Isolation, characterization and mapping of microsatellites from the tomato genome and their application in molecular analysis of centromeric regions

Dissertation

zur Erlangung des akademischen Grades
doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Mathematisch-Naturwissenschaftlich-Technischen Fakultät
der Martin-Luther-Universität Halle-Wittenberg

von Tatyana Areshchenkova
geboren am 1.1.1967 in Kiev, Ukraine

Gutachter:

1. Prof. Dr. Ulla Bonas
2. Dr. habil. Martin Ganal
3. PD Dr. Christiane Gebhardt

Eröffnung des Promotionsverfahrens: April 2000

Datum der öffentlichen Verteidigung: 11. Juli 2000

CONTENTS

List of abbreviations

A	adenine, ampere
AFLP	amplified fragment length polymorphism
Amp	ampicillin
Amp^r	ampicillin-resistant phenotype
bp	base pair
BLAST	basic local alignment search tool
BSA	bovin serum albumin
β-ME	β-mercaptoethanol
C	cytosine
cDNA	complementary DNA
cM	centiMorgans
CTAB	cetyltrimethylammonium bromide
DAF	DNA amplification fingerprinting
DNA	desoxyribonucleic acid
DTT	1,4-Dithiothreitol
E.coli	*Escherichia coli*
EDTA	ethylenediamine tetraacetic acid
EST	expressed sequence tag
FISH	fluorescence in situ hybridization
G	guanosine
g	gram
h	hour
HMW	high molecular weight
IPTG	Isopropylthio-β-D-galactopyranoside
k	kilo
kb	kilobase pair
l	litre
LOD	logarithm of the odds
LTR	long terminal repeat
M	molarity
Mb	megabase pair

m	milli, metre
min	minute
µ	micro
n	nano
OD	optical density
PCR	polymerase chain reaction
PFG	pulsed field gel
PFGE	pulsed field gel electrophoresis
pfu	plaque forming unit
QTL	quantitative trait loci
RAPD	randomly amplified polymorphic DNA
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rpm	rounds per minute
s	second
S	constant of sedimentation
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism
SSR	simple sequence repeat
T	thymine
T_m	melting temperature
TE	Tris/EDTA
TMV	tobacco mosaic virus
u	unit
UV	ultraviolet
V	volt
v/v	volume/volume
w/v	weight/volume
X-gal	5-bromo-4-chloro-3-indoyl-β-D-galactopyranoside
YAC	yeast artificial chromosome

1. Introduction

1.1 DNA markers for genome analysis

In the past two decades, molecular marker techniques have been developed as a direct result of the needs of genome analysis. These techniques range from molecular assays for genetic mapping, gene cloning and marker assisted plant breeding to genome fingerprinting and for the investigation of genetic relatedness. Genetic markers are based on DNA polymorphisms in the nucleotide sequences of genomic regions either defined by restriction enzymes, or two priming sites.

RFLPs

Among the various molecular marker techniques developed, restriction fragment length polymorphisms (RFLPs) were used to construct the first molecular map of the human genome (Botstein et al., 1980). This technique uses cDNA or other cloned single-copy DNA elements as radioactively labeled probes in hybridization with restricted genomic DNA. Usually, several endonucleases and different genotypes are screened. The combination of DNA probe and genotype-specific restriction enzyme pattern reveal a „restriction fragment length polymorphism“. RFLP is a reliable polymorphism which can be used for accurate scoring of genotypes. RFLPs are co-dominant and identify a unique locus and, therefore, are very informative. When cDNAs with known gene function are used as markers, the chromosomal position of the specific gene or genes can be identified. RFLP mapping together with molecular cloning of genes, set the stage for establishing syntenic relationships for a number of plant and animal species. Most comparative maps made to date have relied on RFLP analysis using cDNAs as a probes (Kowalski et al., 1994; Van Deynze et al., 1998; Brubaker et al., 1999; Livingstone et al., 1999; Edwards, 1994). In plants, RFLP remaines the most-widely used DNA marker assay, and is the basis for detailed genetic maps of major crops.

Although it remains widely-used, two basic limitations of the RFLP technique have motivated the development of several alternative technologies. The first limitation is the quantity of DNA required. 50-200 micrograms of DNA per individual are necessary to generate a DNA fingerprint or RFLP analysis of the entire genome. In contrast to RFLPs, PCR-based techniques developed during the last ten years require only approximately 10% of this amount, as template for PCR amplification of large quantities of the target sequence. The second limitation is that closely-related species usually contain the same alleles.

