardi T, Frusciante L, Scotti N, Simon PW, Carputo D: Nuclear and cytoplasmic genome composition of *Solanum bulbocastanum* (+) *S. tuberosum* somatic hybrids. Genome 50:443-450 (2007).

Iovene M, Wielgus SM, Simon PW, Buell CR, Jiang J: Chromatin structure and physical mapping of chromosome 6 of potato and comparative analyses with tomato. Genetics 180:1307-1317 (2008).

Jackson SA, Dong F, Jiang J: Digital mapping of bacterial artificial chromosomes by fluorescence *in situ* hybridization. Plant Journal 17:581-587 (1999).

Jackson SA, Cheng Z, Wang ML, Goodman HM, Jiang J: Comparative fluorescence *in situ* hybridization mapping of a 431-kb *Arabidopsis thaliana* bacterial artificial chromosome contig reveals the role of chromosomal duplications in the expansion of the *Brassica rapa* genome. Genetics 156:833-838 (2000).

Jacobsen E, Reinhout P, Bergervoet JEM, de Looff J, Abidin PE, Huigen DJ, Ramanna MS: Isolation and characterization of potato-tomato somatic hybrids using an amylose-free potato mutant as parental genotype. Theoretical and Applied Genetics 85:159-164 (1992).

Jacobsen E, de Jong JH, Kamstra SA, van den Berg PM, Ramanna MS: Genomic in situ hybridization (GISH) and RFLP analysis for the identification of alien chromosomes in the backcross progeny of potato (+) tomato fusion hybrids. Heredity 74:250-257 (1995).

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P: The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature, 449, 463–467 (2007).

Jellen EN, Gill BS, Cox TS: Genomic in situ hybridization differentiates between A/D- and C-genome chromatin and detects intergenomic translocations in polyploid oat species (genus *Avena*). Genome 37:613-618 (1994).

Ji Y, Pertuzé R, Chetelat RT: Genome differentiation by GISH in interspecific and intergeneric hybrids of tomato and related nightshades. Chromosome Research 12:107-116 (2004).

Ji Y, Schuster JD, Scott WJ: Ty-3, a begomovirus resistance locus near the Tomato yellow leaf curl virus resistance locus Ty-1 on chromosome 6 of tomato. Mol. Breed. 20: 271–284 (2007).

Jiang J, Gill BS: Different species-specific chromosome translocations in *Triticum timopheevii* and *T. turgidum* support the diphyletic origin of polyploid wheats. Chromosome Research 2:59-64 (1994).

Jo SH, Koo DH, Kim JF, Hur CG, Lee S, Yang TJ, Kwon SY, Choi D: Evolution of ribosomal DNA-derived satellite repeat in tomato genome. BMC Plant Biology 2009, 9:42 (2008).

Kaloshian I, Yaghoobi J, Liharska T, Hontelez J, Hanson D, Hogan P, Jesse T, Wijbrandi J, Simons G, Vos P, Zabel P, Williamson VM: Genetic and physical localization of the root-knot nematode resistance locus *Mi* in tomato. Mol. Gen. Genet. 257, 376–385 (1998).

Kamstra SA, de Jong JH, Jacobsen E, Ramanna MS, Kuipers AGJ: Meiotic behaviour of individual chromosomes in allotriploid *Alstroemeria* hybrids. Heredity 93:15-21 (2004).

Kapitonov VV, Jurka J: Helitrons on a roll: eukaryotic rolling-circle transposons. Trends in Genetics 23, 521-529 (2007).

Kato A, Lamb JC, Birchler JA: Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. Proceedings of the National Academy of Sciences of the United States of America 101:13554-13559 (2004).

Kenton A, Parokonny AS, Gleba YY, Bennett MD: Characterisation of the *Nicotiana tabacum* l genome by molecular cytogenetics. Molecular & General Genetics 240:159-169 (1993).

Khrustaleva LI, Kik C: Localization of single-copy T-DNA insertion in transgenic shallots (*Allium cepa*) by using ultra-sensitive FISH with tyramide signal amplification. The Plant Journal 25:699-707 (2001).

Khush GS, Rick CM: Meiosis in hybrids between *Lycopersicon esculentum* and *Solanum pennellii*. Genetica 33:167-183 (1963).

Khush GS, Rick CM, Robinson RW: Genetic activity in a heterochromatic chromosome segment of the tomato. Science, 145, 1432–1434 (1964).

Khush GS, Rick CM: Cytogenetic analysis of the tomato genome by means of induced deficiencies. Chromosoma 23: 452-484 (1968).

Kikudome GY: Studies on the phenomenon of preferential segregation in maize. Genetics 44: 815–831 (1959).

Kim JM, Vanguri S, Boeke JD, Gabriel A, Voytas DF: Transposable Elements and Genome Organization: A Comprehensive Survey of Retrotransposons Revealed by the Complete *Saccharomyces cerevisiae* Genome Sequence. Genome Research 8: 464–478 (1998).

Kimura M: A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. Journal of Molecular Evolution 16:111-120 (1980).

Klášterská I: Structure of eukaryotic chromosomes. The differences between mammalian (mouse), grasshopper (stethophyma) and plant (rosa) chromosomes as revealed at the diffuse stage of meiosis. Hereditas 88:243-253 (1978).

Kocsis E, Trus BL, Steer CJ, Bisher ME, Steven AC: Image averaging of flexible fibrous macromolecules: the clathrin triskelion has an elastic proximal segment. J. Struct. Biol. 107:6–14 (1991).

Koo DH, Jiang J: Extraordinary tertiary constrictions of *Tripsacum dactyloides* chromosomes: Implications for karyotype evolution of polyploids driven by segmental chromosome losses. Genetics 179:1119-1123 (2008).

Koo DH, Jo SH, Bang JW, Park HM, Lee S, Choi D: Integration of cytogenetic and genetic linkage maps unveils the physical architecture of tomato chromosome 2. Genetics 179:1211-1220 (2008).

Koo DH, Jiang J: Super-stretched pachytene chromosomes for fluorescence *in situ* hybridization mapping and immunodetection of DNA methylation. Plant Journal 59:509-516 (2009).

Kopecky D, Lukaszewski AJ, Dolezel J: Meiotic behaviour of individual chromosomes of *Festuca pratensis* in tetraploid *Lolium multiflorum*. Chromosome Research 16:987-998 (2008).

Kosina R, Heslop-Harrison JS: Molecular cytogenetics of an amphiploid trigeneric hybrid between *Triticum durum*, *Thinopyrum distichum* and *Lophopyrum elongatum*. Annals of Botany 78:583-589 (1996).

Koumbaris GL, Bass HW: A new single-locus cytogenetic mapping system for maize (*Zea mays* L.): Overcoming FISH detection limits with marker-selected Sorghum (*S. propinquum* L) BAC clones. Plant J. 35: 647-659 (2003).

Kowles RV, Srienc F, Phillips RL: Endoreduplication of nuclear DNA in the developing maize endosperm. Developmental Genetics 11:125-132 (1990).