RAPDs (randomly amplified polymorphic DNA)

PCR-based techniques for detecting DNA markers require the development of specific DNA primers as a „start“ site for amplification. The widely-used RAPD analysis (Williams et al., 1990) relies on a single 10-base primer of largely-arbitrary sequence, except that primers are selected to have 60% or more G+C content, to obtain stronger binding to the template. PCR amplification would only be expected when the priming site occur twice in opposite orientation within approximately 2,000 bases. In theory, these conditions are met about five times in a higher eukaryotic genome. A similar technique, DAF is based on arbitrary primers of six bases (Baum et al., 1992). RAPD polymorphisms result from DNA sequence variation at primer binding sites and from DNA length differences between primer binding sites. The RAPD assay has alleviated some problems associated with RFLP and has been widely used in screening for DNA sequence-based polymorphisms at a very large number of loci, because it requires small amounts of DNA (15-25 ng), is a nonradioactive assay and can be performed in several hours.

RAPD fragments linked to a trait of interest could easily be identified by using two pooled DNA samples: one from individuals that express the trait, the other from individuals that do not. Any polymorphism between the two pools should be linked to the trait. Identified markers are subsequently confirmed by mapping in a segregating population (Michelmore et al., 1991). This technique, named bulk segregant analysis, is now widely used for mapping simple traits.

Once a marker has been linked to a trait of interest, it is relatively easy to convert the RAPD assay into a more reproducible PCR-type assay based on secondary DNA sequence information, by the use of allele-specific PCR (AS-PCR) or a sequence-characterized amplified region (SCAR) assay (Paran and Michelmore, 1993).

The unpredictable behavior of short primers which is affected by numerous reaction conditions, inheritance in a dominant manner and population specificity are the main disadvantages of RAPDs.

AFLPs

Amplified fragment length polymorphism (AFLP) is based on PCR amplification of restriction fragments generated by specific restriction enzymes and oligonucleotide adapters of a few nucleotide bases (Vos et al., 1995). This method generates a large number of restriction fragments (50-100) facilitating the detection of polymorphisms. The number of DNA fragments which are amplified can be controlled by choosing different base numbers and composition of nucleotides in adapters. Although not many maps have been developed so far using AFLPs, this method is now widely used for developing polymorphic markers. The approach is very useful in saturation mapping and for discrimination between varieties. High reproducibility, rapid generation and high frequency of identifiable polymorphisms make AFLP analysis an attractive technique for determining linkages by analysing individuals from segregating populations. However, AFLPs are predominantly not codominant and still expensive to generate because the fragments are detected by silver staining, fluorescent dye or radioactivity.

Microsatellites (SSRs)

Microsatellites, or simple sequence repeats (SSRs), simple sequence length polymorphisms (SSLPs), short tandem repeats (STRs), simple sequence motifs (SSMs), sequence target microsatellites (STMs), is a class of repetitive sequences which are widely-distributed in all eukaryotic genomes. They consist of arrays of tandemly repeated short nucleotide motifs of 1-4 bases, and are called mono-, di-, tri- or tetranucleotide repeats respectively. It had been known that such arrays of short DNA elements repeated in tandem tend to be imprecisely replicated during DNA synthesis, and generate new alleles with different numbers of repeating units. Variable number of repeats between individuals or array length is a result of slippage of the DNA polymerase during DNA replication (Tautz et al., 1986). This length variation is a source of polymorphisms even between closely related individuals. Such microsatellite sequences can be easily amplified by PCR using a pair of flanking locus-specific oligonucleotides as primers and detect DNA length polymorphisms (Litt and Luty, 1989; Weber and May, 1989). Further, these microsatellite markers use long PCR primers which are specific to a single genetic locus, they are codominant and, most importantly, they are multiallelic and detect a much higher level of DNA polymorphism than any other marker system. Such simple sequence length polymorphisms occur very frequently in a genome, and have proven to be extremely useful as DNA markers. The utilization of microsatellite sites as genetic markers had been proposed for genetic mapping of eukaryotes (Beckmann and Soller, 1990).

Each of these four primary genetic marker systems used in plant genetics not only differs in principle, but also in the type and amount of polymorphism detected. A comparison of the utility of each marker assay for germplasm analysis was performed by evaluating information content (expected heterozygosity), number of loci simultaneously analysed per experiment (multiplex ratio) and effectiveness in assessing relationships between soybean accessions (Powell et al., 1996). This study has demonstrated that all four types of marker assays have different properties. Microsatellites have the highest expected heterozygosity, while AFLPs are characterized by a very high multiplex ratio, RAPDs are intermediate in heterozygosity and multiplex ratio, while RFLPs have moderate heterozygosity and are uniquely appropriate for studies of synteny.

The utility of PCR-based AFLPs, RAPDs and SSRs marker systems in tetraploid outbreeder potato was examined by Milbourne et al. (1997) in terms of multiplex ratio and the amount of polymorphism detected (diversity index). SSRs were characterized by the highest diversity index which was enhanced by the codominant nature of the markers. AFLPs and RAPDs have very similar average diversity indices in potato that is consistent with findings by Powell et al. (1996) in the diploid inbreeder soybean. AFLPs had the highest multiplex ratio in both soybean and potato, but the value for RAPDs exceeded that of SSRs in potato and the order was reversed in soybean.