Lamb J: Plant and Animal Genome Conference, San Diego 2009.

Lamb JC, Danilova T, Bauer MJ, Meyer UM, Holland JJ, Jensen MD, Birchler JA: Single-gene detection and karyotyping using small-target fluorescence *in situ* hybridization on maize somatic chromosomes. Genetics 175:1047-1058 (2007a).

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

Lamoureux D, Peterson DG, Li W, Fellers JP, Gill BS: The efficacy of Cot-based gene enrichment in wheat (*Triticum aestivum* L.) Genome 48: 1120-1126 (2005).

Landgeva S, Ganeva G: N-banded karyotype of *Aegilops ovata* and chromosomal constitution of its amphiploid with *Triticum aestivum*. Plant Breeding 115:330-334 (1996).

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

Lapitan NLV, Ganal MW, Tanksley SD: Organization of the 5S ribosomal RNA genes in the genome of tomato. Genome, 34, 509–514 (1991).

Levin AD: The role of chromosomal change in plant evolution. (Oxford Univerity Press, 2002).

Levin RA, Whelan A, Miller JS: The utility of nuclear conserved ortholog set II (COSII) genomic regions for species-level phylogenetic inference in *Lycium* (Solanaceae). Molecular Phylogenetics and Evolution 50:608-617 (2009).

Liehr T, Weise A, Heller A, Starke H, Mrasek K, Kuechler A, Weier H-U, Claussen U: Multicolor chromosome banding (MCB) with YAC/BAC-based probes and region-specific microdissection DNA libraries. . Cytogenetic and Genome Research 97:43-50 (2002).

Liharska T, Koornneef M, Van Wordragen M, Van Kammen A, Zabel P: Tomato chromosome 6: Effect of alien chromosomal segments on recombinant frequencies. Genome 39:485-491 (1996).

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

Lim KB, Ramanna MS, de Jong JH, Jacobsen E, van Tuyl JM: Indeterminate meiotic restitution (IMR): A novel type of meiotic nuclear restitution mechanism detected in interspecific lily hybrids by GISH. Theoretical and Applied Genetics 103:219-230 (2001).

Lim KY, Leitch IJ, Leitch AR: Genomic characterisation and the detection of raspberry chromatin in polyploid *Rubus*. Theoretical and Applied Genetics 97:1027-1033 (1998).

Linc G, Friebe BR, Kynast RG, Molnar-Lang M, Köszegi B, Sutka J, Gill BS: Molecular cytogenetic analysis of *Aegilops cylindrica* host. Genome 42:497-503 (1999).

Liu Z, Wang Y, Shen Y, Guo W, Hao S, Liu B: Extensive alterations in DNA methylation and transcription in rice caused by introgression from *Zizania latifolia*. Plant Mol Biol (2004).

Livingstone KD, Lackney VK, Blauth JR, van Wijk R, Jahn MK: Genome mapping in *Capsicum* and the evolution of genome structure in the Solanaceae. Genetics 152:1183-1202 (1999).

Lysak MA, Koch MA, Pecinka A, Schubert I: Chromosome triplication found across the tribe Brassiceae. Genome Research 15:516-525 (2005).

Lysak MA, Lexer C: Towards the era of comparative evolutionary genomics in Brassicaceae. Plant Systematics and Evolution 259:175-198 (2006).

Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I: Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. PNAS 103:5224-5229 (2006).

Lysak MA, Cheung K, Kitschke M, Bures P: Ancestral chromosomal blocks are triplicated in Brassiceae species with varying chromosome number and genome size. Plant Physiology 145:402-410 (2007).

Lysak MA, Fransz PF, Ali HBM, Schubert I: Chromosome painting in *Arabidopsis thaliana*. The Plant Journal 28:689-697 (2001).

Lysak MA, Pecinka A, Schubert I: Recent progress in chromosome painting of *Arabidopsis* and related species. Chromosome Research 11: 195-204 (2003).

Masoud SA, Gill BS, Johnson LB: C-banding of alfalfa chromosomes: Standard karyotype and analysis of a somaclonal variant. Journal of Heredity 82:335-338 (1991).

McClean PE, Hanson MR: Mitocondrial DNA sequence divergence among *Lycopersicon* and related *Solanum* species. Genetics 112:649-667 (1986).

McClintock B: The significance of responses of the genome to challenge. Science 226:792-801 (1984).

McDonald BA, Linde C: The population genetics of plant pathogens and breeding strategies for durable resistance. Euphytica 124(2): 163-180 (2002).

Milligan SB, Bodeau J, Yaghoobi J, Kaloshian I, Zabel P, Williamson VM: The root knot nematode resistance gene *Mi* from tomato is a member of the leucine zipper, nucleotide binding, leucine-rich repeat family of plant genes. Plant Cell 10: 1307-1319 (1998).

Mittelsten Scheid O, Paszkowski J, Potrykus I: Reversible inactivation of a transgene in *Arabidopsis thaliana*. Mol Gen Genet 228:104-112 (1991).

Mittelsten Scheid O, Jakovleva L, Afsar K, Maluszynska J, Paszkowski J: A change of ploidy can modify epigenetic silencing. Proceedings of the National Acad. Sciences USA Proc Natl Acad Sci U S A. 93(14):7114–7119 (1996).

Montijn MB, Houtsmuller AB, Ten Hoopen R, Oud JL, Nanninga N: The 5S rRNA gene clusters have a defined orientation toward the nucleolus in *Petunia hybrida* and *Crepis capillaris*. Chromosome Research 7:387-399 (1999).

Moscone EA, Scaldaferro MA, Grabiele M, Cecchini NM, García YS, Jarret R, Davina JR, Ducasse DA, Barboza GE, Ehrendorfer F: The evolution of chili peppers (*Capsicum* -Solanaceae): A cytogenetic perspective: Acta Horticulturae, pp 137-169 (2007).

Mroczek RJ, Dawe RK: Distribution of Retroelements in Centromeres and Neocentromeres of Maize. Genetics 165: 809–819 (2003).

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: The SOL Genomics Network: a comparative resource for Solanaceae biology and beyond. Plant Physiol 138(3): 1310-7 (2005a).

Mueller LA, Steven D. Tanksley , Jim J. Giovannoni , Joyce van Eck , Stephen Stack, Doil Choi, Byung Dong Kim, Mingsheng Chen, Zhukuan Cheng, Chuanyou Li, Hongqing Ling, Yongbiao Xue, Graham Seymour, Gerard Bishop, Glenn Bryan, Rameshwar Sharma, Jiten Khurana, Akhilesh Tyagi, Debasis Chattopadhyay, Nagendra K. Singh, Willem Stiekema, P. Lindhout, Taco Jesse, Rene Klein Lankhorst, Mondher Bouzayen, Daisuke Shibata, Satoshi Tabata, Antonio Granell, Miguel A. Botella, Giovanni Giuliano, Luigi Frusciante, Mathilde Causse and Dani Zamir: The Tomato Sequencing Project, the first cornerstone of the International Solanaceae Project (SOL). Comparative and Functional Genomics 6: 153 -158 (2005b).