Informativeness and ease of genotyping are the most important criteria determining choice of assay. SSR analysis appears to be the most polymorphic marker system and dominates mammalian genome research and is likely to have a major positive impact on plant genome analysis and plant breeding programs. Recent developments in fluorescence-based marker technology offer the possibility of increasing the value of SSRs via multiplexing (the use of multiple primer sets labelled with different fluorophors and producing a range of allele sizes).

For many future genetic mapping studies and other applications, it will be necessary to characterize polymorphisms at a density higher than that available with microsatellite markers, which normally occur once every 6 kb in mammalian and every 30 kb in plant genomes.

SNPs

Polymorphisms corresponding to differences at a single nucleotide position (substitution, deletion, or insertion) occur approximately every 1.3 kb (Cooper et al., 1985; Kwok et al., 1996) and are referred to as single nucleotide polymorphisms or SNPs. Most polymorphisms of this type have only two alleles and are also called biallelic loci. There has been recent interest in the development of high-density linkage maps based on biallelic markers that can be assayed by PCR. With only two alleles, SNP markers are generally less informative than SSRs. Human genetic maps consisting of SNPs used in linkage studies need at least three times more markers than those containing SSRs at comparable resolution (Kruglyak, 1997). Although most SNPs reside within noncoding genomic regions, an important subset corresponds to mutations in genes that are associated with diseases or other phenotypes. Positional cloning based on SNPs may accelerate the identification of disease traits and a range of biologically informative mutations (Wang et al., 1998). The increasing production of genomic sequence data in conjunction with improved methods for SNP analysis are leading to the systematic generation of genetic maps consisting of SNPs.

1.2 Generation and use of microsatellites as genetic markers

Generation of microsatellite markers remains a relatively complex technique, because it requires isolation of flanking sequences specific for each microsatellite locus by the construction and screening of different types of libraries, including physically sheared and enzyme-digested genomic, cDNA libraries, and microsatellite enriched libraries. The process of SSR marker development includes following major steps:

- Construction of small-insert genomic libraries by cloning of a particular or specific size fraction of genomic DNA fragments into vector DNA. Phage and plasmid vectors with M13 priming sites are usually preferable, because M13 primers yield the most consistent high-quality sequence information. The size range of cloned inserts should be not more than 1,500 bp in order to facilitate complete sequencing.
- Detection of microsatellite-containing clones by hybridization with synthetic oligonucleotide probes complemetary to simple sequence repeats.
- Complete sequencing of microsatellite-containing clones and design of locus-specific oligonucleotide PCR primers (usually 20 bp long) in regions adjacent to the microsatellite.
- Amplification of the respective region from different sources of genomic DNA by PCR and detecting differences in the size of the amplified fragments using gel electrophoresis.

Therefore, this process is time-consuming and relatively expensive. Although these markers are costly to identify, they are relatively cheap and simple in further use. The development of

SSR markers has now been facilitated by the increasing amount of sequence information available in databases.

The use of microsatellite markers requires only small amounts of DNA and nonradioactive methods. PCR, product separation and data analysis can be almost completely automated, so that up to several hundred thousand data points can be generated in a relatively short time. Such high throughput analysis is of great interest for plant breeding applications. Microsatellites as locus-specific, multiallelic and highly informative molecular markers are extremely useful for genetic mapping and fingerprinting, variety identification and protection, monitoring of seed purity and hybrid quality, germplasm evaluation, population and evolutionary studies. Mapped microsatellite markers as an ordered set of sequence-tagged sites (STSs) provide a means for combining genetic linkage maps and physical maps (Olson et al., 1989).

A dense SSLP map containing 5,264 microsatellite loci already exists for human (Dib et al., 1996). A genetic map with over 7,300 SSRs at an average resolution of 1.1 cM has recently been constructed for mouse (Dietrich et al., 1996). Unfortunately, not many microsatellite markers have been developed in plants. In the last few years, initial studies on the frequences of microsatellite repeats and the generation of a number of markers have been reported for several plant species. The best-investigated plants regarding microsatellite marker isolation and mapping are mostly within the Gramineae. More than 120 microsatellites isolated from rice genomic libraries and sequence databases were integrated into the RFLP map and showed an even distribution throughout the 12 rice chromosomes (McCouch et al., 1997). A total of 279 loci amplified by 230 microsatellite markers were placed onto the wheat genetic linkage map. The markers were mapped to all three A, B and D genomes of wheat and were randomly distributed along the linkage map with some clustering in several centromeric regions (Röder et al., 1998b). Fourty SSRs have been isolated from barley genomic libraries and database sequences and mapped to seven barley chromosomes (Becker and Heun, 1995; Liu et al., 1996). This has been followed by the development of 568 new barley SSRs from enriched small-insert genomic libraries and from new sequences in the public databases. Using these markers, a second-generation SSR-based linkage map of barley genome has been constructed (Ramsay et al., submitted). Recently, an integrated genetic linkage map of the soybean genome consisting of 606 SSR, 689 RFLP, 79 RAPD, 11 AFLP, 10 isozyme, and 26 classical loci has been published (Cregan et al., 1999). Thirty-four highly informative microsatellite markers were isolated earlier from maize genomic libraries and eighteen were placed on a corn RFLP genetic map with relatively random distribution across the genome (Taramino and Tingey, 1996). To date, about 1,000 mapped microsatellite primer sets are available in maize database (http://www.agron.missouri.edu/ssr.html). Ninety-eight highly informative potato microsatellite markers were isolated from the EMBL nucleotide sequence database, cDNA and selectively enriched small-insert genomic libraries, and 65 markers were localized at 89 genetic loci across the twelve chromosomes giving reasonable genome coverage (Milbourne et al., 1998). A numbers of highly polymorphic markers were isolated and mapped for *Arabidopsis* (Bell and Ecker, 1994), *Sorghum bicolor* (Taramino et al., 1997), rapeseed (Kresovich et al., 1995), rye (Saal and Wricke, 1999) and some other plant species. In most of these cases, microsatellite markers were nearly randomly distributed throughout the genome based on genetic and physical data.

The number, abundance and composition of microsatellite repeats differ in plants and animals. The frequency of repeats longer than 20 bp has been estimated every 30 kb in plant genomes, while in mammals microsatellites occur every 6 kb. In general, plants have about ten times less SSRs than humans. In humans, GT is a very common repeat unit, but it is less abundant in plants, where AT repeats are much more abundant followed by A and GA motifs. The GT motif is significantly less abundant in plants compared to mammalian DNA (Weber, 1990). GT and GA microsatellites have been mostly used for the generation of markers for genetic mapping.

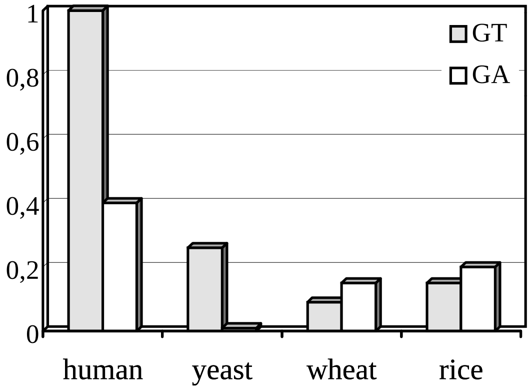


Figure 1. Relative abundancies of GT and GA microsatellite in different species based on plaque or colony hybridization with polyGT and polyGA oligonucleotides. The estimates are given per unit DNA in relation to human DNA (Tautz and Renz, 1984).

Tri- and tetranucleotide repeats occur at a much lower frequency. In plants, ATT repeats prevailed among trinucleotides and GATA among tetranucleotides. Distribution of microsatellites with 3-4 bases is different from that of dinucleotide repeating units. These repeats appear to be clustered in the regions of suppressed recombination in the genome such as centromeres and telomeres. Nonrandom distribution throughout genomes has been shown by the genetic mapping of the tetranucleotide motif GATA onto tomato (Arens et al., 1995) which are clustered around the tomato centromeres. Furthermore, *in situ* hybridizations with microsatellite motifs consisting of di-, tri- and tetranucleotide repeats in sugar beet (Schmidt and Heslop-Harrison, 1996) and trinucleotide repeats in the Gramineae (Pedersen and Linde-Laursen, 1994) suggest that some microsatellite repeats are highly clustered in genomes of some plant species. In contrary to di- and tetranucleotides, trinucleotide repeats were found to be frequent in coding regions (Smulders et al., 1997). Wang et al. (1994) found that as much as 57% of trinucleotide repeats rich in GC basepairs were located within coding sequences.

1.3 Tomato as a genetic system

The genus *Lycopersicon* includes the cultivated tomato (*L. esculentum* Mill.) together with its wild relatives. The wild species bear a wealth of genetic variability. Less than 10% of the total genetic diversity in the *Lycopersicon* gene pool is found in *L. esculentum* (Miller and Tanksley, 1990). The center of diversity for tomato is located in western South America, and the cherry tomato *L. esculentum* var. *cerasiforme* is considered as the most likely ancestor of cultivated tomatoes. Karyotypes of the *Lycopersicon* species are very similar with little or no structural difference among species (Barton, 1950).

As a crop plant, tomato is one of the best characterized plant systems. It has a relatively small genome of 0.95 pg or 950 Mb per haploid nucleus (Arumuganathan and Earle, 1991), and features such as diploidy, selfpollination, and a relatively short generation time make it amenable to genetic analysis.

Classical genetics has created one of the largest stocks of morphological mutations induced by radium, X-rays, UV-light, neutrons and chemical mutagenesis. A major contributor in the mutagenesis area was Hans Stubbe who developed over 300 *L. esculentum* mutants and 200 in *L. pimpinellifolium* (for a summary see Rick, 1975). A particularly interesting example of induced mutagenesis was the directed manipulation of fruit size of *L. esculentum* and *L. pimpinellifolium* (Stubbe, 1971).