Mueller LA, Klein Lankhorst R, Tanksley SD, Giovannoni JJ, White R, Vrebalov J, Fei Z, van Eck J, Buels R. Mills AA, Menda N, Tecle IY, Bombarely A, Stack S, Royer SM, Chang S-B, Shearer LA, Kim BD, Jo SH, Hur CG, Choi D, Li CB, Zhao J, Jiang H, Geng Y, Dai Y, Fan H, Chen J, Lu F, Shi J, Sun S, Chen J, Yang X, Lu C, Chen M, Cheng Z, Li C, Ling H, Xue Y, Wang Y, Seymour GB, Bishop GJ, Bryan G, Rogers J, Sims S, Butcher S, Buchan D, Abbott J, Beasley H, Nicholson C, Riddle C, Humphray S, McLaren K, Mathur S, Vyas S, Solanke AU, Kumar R, Gupta V, Sharma AK, Khurana P, Khurana JP, Tyagi A, Chowdhury SP, Shridhar S, Chattopadhyay D, Pandit A, Singh P, Kumar A, Dixit R, Singh A, Praveen S, Dalal V, Yadav M, Ghazi IA, Gaikwad K, Sharma TR, Mohapatra T, Singh NK, Szinay D, de Jong H, Peters S, van Staveren M, Datema E, Fiers MWEJ, van Ham RCHJ, Lindhout P, Philippot M, Frasse P, Regad F, Zouine M, Bouzayen M, Asamizu E, Sato S, Fukuoka H, Tabata S, Shibata D, Botella MA, Perez-Alonso M, Fernandez-Pedrosa V, Osorio S, Mico A, Granell A, Zhang Z, He J, Huang S, Du Y, Qu D, Liu L, Liu D, Wang J, Ye Z, Yang W, Wang G, Vezzi A, Todesco S, Valle G, Falcone G, Pietrella M, Giuliano G, Grandillo S, Traini A, D'Agostino N, Chiusano ML, Ercolano M, Barone A, Frusciante L, Schoof H, Jöcker A, Bruggmann R, Spannagl M, Mayer KXF, Guigó R, Camara F, Rombauts S, Fawcett JA, Van de Peer Y, Knapp S, Zamir D, Stiekema W: A Snapshot of the Emerging Tomato Genome Sequence. The Plant Genome 2:78–92. (2009).

Muller S, Neusser M, Wienberg J: Towards unlimited colors for fluorescence *in-situ* hybridization (FISH). Chromosome Research 10: 223-232 (2002).

Nagl W: Endopolyploidy and polyteny in differentiation and evolution, North Holland, Amsterdam (1978).

Navrátilová A, Koblizková A, Macas J: Survey of extrachromosomal circular DNA derived from plant satellite repeats. BMC Plant Biology 8 (2008).

Nilsson M, Krejci K, Koch J, Kwiatkowski M, Gustavsson P, Landegren U: Padlock probes reveal single-nucleotide differences, parent of origin and *in situ* distribution of centromeric sequences in human chromosomes 13 and 21. Nature Genetics 16:252-255 (1997).

Painter TS: A new method for the study of chromosome rearrangements and the plotting of chromosome maps. Science 78:585-586 (1933).

Paterson AH: Leafing through the genomes of our major crop plants: strategies for capturing unique information. Nature Review Genetics 7 174-184 (2006).

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S, Berges H, Eversole K, Ap-

pels R, Safar J, Simkova H, Dolezel J, Bernard M, Feuillet C: A physical map of the 1-gigabase bread wheat chromosome 3b. Science 322:101-104 (2008).

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

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

Pedrosa-Harand A, Kami J, Gepts P, Geffroy V, Schweizer D: Cytogenetic mapping of common bean chromosomes reveals a less compartmentalized small-genome plant species. Chromosome Research 17:405-417 (2009).

Pereira V: Bias and purifying selection of retrotransposons in the *Arabidopsis thaliana* genome. Genome Biology, 5:R79 (2004).

Perez F, Menendez A, Dehal P, Quiros CF: Genomic structural differentiation in *Solanum*: comparative mapping of the A- and E-genomes. Theor Appl Genet 98: 1183-1193 (1999).

Pertuzé RA, Ji Y, Chetelat RT: Comparative linkage map of the *Solanum lycopersicoides* and *S. sitiens* genomes and their differentiation from tomato. Genome 45:1003-1012 (2002).

Peters SA, Datema E, Szinay D, Van Staveren MJ, Schijlen EGWM, Van Haarst JC, Hesselink T, Abma-Henkens MHC, Bai Y, De Jong H, Stiekema WJ, Klein Lankhorst RM, Van Ham RCHJ: *Solanum lycopersicum* cv. Heinz 1706 chromosome 6: Distribution and abundance of genes and retrotransposable elements. Plant Journal 58:857-869 (2009).

Peters SA, van Haarst JC, Jesse TP, Woltinge D, Jansen K, Hesselink T, van Staveren MJ, Abma-Henkens, MCJ, Klein Lankhorst RM: TOPAAS, a tomato and potato assembly assistance system for selection and finishing of bacterial artificial chromosomes. Plant Physiol. 140, 805–817 (2006).

Peterson DG, Price HJ, Johnston JS, Stack SM: DNA content of heterochromatin and euchromatin in tomato (*Lycopersicon esculentum*) pachytene chromosomes. Genome, 39: 77-82 (1996).

Peterson DG, Pearson WR, Stack SM: Characterization of the tomato (*Lycopersicon esculentum*) genome using in vitro and in situ DNA reassociation. Genome 41:346-356 (1998).

Peterson DG, Lapitan NLV, Stack SM: Localization of single- and low-copy sequences on tomato synaptonemal complex spreads using fluorescence in situ hybridization (FISH). Genetics 152:427-439 (1999).

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

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

Pont G, Degroote F, Picard C: Some extrachromosomal circular DNA from *Drosophila* embryos are homologous to tandemly repeated genes. J. Mol. Biol 195, 447-451 (1987).

Presting GG, Frary A, Pillen K, Tanksley SD: Telomere-homologous sequences occur near the centromeres of many tomato chromosomes. Mol Gen Genet 251:526-531 (1996).

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

Raap AK, Tanke HJ: Combined binary ratio fluorescence in situ hybridiziation (COBRA-fish): Development and applications. Cytogenetic and Genome Research 114:222-226 (2006).

Rabinowicz PD: Plant genomic sequencing using gene-enriched libraries. Chemical Reviews 107:3377-3390 (2007).

Rabl C: Über Zellteilung. Morphologisches Jahrbuch 10:214-230 (1885).

Ramanna M, Prakken R: Structure of and homology between pachytene and somatic metaphase chromosomes of the tomato. Genetica 38:115-133 (1967).

Ramanna MS, Wagenvoort M: Identification of the trisomic series in diploid *Solanum tuberosum* L., group *tuberosum*. I. Chromosome identification. Euphytica 25:233-240 (1976).

Ren Y, Zhang Z, Liu J, Staub JE, Han Y, Cheng Z, Li X, Lu J, Miao H, Kang H, Xie B, Gu X, Wang X, Du Y, Jin W, Huang S: An integrated genetic and cytogenetic map of the cucumber genome. PLoS ONE 4 (2009).

Rens W, Fu B, O'Brien PCM, Ferguson-Smith M: Cross-species chromosome painting. Nature Protocols 1:783-790 (2006a).

Rens W, Moderegger K, Skelton H, Clarke O, Trifonov V, Ferguson-Smith MA: A procedure for image enhancement in chromosome painting. Chromosome Research 14:497-503 (2006b).

Rhoades MM, Dempsey E, The effect of abnormal 10 on preferential segregation and crossing over in maize. Genetics 53: 989–1020 (1966).

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

Saitou N, Nei M: The neighbor-joining method: A new method for reconstructing phylogenetic trees. Molecular Biology and Evolution 4:406-425 (1987).

Sánchez-Morán E, Benavente E, Orellana J: Simultaneous identification of A, B, D and R genomes by genomic in situ hybridization in wheat-rye derivatives. Heredity 83:249-252 (1999).

Schermelleh L, Carlton PM, Haase S, Shao L, Winoto L, Kner P, Burke B, Cardoso MC, Agard DA, Gustafsson MGL, Leonhardt H, Sedat JW: Subdiffraction Multicolor Imaging of the Nuclear Periphery with 3D Structured Illumination Microscopy. Science 320 (5881):1332 – 1336, DOI: 10.1126/science.1156947 (2008).

Schmidt T, Heslop-Harrison JS: Genomes, genes and junk: The large-scale organization of plant chromosomes. Trends in Plant Science 3:195-199 (1998).

Schubert I, Rieger R: Asymmetric banding of *Vicia faba* chromosomes after BrdU incorporation. Chromosoma 70:385-391 (1979).

Schubert I, Fransz PF, Fuchs J, De Jong JH: Chromosome painting in plants. Methods in Cell Science 23:57-69 (2001).

Schwarzacher T, Heslop-Harrison P: Practical in situ Hybridization. BIOS Scientific Publishers Ltd, Oxford. ISBN 1-85996-138-x, pp. 1-203 (2000).

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

Seah S, Yaghoobi J, Rossi M, Gleason CA, Williamson VM: The nematode-resistance gene, *Mi*-1, is associated with an inverted chromosomal segment in susceptible compared to resistant tomato. Theoretical and Applied Genetics 108:1635-1642 (2004).

Sepsi A, Molnár I, Szalay D, Molnár-Láng M: Molecular cytogenetic analysis of the wheat-*Agropyron elongatum* partial amphiploid BE-1. Acta Biologica Szegediensis 52:139-141 (2008).

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

Shaked H, Kashkush K, Ozkan H, Feldman M, Levy AA: Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell 13, 1749–1759 (2001).

Sherman JD, Stack SM: Two-dimensional spreads of synaptonemal complexes from solanaceous plants. High-resolutio recombination nodule map for tomato (*Lycopersicon esculentum*). Genetics, 141, 683–708 (1995).

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

Smith MJ, Britten RJ, Davidson EH: Biochemistry Studies on nucleic acid reassociation kinetics: Reactivityof single-stranded tails in DNA-DNA renaturation. Proc.Nat.Acad.Sci.USA Vol 72 (12): 4805-4809 (1975).

Spooner DM, Peralta IE, Knapp S: Comparison of aflps with other markers for phylogenetic inference in wild tomatoes [*Solanum* l. Section *Lycopersicon* (mill.) wettst.]. Taxon 54:43-61 (2005).

Stack S, Anderson L: Two-dimensional spreads of synaptonemal complexes from solanaceous plants - III. Recombination nodules and crossing over in *Lycopersicon esculentum* (tomato). Chromosoma 94:253-258 (1986).

Stack SM, Anderson LK: Crossing over as assessed by late recombination nodules is related to the pattern of synapsis and the distribution of early recombination nodules in maize. Chromosome Research 10:329-345 (2002).

Stamova BS, Chetelat RT: Inheritance and genetic mapping of cucumber mosaic virus resistance introgressed from Lycopersicon chilense into tomato. Theoretical and Applied Genetics 101:527-537 (2000).

Sturtevant AH: Contributions to the genetics of *Drosophila melanogaster* III Inherited linkage variations in the second chromosome. Carnegie Inst Wash Publ 278: 305-341 (1919).

Sumner AT: Chromosome banding. Unwin Hyman Publishing Inc., 448 pg. (1991).
Sykorová E, Lim KY, Fajkus J, Leitch AR: The signature of the *Cestrum* genome suggests an evolutionary response to the loss of (TTTAGGG)n telomeres. Chromosoma 112:164-172 (2003).
Szinay D, Chang SB, Khrustaleva L, Peters S, Schijlen E, Bai Y, Stiekema WJ, Van Ham RCHJ, de Jong H, Klein Lankhorst RM: High-resolution chromosome mapping of BACs using multicolour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6. Plant Journal 56:627-637 (2008).
Szinay D, Bai Y, Visser R and de Jong JH: FISH applications for genomics and plant breeding strategies in tomato and other Solanaceous crops. Cytogenetics and Genome Research (in press) (2010).
Szuhai K, Tanke HJ: Cobra: Combined binary ratio labeling of nucleic-acid probes for multicolor fluorescence *in situ* hybridization karyotyping. Nature Protocols 1:264-275 (2006).
Takahashi C, Marshall JA, Bennett MD, Leitch IJ: Genomic relationships between maize and its wild relatives. Genome 42:1201-1207 (1999).
Tam SM, Mhiri C, Vogelaar A, Kerkveld M, Pearce SR, Grandbastien M-A: Comparative analyses of genetic diversities within tomato and pepper collections detected by retrotransposon-based SSAP, AFLP and SSR. TAG Theoretical and Applied Genetics 110:819-831 (2005).
Tamura K, Dudley J, Nei M, Kumar S: MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Molecular Biology and Evolution 24: 1596-1599 (2007).
Tang X, Szinay D, Lang C, Ramanna MS, Van Der Vossen EAG, Datema E, Lankhorst RK, De Boer J, Peters SA, Bachem C, Stiekema W, Visser RGF, De Jong H, Bai Y: Cross-species bacterial artificial chromosome-fluorescence in situ hybridization painting of the tomato and potato chromosome 6 reveals undescribed chromosomal rearrangements. Genetics 180:1319-1328 (2008).
Tang X, de Boer JM, van Eck HJ, Bachem C, Visser RGF, de Jong H: Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. Chromosome Research 17:899-915 (2009).
Tanke HJ, Dirks RW, Raap T: FISH and immunocytochemistry: Towards visualising single target molecules in living cells. Current Opinion in Biotechnology 16:49-54 (2005).
Tanksley SD, Ganal MW, Prince JP, Devicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Roder MS, Wing RA, Wu W, Young ND: High-density molecular linkage maps of tomato and potato genomes. Genetics 132:1141-1160 (1992).
Tek AL, Song J, Macas J, Jiang J: Sobo, a recently amplified satellite repeat of potato, and its implications for the origin of tandemly repeated sequences. Genetics 170:1231-1238 (2005).
Tek AL, Jiang J: The centromeric regions of potato chromosomes contain megabase-sized tandem arrays of telomere-similar sequence. Chromosoma 113: 77-83 (2004).
Tessadori F, Chupeau MC, Chupeau Y, Knip M, Germann S, van Driel R, Fransz P, Gaudin V: Large-scale dissociation and sequential reassembly of pericentric heterochromatin in differentiated *Arabidopsis* cells. Journal of Cell Science 120:1200-1208 (2007).