A considerable proportion of these mutations have been mapped onto the classical genetic map. By 1988, the classical linkage map of the tomato genome comprised of 233 morphological and isozyme loci. An additional 86 have been assigned to their respective chromosomes via two-point or trisomic tests. The number of mapped genes in the form of cDNAs has increased considerably with the introduction of RFLP markers. The current tomato RFLP map was constructed using an F_2 population of the interspecific cross *L.esculentum* x *L.pennellii* and contained more than 1030 markers which were distributed over 1276 cM (Tanksley et al., 1992). A number of morphological and isozyme markers have also been mapped with respect to RFLP markers orienting the molecular linkage map with both the classical morphological and cytological maps of tomato. An integrated high-density RFLP-AFLP map of tomato based on two independent *L.esculentum* x *L.pennellii* F_2 populations has been constructed (Haanstra et al., 1999). This map spanned 1482 cM and contained 67 RFLP and 1175 AFLP markers. Both RFLP and AFLP maps show clusters of markers associated with almost all centromeres and some telomeres indicating that recombination is suppressed in those regions.

The current tomato map is considered to be complete. All molecular and classical markers could be mapped to linkage groups indicating that no loci failed to link up with the map.

The average relationship between genetic and physical distance in tomato is about 750 kb per cM. The actual ratio of genetic and physical distance varies considerably depending on the chromosomal region. High-resolution genetic and physical mapping around the *Tm-2a* region, which is located close to the centromere of chromosome 9, indicates that one cM in this area corresponds to more than five million base pairs (Pillen et al., 1996), approximately a sevenfold suppression of recombination over the expected value based on the estimated physical size of the region. In contrast, map-based cloning of the *chloronerva* gene which is involved in iron uptake and located in euchromatin of chromosome 1 demonstrated, that the ratio of genetic to physical distance in the *chloronerva* region is 160 kb per 1 cM (Ling et al., 1999) suggesting much higher levels of recombination in this area of the genome. By determining frequency and distribution of recombination nodules on tomato synaptonemal complexes, Sherman and Stack (1995) observed a much lower frequency of recombination nodules in heterochromatic regions around the centromeres compared to euchromatin. Suppression of recombination near the centromeres and higher values of recombination in distal chromosomal regions were also observed in potato (Bonierbale, 1988; Tanksley et al., 1992) and many other plant and animal species.

The tomato genome at the DNA level is comprised of approximately 78% single copy sequences, as evaluated under high stringency hybridization conditions (Zamir and Tanksley, 1988). In other plant species with large genome sizes, such as wheat or pea the single copy fraction is less than 20%, and in barley and rye it is less than 50%. The remaining part of the tomato sequences is repetitive DNA of which four major classes have been characterized. Ribosomal DNA represents the most abundant repetitive DNA family and comprises approximately 3% of the tomato genome. Both 5S and 45S rRNA genes are tandemly repeated with 1,000 and 2,300 copies and map to single loci on chromosome 1 and 2 respectively (Vallejos et al., 1986; Lapitan et al., 1991). As confirmed by *in situ* hybridization, a 162 bp tandem repeat TGRI with 77,000 copies in the genome is localized within a few hundred kb of the terminal 7 bp telomeric repeat TT(T/A)AGGG at 20 of 24 chromosome ends (Ganal et al., 1988) and, in addition, is also found at a few centromeric and interstitial sites (Lapitan et al., 1989; Ganal et al., 1992). Two other tomato genomic repeats, TGRII and TGRIII, are less abundant with 4,200 and 2,100 copies respectively. TGRII is apparently randomly distributed with an average spacing of 133 kb, and TGRIII is predominantly clustered in the centromeric regions of chromosomes. Except TGRIII, these repeats are only present in *Lycopersicon* species (Ganal et al., 1988). Zamir and Tanksley (1988) also reported a positive correlation between copy number and rate of divergence of repeats among DNA sequences from related solanaceous species. The more highly repeated sequences evolve more rapidly, whereas single copy coding regions are more conserved among different species. 43% of cloned low copy telomere-homologous sequences which were mapped near the tomato centromeres, hybridized to DNA from *L. esculentum* but not to *L. pennellii*, whereas single copy probes hybridized to both *L. esculentum* and *L. pennellii* (Presting et al., 1996) indicating rapid evolution of centromere-proximal sequences.

Cytologically, the centromere of higher eukaryotes is a constriction on condensed metaphase chromosomes surrounded by large blocks of pericentric heterochromatin. At the primary constriction various proteins associate with the centromeric DNA and form the kinetochore, the attachment point for the spindle apparatus. Thus, centromeres composed of both DNA sequences and proteins organized in a structurally and functionally unique manner and are complex genetic loci. Kinetochore protein components appear to be more conserved than centromeric sequences. As a general rule, plant centromeric DNA is heterogenous, composed of megabases of satellite DNA with poor conservation of primary repeat sequence across distantly related plant species, and includes also low-copy sequences, transposable elements and telomere-similar repeats. Each tomato chromosome has heterochromatin concentrated around its centromere. Using Feulgen densitometry and SC karyotype data, it was determined that 77% of the DNA in tomato pachytene chromosomes is packaged in heterochromatin which is similar to an earlier estimate (75.3%) in mitotic metaphase chromosomes (Peterson et al., 1996). In association with findings of Zamir and Tanksley (1988), these data suggest that a large fraction of tomato heterochromatic DNA is composed of single- and/or low-copy sequences and makes tomato heterochromatin unusual and probably genetically active.

The approximate map position of the centromere is now known for each tomato chromosome. For chromosomes 1 and 2, the centromere positions have been identified by RFLP mapping and by *in situ* hybridization with 5S rDNA and 45S rDNA respectively (Lapitan et al., 1991; Tanksley et al., 1988). The centromeres of chromosomes 3 and 6 have been located on the integrated molecular-classical map and by deletion mapping (Van der Biezen et al., 1994; Van Wordragen et al., 1994). Since there is evidence that the potato/tomato inversions on chromosomes 5, 10, 11 and 12 involve entire chromosome arms, the respective centromeres are most likely located at the inversion breakpoints (Tanksley et al., 1992). Map positions of the centromeres of chromosomes 4 and 8 were predicted based on the relationship among the cytological, genetic and molecular tomato maps. A more precise localization of the centromeres of chromosomes 7 and 9 has been achieved by RFLP hybridization and dosage analysis of telo-, secondary and tertiary trisomic stocks (Frary et al., 1996).

Despite their functional importance, the molecular characteristics of the centromeres of higher eukaryotes remain ill-defined. The most extensively studied DNA sequence is the 171 bp alpha satellite sequence which is located exclusively at the primary constriction of human chromosomes and thought to play a major structural and/or functional role at human centromeres (Haaf et al., 1992; Harrington et al., 1997). So far, no plant DNA sequences essential for centromere function have been identified. The only plant sequence to which a centromere function could be attributed is a DNA repeat from the centric region of the maize B chromosome. Sequence analysis revealed that this repeat contains several motifs. One is a stretch of repeats which has high similarity to the telomeric repeat of plants. The other has a significant homology to the 180 bp repeat that comprises the telomeric heterochromatic knob. Under certain conditions such knobs can function as spindle attachment sites and form neocentromeres (Alfenito and Bichler, 1993). Two repetitive sequences CentA and CentC were characterized in the centromeric region of the maize chromosome 9 (Ananiev et al., 1998). CentA has a structural similarity to retroelements and CentC is a tandem repeat which forms clusters of different sizes at centromeric sites of all maize chromosomes without obvious homology to the maize knob-associated tandem repeat.

Several types of DNA sequences located at pericentromeric regions have been identified (Martinez-Zapater et al., 1986; Ganal et al., 1988; Maluszynska and Heslop-Harrison, 1991; Richards et al., 1991; Aragón-Alcaide et al., 1996; Jiang et al., 1996; Presting et al., 1996; Thompson et al., 1996a,b), but their role in centromere function remains elusive.

Centromere research has a potentionally important application in the production of artificial chromosomes for use as plant cloning vectors. Studying of plant centromeric DNA sequences provides an opportunity to look for conserved structural patterns or primary nucleotide sequence motifs that may contribute to centromere function.

Recently, with the use of microsatellites in genome mapping interest has been directed to the location of microsatellite sequences on tomato chromosomes. Microsatellite polymorphism and genomic distribution were studied by fingerprinting of the tomato genome using labelled oligonucleotide probes complementary to GATA or GACA microsatellites. The copy number and the size of microsatellite containing restriction fragments were highly variable between tomato cultivars (Vosman et al., 1992). The mapping of individual fingerprint bands containing GATA or GACA microsatellites showed predominant association of these repeats with tomato centromeres (Arens et al., 1995). Structure, abundance, variability and location were evaluated for a number of different simple sequence repeats isolated from genomic libraries (Broun and Tanksley, 1996). Ten generated microsatellite markers (6(GT)n, 3(GA)n and 1(ATT)n) were tested for polymorphism in a set of ten tomato cultivars. Only two microsatellite loci ($(GA)_{16}$ and an imperfect ATT repeat) did reveal significant polymorphism in the tested cultivars and were mapped on the high resolution molecular map near the putative centromeres of chromosomes 3 and 12 respectively. In addition, nine polymorphic GATA-containing RFLP fragments (9-15 kb) were scored as dominant markers and mapped within clusters of markers adjacent to centromeres. Centromeric location was also observed for six polymorphic GATA-microsatellite loci which were analysed in a cross between *L.esculentum* and *L.pimpinellifolium* (Grandillo and Tanksley, 1996). Thus, GATA clusters are now known to be located in most if not all of the centromeric regions of the tomato chromosomes. A number of polymorphic microsatellite markers generated from database sequences have been used succesfully for genotyping tomato cultivars and accessions (Smulders et al., 1997; Bredemeijer et al., 1998) but their map positions have not been published to date.