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

Tor M, Manning K, King GJ, Thompson AJ, Jones GH, Seymour GB, Amstrong SJ: Genetic analysis and FISH mapping of the colourless non-ripening locus of tomato. Theor. Appl. Genet. 104, 165–170 (2002).

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

Valárik M, Bartos J, Kovárová P, Kubaláková M, de Jong JH, Doležel J: High-resolution FISH on super-stretched flow-sorted plant chromosomes. Plant Journal 37:940-950 (2004).

van Daelen RAJJ, Gerbens F, van Ruissen F, Aarts J, Hontelez J, Zabel P: 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 (1993).

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

van der Knaap E, Sanyal A, Jackson SA, Tanksley SD: High-resolution fine mapping and fluorescence in situ hybridization analysis of sun, a locus controlling tomato fruit shape, reveals a region of the tomato genome prone to DNA rearrangements. Genetics Society of America 168:2127-2140 (2004).

van der Voort JNAMR, van Zandvoort P, van Eck HJ, Folkertsma RT, Hutten RCB, Draaistra J, Gommers FJ, Jacobsen E, Helder J, Bakker J: Use of allele specificity of comigrating AFLP markers to align genetic maps from different potato genotypes. Molecular and General Genetics MGG 255:438-447 (1997).

van der Vossen E, Gros J, Sikkema A, Muskens M, Wouter D et al.: The *Rpi-blb*2 gene from *Solanum bulbocastanum* is an *Mi*-1 gene homolog conferring broad-spectrum late blight resistance in potato. Plant Journal 44:208-222 (2005).

van Os H, Andrejewski S, Bakker E, Barrena I, Bryan GJ: Construction of a 10,000-marker ultradense genetic recombination map of potato: providing a framework for accelerated gene isolation and a genomewide physical map. Genetics 173:1075-1087 (2006).

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

van Wordragen MF, Weide RL, Coppoolse E, Koornneef M, Zabel P: 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 (1996).

Vega JM, Abbo S, Feldman M, Levy AA: Chromosome painting in plants: In situ hybridization with a DNA probe from a specific microdissected chromosome arm of common wheat. Proceedings of the National Academy of Sciences of the United States of America 91:12041-12045 (1994).

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: 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. American Journal of Potato Research:1-13 (2009).

Vosman B, Arens P: Molecular characterization of GATA/ GACA microsatellite repeats in tomato. Genome 40: 25-33 (1997).

Vrana J, Kubalakova M, Simkova H, Cihalikova J, Lysak MA, Doležel J: Flow sorting of mitotic chromosomes in common wheat (Triticum aestivum L.). Genetics Society of America 156:2033-2041 (2000).

Wang Y, Tang X, Cheng Z, Mueller L, Giovannoni J, Tanksley SD: Euchromatin and pericentromere heterochromatin: comparative composition in the tomato genome. Genetics 172: 2529-2540 (2006).

Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD: Sequencing and comparative analysis of a conserved syntenic segment in the solanaceae. 180:391-408 (2008).

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

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

Weide R, Hontelez J, van Kammen A, Koornneef M, Zabel P: 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 (1998).

Werner JE, Kota RS, Gill BS, Endo TR: Distribution of telomeric repeats and their role in the healing of broken chromosome ends in wheat. Genome 35: 844-848 (1992).

Wiegant J, Bezrookove V, Rosenberg C, Tanke HJ, Raap AK, Zhang H, Bittner M, Trent JM, Meltzer P: Differentially painting human chromosome arms with combined binary ratio-labeling fluorescence *in situ* hybridization. Genome Research 10:861-865 (2000).

Wright DA, Voytas DF: Athila of Arabidopsis and Calypso of Soybean Define a Lineage of Endogenous Plant Retroviruses. Genome Res. 12: 122-131 (2002).

Xu J, Earle ED: High resolution physical mapping of 45S (5.8S, 18S and 25S) rDNA gene loci in the tomato genome using a combination of karyotyping and FISH of pachytene chromosomes. Chromosoma 104:545-550 (1996).

Xu SJ, Singh RJ, Kollipara KP, Hymowitz T: Primary trisomics in soybean: Origin, identification, breeding behavior and use in linkage mapping. Crop Science 40:1543-1551 (2000).

Yang TJ, Lee S, Chang SB, Yu Y, de Jong H, Wing RA: In-depth sequence analysis of the tomato chromosome 12 centromeric region: Identification of a large CAA block and characterization of pericentromere retrotranposons. Chromosoma 114:103-117 (2005).

Yasuhara JC, DeCrease CH, Wakimoto BT: Evolution of heterochromatic genes of *Drosophila*. Proc. Natial. Acad. Science USA 102:10958–10963 (2005).

Yasuhara JC, Wakimoto BT: Oxymoron no more: The expanding world of heterochromatic genes. Trends in Genetics 22:330-338 (2006).

Yasuhara JC, Wakimoto BT: Molecular landscape of modified histones in *Drosophila* heterochromatic genes and euchromatin-heterochromatin transition zones. PLoS Genet 4(1): e16. doi:10.1371/journal.pgen.0040016 (2008).

Yeh BP, Peloquin SJ: Pachytene chromosomes of the potato (*Solanum tuberosum*, Group Andigena). American Journal of Botany 52:1014-1020 (1965).

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S: Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiology 137:1174-1181 (2005).

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.

Yuan Y, SanMiguel PJ, Bennetzen JL: High-Cot sequence analysis of the maize genome. The Plant J 34:249-255 (2003).

Zabel P, Meyer D, van de Stolpe O, et al.: Towards the construction of artificial chromosomes for tomato. In: van Vloten-Doting L, Groot GSP, Hall TC, eds. Molecular Form and Function of the Plant Genome. New York: Plenum, 609-624 (1985).

Zamir D, Tanksley SD: Tomato genome is comprised largely of fast-evolving, low copy-number sequences. Molecular & General Genetics 213:254-261 (1988).

Zhang P, Li W, Fellers J, Friebe B, Gill BS: BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. Chromosoma 112:288-299 (2004a).

Zhang P, Li W, Friebe B, Gill BS: Simultaneous painting of three genomes in hexaploid wheat by BAC-FISH. Genome 47:979-987 (2004b).

Zhang W, Lee HR, Koo DH, Jiang J: Epigenetic modification of centromeric chromatin: Hypomethylation of DNA sequences in the CENH3-associated chromatin in *Arabidopsis thaliana* and maize. Plant Cell 20:25-34 (2008).

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

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

Zhong XB, Fransz PF, Eden JWV, Zabel P, Van Kammen A, De Jong JH: High-resolution mapping on pachytene chromosomes and extended DNA fibres by fluorescence *in-situ* hybridisation. Plant Molecular Biology Reporter 14:232-242 (1996b).

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

Zhong XB, Bodeau J, Fransz PF, Williamson VM, van Kammen A, de Jong JH, Zabel P: 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 (1999).

Zhong XB, Lizardi PM, Huang XH, Bray-Ward PL, Ward DC: Visualization of oligonucleotide probes and point mutations in interphase nuclei nuclei and DNA fibers using rolling circle DNA amplification. Proceedings of the National Academy of Sciences 98 (7):3940–3945 (2001)

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.

Summary

In this thesis various Fluorescence *in situ* Hybridization (FISH) technologies are described to support genome projects, plant breeding and phylogenetic analysis on tomato (*Solanum lycopersicum*, 2n=24). Its genome is 980 Mb and only 30 % are single copy sequences, which are mostly found in the euchromatin regions. These regions in all 12 chromosomes were therefore focus of the International Solanaceae Genome Sequencing Project. Based on the F2.2000 linkage map bacterial artificial chromosomes (BACs) were selected from three libraries for validating their physical locations by FISH. Chapter 2 deals with a five-colour high-resolution BAC FISH approach and presents the results of the mapping of 75 seed BACs on pachytene complements of chromosome 6. There were obvious differences between the cytogenetic map and the linkage map. Most of the discrepancies occurred in the pericentromeric heterochromatin where recombination is highly suppressed. For establishing the BAC coverage of chromosome 6 a pooled BAC FISH method was used to hybridize all seed BACs simultaneously. A few larger gaps were discovered mostly on the long arm, where our 'BAC-by-BAC' sequencing approach could not manage to close the gaps by extending contigs. Afterwards new candidate BACs were tested by pooled-BAC FISH. Finally we demonstrated the heterochromatin/euchromatin distribution focusing on its borders by mapping pooled repetitive sequences (Cot 100) together with border BACs. In Chapter 3 the repeat content of chromosome 7 was analyzed by combining BAC and extended fiber FISH mapping with bioinformatics of 169 BACs. Repeats are important due to their challenging interpretations in genome sequencing. Tandem arrays of Tomato Genome Repeat I (TGRI) were found in BACs close to the distal end of chromosome 7 as well as on the long arm interstitial knobs. Phylogenetic analysis by Neighbor-Joining approach showed clustering of the TGRI blocks that suggested their independent origin. TGRI is likely to be transposed by extrachromosomal circular DNA molecules during interphase. The dispersed TGR repeats (TGRII, TGRIII, TGRIV) all belong to the Ty3-*Gypsy* LTR class of retrotransposons. All of them cover the pericentromeric heterochromatin but overlap only partly as shown by FISH and BAC sequencing. TGRII hybridized through the whole pericentromere, TGRIII overlapped with TGRII except for the distal regions of the heterochromatin on the long arm, whereas TGRIV showed coverage in the most proximal parts of the short arm heterochromatin. BAC sequences corresponded well to the FISH data except that there were solo LTRs of TGRII found in the euchromatin. In the pericentromere heterochromatin truncated and solo LTRs were present of both TGRII and TGRIII. The TGRIV repeat could not be further investigated due to too high repeat content of the BACs in that region. Furthermore, this chapter offers some clues about the TGR repeats distributed in the pericentromere.

In Chapter 4 a comparative mapping study was carried out between tomato and potato (*Solanum tuberosum*) chromosome 6 using BACs from both species. The BACs were hybridized on both species by FISH. Due to some repeat-rich BACs Cot 100 blocking was necessary as well as lowered stringent washing to achieve unique and clear signals. We detected a novel paracentric inversion on the short arm of chromosome 6. The two break points are close to the

distal heterochromatin end and to the eu- heterochromatin border. The BAC order revealed colinearity on the long arm. The two investigated tomato cultivars- Heinz 1706 and Cherry VFNT- were colinear for all of the used BACs. One (RH98-856-18) out of six potato clones differed by a small rearrangement in the middle of the inversion. This study gave a first idea for evolutionary studies in the *Solanum* genus using chromosomal rearrangements as detected by FISH and which are elaborated in Chapter 5. It is known that chromosomal rearrangements happen frequently, but rarely get fixed during evolution. The reason is that chromosomal rearrangements have often a negative influence on fertility and on the progeny. In *Solanum* most of the large inversions were previously reported. We selected repeat poor and evenly distributed tomato and potato BACs and after labeling those by fluorescence dyes we hybridized them across related wild species, tomato breeding lines, potato, eggplant and pepper (which is a close relative outside of the genus). We could reveal synteny between these species. In this way we discovered five undescribed inversions and also found discrepancies with previous literature claiming chromosomal rearrangements. Our results correspond well to published phylogeny on *Solanum*, suggesting that our approach would be suitable for studying unknown genomes and resolving relationships on a higher level, such as sections.

Finally this thesis discusses the crucial points of FISH technology; such as spatial resolution, detection sensitivity and applicability. It highlights the strength, weaknesses, opportunities and threads of FISH. In conclusion FISH is an indispensible technique for sequencing large genomes and defining repeat content with support of bioinformatics. Moreover, hidden chromosomal rearrangements can be visualized in regions where recombination is suppressed, which is important for plant breeding and definitely for phylogenetic studies.

Samenvatting

In dit proefschrift worden verschillende fluorescente in situ hybridisatie (FISH) technieken beschreven ter ondersteuning van genoomprojecten, plantveredeling en fylogenetische analyses gericht op tomaat (*Solanum lycopersicum*, 2n=24). Dit genoom is 980 Mb groot en bestaat voor slechts 30% uit unieke sequenties, welke voornamelijk in de euchromatine gebieden te vinden zijn. Deze gebieden in alle 12 chromosomen waren daarom het brandpunt van het internationale Solanaeceae Genome Sequencing project. Aan de hand van de F2.2000 genetische kaart werden Bacterial Artificial Chromosomes (BACs) geselecteerd uit drie bibliotheken teneinde hun fysieke locaties te valideren aan de hand van Fluorescence in situ Hybridization (FISH).