1.4 Research objectives

The utility of microsatellites as tools for genetic mapping and genome analysis has been proven in human, animals and plants, because these PCR-based markers detect a much higher level of DNA polymorphisms than any other marker assay and these markers are amenable for automation. In the recent years, microsatellite markers were integrated into the existing molecular maps of a number of plant species and successfully used for gene mapping, study of genetic diversity, population and evolutionary studies.

Cultivated tomato is well-known for its low level of DNA-polymorphism, because of this, highly polymorphic markers are required for genome analysis within *L. esculentum*. Until now, microsatellite markers consisting of GATA repeats and dinucleotide motifs extracted from sequence databases have been studied (Arens et al., 1995; Broun and Tanksley, 1996; Smulders et al., 1997). However, no detailed studies regarding microsatellites directly extracted from the tomato genome by clone isolation have been reported and basically no microsatellite markers have been mapped onto the genetic map.

The goal of this research was to investigate the organization of simple sequence repeats in the tomato genome and to evaluate how useful these sequences might be as genetic markers in tomato. For that, this work was aimed at the isolation and mapping of a number of microsatellite containing sequences from genomic libraries. The investigation on abundance and structure of microsatellite repeats in tomato and their allelic variation and genomic distribution should extend our knowledge about this class of tandem repeats in the tomato genome. If tomato microsatellites are associated with centromeric regions like the previously isolated GATA repeats, they could be used to characterize tomato centromeric sequences in more detail. A high level of variability and assignment to the linkage map would provide additional opportunities for applications of newly developed markers in genome analysis and breeding of tomato.

2. Materials and Methods

2.1 Plant material

For the construction of tomato genomic libraries and as a reference stock the *L. esculentum* cv. VFNTcherry was used. The survey of tomato cultivars included the following *L. esculentum* varieties: TA55 (VF36 *Tm2a*); Moneymaker, Momor and Monita (related by decent, Laterrot, 1987); Rio Grande; New Yorker; Piline and Fline (related by decent, Laterrot, 1989); Angela; Puz11 and TA205. *L. esculentum* TA55 and *L. pennellii* accession LA716 were the parents of the standard tomato mapping population. All seed material was originally obtained from the USDA germplasm center at Geneva, New York, USA or the Tomato Genetics Stock Center at the University of California-Davis, USA.

For genetic mapping, DNA from fourty-three segregating F_2 plants derived from the interspecific cross between the two inbred accessions *L.esculentum* TA55 and *L.pennellii* LA716 (Tanksley et al., 1992) was used.

2.2 Phage, bacterial and yeast strains

Phage strains

Lambda Zap Express^{TM} vector arms	*Eco*RI and *Bam*HI predigested and phosphatased phage arms	The Zap Express^{TM} vector has 12 unique cloning sites and can accomodate DNA inserts of 0-12 kb, which can be excised out of the phage in the form of the kanamycin-resistant pBK-CMV phagemid	Stratagene
ExAssist ^{TM}	contains an amber mutation that prevents replication of the phage genome in a nonsuppressing XLOLR cells	Filamentous (M13) interference-resistant helper phage for excision of the pBK-CMV phagemid vector from the Lambda ZAP express vector	Stratagene

Bacterial and yeast strains

K 802	*supE hsdR gal metB*	A suppressing strain used to propagate bacteriophage lambda vectors and their recombinants	Stratagene
XLOLR	*hsdR endA1 thi-1 recA1 lac* λ^r *Su*$^-$	*E.coli* lambda resistant nonsuppressing host strain for plating excised phagemids	Stratagene
Epicurian Coli XL10-Gold	*endA1 recA*$^-$ *McrF supE*	*E.coli* ultracompetent cells with the Hte phenotype which increases transformation efficiency of ligated DNA	Stratagene
S.cerevisiae AB1380	*MATa* ϕ^+ *ura3 trp1 ade2-1*	yeast host strain for pYAC4 vector, chromosome preparation was used for size determination of YAC clones	Burke et al., 1987

2.3 Vectors and primers

Plasmids

pUC18 *Bam*HI/BAP pUC18 *Eco*RI/BAP	predigested vector used for cloning and sequencing of foreign DNA, Amp^r	Amersham Pharmacia
pBluescript II SK	phagemid vector used in standard PCR	Stratagene
pYAC4 vector	11.5 kb, *ori*, Amp^r, *Eco*RI	Burke et al., 1987
pBR322	ϕ^+ *ura3 trp1 ade2-1 can1-100 lys2-1 his5* Amp^r	Bolivar et al., 1977

Sequencing and PCR primers (Pharmacia Biotech)