In hoofdstuk 2 beschrijf ik een vijf-kleuren, hoge-resolutie BAC FISH aanpak en de resultaten van het plaatsen van 75 'seed' BACs op pachyteencomplementen van chromosoom 6. We hebben verschillen gevonden tussen de cytogenetische kaart en de genetische kaart. Het merendeel van de discrepanties werd gevonden in het pericentromerische heterochromatine, waar recombinatie sterk onderdrukt wordt. Om de dekking van chromosoom 6 door BACs vast te stellen is een 'pooled BAC FISH' methode toegepast waarin alle BACs tegelijkertijd gehybridiseerd worden. Een aantal grote gaten werd ontdekt, voornamelijk op de lange arm, waar onze 'BAC-by-BAC' sequentie aanpak niet in staat was om de gaten te sluiten door het uit te breiden van contigs. Naderhand werden nieuwe kandidaat BACs getest met behulp van 'pooled-BAC FISH'. Ten slotte hebben we de distribute van euchromatine en heterochromatine, en in het bijzonder de grenzen hiervan, laten zien aan de hand van het plaatsen van samengevoegde repetitieve sequenties (Cot 100) samen met grens-BACs.

In hoofdstuk 3 is het gehalte aan repetitieve sequenties van chromosoom 7 geanalyseerd door het combineren van BAC en 'extended fiber FISH kartering' met bioinformatische analyses van 169 BACs. Repetitieve sequenties zijn van belang vanwege hun lastige interpretatie bij het sequencen van genomen. Opeenvolgende gebieden van Tomato Genome Repeat I (TGRI) werden zowel in BACs dichtbij de distale einden van chromosoom 7 als op de interstitiële knopen op de lange arm gevonden. Phylogenetische analyse aan de hand van een 'Neighbour-Joining' aanpak toonde een groepering van de TGRI blokken aan die een onafhankelijke oorsprong van deze blokken suggereerde. TGRI transposeert waarschijnlijk door middel van extrachromosomale, circulaire DNA moleculen tijdens de interfase. De verspreidde TGR repetitive elementen (TGRII, TGRIII, TGRIV) behoren allen tot de Ty3-*Gypsy* klasse van de retrotranspons. Al deze elementen bedekken het pericentromerische heterochromatine maar overlappen elkaar slechts gedeeltelijk, zoals is aangetoond door middel van FISH en BAC sequencen. TGRII hybridiseerde in het gehele pericentromeer gebied, TGRIII overlapte met TGRII met uitzondering van de distale regionen van het heterochromatine op de lange arm, terwijl TGRIV dekking vertoonde van de meest proximale gebieden van het heterochromatine op de korte arm. BAC sequenties kwamen goed overeen met de FISH resultaten, echter er werden ook solo LTRs van TGRII gevonden in het euchromatine. In het pericentromere heterochroma-

tine waren getrunceerde en solo LTRs aanwezig van zowel TGRII als TGRIII. Het TGRIV element kon niet verder onderzocht worden vanwege een te hoog gehalte aan repetitieve sequenties in het chromosoomgebied. Verder geeft dit hoofdstuk enkele aanwijzingen over de verdeling van de TGR elementen in het pericentromeer gebied.

In hoofdstuk 4 is een vergelijkende karteringstudie uitgevoerd tussen chromosoom 6 van tomaat en aardappel (*Solanum tuberosum*) aan de hand van BACs van beide soorten. De BACs werden gehybridiseerd op beide soorten door middel van FISH. Vanwege enkele BACs die veel repetitieve sequenties bevatten was het noodzakelijk om Cot-100 blokkering toe te passen alsmede het wassen na de hybridisatie bij een lagere stringentie teneinde unieke en duidelijke signalen te verkrijgen. We hebben een nieuwe paracentrische inversie waargenomen op de korte arm van chromosoom 6. De twee breekpunten bevinden zich dichtbij het distale heterochromatine einde en dichtbij de grens tussen eu- en heterochromatine. De BAC volgorde toonde colineariteit aan op de lange arm. De twee tomaat cultivars die onderzocht werden – Heinz 1706 en Cherry VFNT – waren colineair voor alle gebruikte BACs. Eén (RH98-856-18) van de zes aardappel klonen was verschillend met een kleine herschikking in het midden van de inversie. Deze studie gaf een eerste idee voor evolutionaire onderzoeksstudies in de genus *Solanum* aan de hand van chromosomale herschikkingen zoals gedetecteerd door FISH en zoals beschreven in hoofdstuk 5. Het is bekend dat chromosomale herschikkingen frequent plaatsvinden, maar slechts zelden gefixeerd worden tijdens evolutie. De oorzaak hiervan is dat chromosomale herschikkingen vaak een negatieve invloed hebben op vruchtbaarheid en op de nakomelingen. Het merendeel van de inversies in *Solanum* werden reeds eerder gerapporteerd. Wij hebben BACs uit tomaat en aardappel geselecteerd die weinig repetitieve sequenties bevatten en evenredig gedistribueerd zijn, en hebben deze na etikettering met fluorescente kleurstoffen gehybridiseerd tegen wilde soorten, veredelingslijnen van tomaat, aardappel, aubergine en peper (een nauw verwante soort buiten de genus). We hebben syntenie aangetoond tussen deze soorten. Op deze manier hebben we vijf nog niet eerder beschreven inversies ontdekt en discrepanties met eerder gepubliceerde chromosomale herschikkingen gevonden. Onze resultaten kwamen goed overeen met de gepubliceerde fylogenie van *Solanum*, hetgeen suggereert dat onze aanpak geschikt zou zijn om onbekende genomen te bestuderen en om relaties op een hoger niveau, zoals tussen secties, op te lossen.

Tenslotte bespreekt dit proefschrift de cruciale punten van de FISH technologie, zoals ruimtelijke resolutie, detectiegevoeligheid en toepasbaarheid. Het laat de sterke en zwakke punten, en de mogelijkheden en beperkingen van FISH zien. Samengevat is FISH een onmisbare techniek voor het sequencen van grote genomen en het definieren van het gehalte aan repetitieve sequenties met ondersteuning van de bioinformatica. Bovendien kunnen verborgen chromosomale herschikkingen, in gebieden waar recombinatie onderdrukt wordt, zichtbaar gemaakt worden, hetgeen belangrijk is voor plantenveredeling en voor fylogenetische studies in het bijzonder.

Acknowledgements

I would like to thank all who contributed to my life and PhD study in Wageningen during the last four years. First of all I wish to mention the great help of my supervising professor, Dr. Richard Visser, who always encouraged me with the experimental work and the manuscript writing, and who built for me a bridge between the Laboratories of Plant Breeding and Genetics. I am thankful to my co-supervising professor, Dr. Yuling Bai, who kept track of my progress; pushed me where I needed and who gave me useful suggestions for the writing process of the thesis chapters. I also appreciate the help of my second co-supervisor, Dr. Hans de Jong for accepting me in his team, his patience in teaching me the secrets of science and chromosomes, having fruitful scientific discussions, and critical reading and correcting my thesis. Hans: thank you for trusting me and for supporting me in finding my way in an entirely new Dutch life and living.

I enjoyed always the good atmosphere in the cytogenetic group, which is a small, yet very powerful research team. I express my special appreciation for Erik Wijnker, who challenged me often in critical scientific thinking and reasoning, in logic writing and improving my English skills. Erik, it has been great fun to work with you and the moments outside of the lab with you and Ilse, and the cat as well. I also thank the other colleagues: Hetty, Henny, Penka, Xianwen and Xiaomin, who have always been nice to me. I am glad to thank those who provided the financial funds for my PhD, first CBSG (Center for BioSystems Genomics), then the European EU-SOL genome consortia, and finally Wageningen UR Plant Breeding laboratory. I appreciate the support from Prof. Willem Stiekema, Dr. René Klein Lankhorst and Dr. Roeland van Ham. I would like to acknowledge Dr. Sander Peters and Erwin Datema for their sequencing and bioinformatics expertise and support. I thank all my European partners in EU-SOL but especially I express my appreciation for Dr. Mondher Bouzayen and Murielle Philippot for their outstanding and precise work, and their warm friendship that I experienced last few years. I am thankful for Dr. Ronald van den Berg for his precise and great help in the *Solanum* phylogeny analysis. My gratefulness also goes to Dirk Jan Huigen, Fien Meijer-Dekens, Bert Essenstam, and Michel Hagendoorn, who kindly took care of my plant material. I mention the great help of Corrie Hanhart, who made 'lab life' easy by taking care of ordering chemicals and the nice and cheerful talks during the breaks. Boudewijn van Veen always was eager to solve any computer or fluorescence microscope problem and Marijke Hartog who is a kind neighbor from the Molecular Biology Laboratory who helped with molecular questions. I would like to thank for Dr. Douwe Zuidema, who was always kind and helpful with every questions related to the PhD study. I am glad to say thank you for the secretariat to Aafke van der Kooi, Corrie Eekelder, Wytske Nijenhuis and Annie Marchal for their great help with all kinds of financial, social and administrative problems. I express my special appreciation for my former boss Prof. Ervin Balázs at the Agricultural Research Institute of the Hungarian Academy of Science in Martonvásár, who gave me the first 'kick' in science and Dr. Gabriella Linc who encouraged me to go to Wageningen University. Finally, I mention all other great colleagues of the Laboratories of

Genetics and Plant Breeding, who were part of my daily life in the lab, at coffee breaks, the scientific meetings, or just a chat at the door.

I would like to extend my gratitude to colleagues who became my friends, Dr. Károly Pál who helped me to survive the first difficult moments in the Netherlands, Chunting Lang who always cheered me up and Dr. Christiane Kiefer whom I always could rely on. All of these would not have been carried out without my wonderful parents and my family who guided me along the way to reach my goals in the Netherlands and my study. *Drága Szüleim, köszönöm hogy mellettem álltok és álltatok jóban – rosszban és a távolság ellenére sem csökkent a köztünk lévő szeretet és összetartás.* Finally, and not in the least, I would like to thank my boyfriend – Jeroen – for his continuous and unsurpassed help, understanding and love towards me, and his great help to complete my PhD. You and your family are more than fantastic and give me the feeling of being so welcome in their Dutch life!

Education Statement of the Graduate School Experimental Plant Sciences

Issued to: Dóra Szinay
17 May 2010
Laboratory of Genetics, Plant Breeding, Wageningen University

1) Start-up phase	date
• **First presentation of your project**	
The power of FISH for tomato	Nov 29 , 2007
• **Writing or rewriting a project proposal**	
• **Writing a review or book chapter**	
Szinay D., Bai Y., Visser R. and de Jong H (2010): FISH applications for genomics and plant breeding strategies in tomato and other Solanaceous crops (Cytogenetics and Genome Research)	2009-2010
MSc courses	
GATC	2008-2009
Laboratory use of isotopes	
Subtotal Start-up Phase	10.5 credits*

2) Scientific Exposure	date
• **EPS PhD student days**	
PhD student day, Wageningen University	Sep 13, 2007
PhD student day, Leiden University	Feb 26, 2009
• **EPS theme symposia**	
Theme 4 symposia 'Genome Plasticity', Leiden University	Dec 07, 2007
Theme 4 symposia 'Genome Plasticity', Wageningen University	Dec 12, 2008
Theme 4 symposia 'Genome Plasticity', Radboud University Nijmegen	Dec 11, 2009
• **NWO Lunteren days and other National Platforms**	
NWO Lunteren days	Apr 02-03, 2007
NWO Lunteren days	Apr 06-07, 2009
• **Seminars (series), workshops and symposia**	
EU-SOL meeting, Rome	Nov 18-19, 2007
Symposium Flowers and Fruits of the Solonaceae	Jun 12, 2008
Revolution in Evolution? Epigenetics in Ecology and Evolution	Sep 18, 2009
• **Seminar plus**	
• **International symposia and congresses**	
Int. Chromosome Conference, Amsterdam	Aug 25-29, 2007

XVI.Plant & Animal Genome Conference, San Diego	Jan 12-16, 2008
Int. Solonaceae meeting, Cologne	Oct 11-16 , 2008
XVII.Plant & Animal Genome Conference, San Diego	Jan 10-14, 2009
• **Presentations**	
poster: FISH strategies for the genetics projects of tomato and potato	Apr 02-03, 2007
presentation: Multicolour BAC-FISH technology in tomato	Dec 07, 2007
poster: Applications of Multicolor FISH Technology For the Tomato and Potato Genome Projects	Jan 12-16, 2008
poster: Applications of FISH Technology for the ongoing Solanum sequencing project	Oct 11-16 , 2008
presentation: FISH characterization of two repeat types in tomato chromosomes	Jan 10-14, 2009
presentation: BAC FISH maps link comparative genomics with phylogeny and breeding in Solanum species	Apr 06-07, 2009
• **IAB interview**	Dec 02, 2009
• **Excursions**	
Subtotal Scientific Exposure	15.1 credits*
3) In-Depth Studies	date
• **EPS courses or other PhD courses**	
Plant Genetics : Natural variation	Aug 25-29, 2008
Molecular Phylogenies: Reconstruction & Interpretation	Oct 19-23, 2009
• **Journal club**	
Participation in our lab literature discussion groups (Cytogenetics lab and Plant Breeding)	2007-2009
• Indivi**dual research training**	
Subtotal In-Depth Studies	6.0 credits*
4) Personal development	date
• **Skill training courses**	
Information Literacy, including Introduction Endnote	Sep 11-12, 2007
Project and Time Management	Jan 22, Feb 05, March 04, 2008
Mobilising your scientific network	May 23, Jun 06, 2008
Guide to digital scientific artwork	Dec 15-16, 2009
Effective behaviour in your professional surrounding	Feb 23, March 23, 2009
TOEFL iBT	Spring 2009
Carreer Assessment	Nov 09, 2009
• **Organisation of PhD students day, course or conference**	
• **Membership of Board, Committee or PhD council**	

Subtotal Personal Development	8.5 credits*
TOTAL NUMBER OF CREDIT POINTS*	40.1

Herewith the Graduate School declares that the PhD candidate has complied with the educational requirements set by the

Educational Committee of EPS which comprises of a minimum total of 30 credits

* A credit represents a normative study load of 28 hours of study