MVL	5´-GCCGCTCTAGAAGTACT-3´
MVR	5´-CTAAAGGGAACAAAAGC-3´
M13-20 Universal primer	5´-GTAAAACGACGGCCAGT-3´
M13 Reverse primer	5´-GGAAACAGCTATGACCATG-3´
KS II	5´-CGAGGTCGACGGTATCG-3´
SK L	5´-CGCTCTAGAACTAGTGGATC-3´
KS I	5´- TCGAGGTCGACGGTATC-3´
SK S	5´- GCCGCTCTAGAACTAGTG-3´
T3	5´- AATTAACCCTCACTAAAGGG-3´
M13F L	5´- CGTTGTAAAACGACGGCCAGT-3´
M13R XL	5´- GGAAACAGCTATGACCATG-3´

2.4 Plant DNA extraction

Total genomic DNA was isolated from leaf tissue according to Bernatzky and Tanksley (1986). 20-30 g of tomato leaves were homogenized in 150 ml of DNA extraction buffer (0.35M Sorbitol, 0.1M Tris-HCl, 5mM EDTA-Na salt, pH 7.5, and 20mM sodium bisulfite). The homogenized suspension was filtered through miracloth into centrifuge bottles on ice and centrifuged 15 min. at 2,000 rpm and 4°C. The pellet was resuspended in 5 ml of extraction buffer and 5 ml of nuclei lysis buffer (200mM Tris-HCl, 50mM EDTA, 2M NaCl, 2% w/v CTAB) and then 2 ml of sarcosyl (5% w/v) were added. After gentle mixing, lysis was performed for 20-30 min. at 50°C, and subsequently DNA was extracted with 15 ml of chloroform:isoamyl alcohol (24:1). After centrifugation for 15 min. at 3,000 rpm, the aqueous phase was transferred into a new tube. The DNA was precipitated with 2/3 volume of isopropanol and hooked out with a bend pasteur pipette into 5-10 ml of 70% ethanol. The isolated DNA could be stored in this way or centrifuged, dried for a short time, and dissolved at 50-60°C in 1 ml of TE (10mM Tris-HCl, 1mM EDTA, pH 8.0) buffer. The DNA concentration was approximately 250-500 µg/ml.

2.5 Construction of tomato genomic libraries

2.5.1 Lambda libraries construction and screening

Lambda library construction

For the isolation of microsatellite containing sequences from tomato genome, Lambda ZapExpress libraries were constructed. Tomato cultivar VFNTcherry total DNA was digested with *Mbo*I (GibcoBRL).

Digestion reaction:	VFNT cherry DNA	- 15µg
	React2 (GibcoBRL)	- 10µl
	40mM Spermidine	- 6µl
	10mM DTT	- 4µl
	*Mbo*I (10u/µl)	- 3µl
	H_2O to a final volume of 100µl	
	at 37°C for 4h	

Restriction fragments were purified by precipitation with 1/10 volume of 3M sodium acetate and 2.5 volumes of absolute ethanol. After centrifugation at 12,000 rpm at room temperature for 10 min., the air dried DNA pellet was dissolved in water to the final concentration of 200ng/µl. Purified restriction fragments were ligated into the *BamH*I site of the Lambda ZapExpress vector arms (Stratagene).

The ligation reaction of the insert into the Zap Express vector was performed as follows:

Lambda Zap Express^{TM} vector arms predigested with *BamH*I (1µg/µl)	- 1µl
10x ligation buffer (Boehringer Mannheim)	- 0.5µl
insert (50-200 ng)	- 1µl
T4 DNA ligase (5u/µl, Boehringer Mannheim)	- 0.5µl

H_2O to a final volume of 5µl

at 14°C overnight

Packaging of recombinant bacteriophage lambda DNA into infectious particles in vitro (Gigapack II Gold packaging extract, Stratagene), phage propagation in *E.coli* host strain, and titering of the library was performed according to the instruction manual of Zap Express vector cloning kit (Stratagene).

Recombinant lambda phages were stored in SM buffer with 0.3% (v/v) chloroform at 4°C, propagated using bacterial host strain K802 grown in NZY broth and plated onto NZY Top Agar medium.

SM buffer (per liter):

5.8g	NaCl
2g	$MgSO_4$ $7H_2O$
50ml	1M Tris-HCl (pH 7.5)
5ml	2%(w/v) gelatine

NZY broth (per liter):

5g	NaCl
2g	$MgSO_4$ $7H_2O$
5g	yeast extract
10g	NZ amine (casein hydrolysate)
pH 7.5	

NZY Top Agar (per liter):

1l NZY broth

0.7% (w/v) agar

Phage library screening

Approximately 10^5 pfu were mixed with fresh K802 cells grown in NZY broth (overnight bacterial culture OD_{600}=0.5), incubated at 37°C for 15 min. and plated onto NZY top agar plates for propagation at 37°C. Plaques (zones of lysis) became visible after 8 hours.

Plaques were screened by hybridization with ^{32}P-labelled probes. Hybond N+ membranes (Amersham Buchler) were gently placed onto the surface of the top agarose. Bacteriophage particles and DNA were transfered to the filter by capillary action in an exact replica of the pattern of plaques. After denaturation with 0.4N NaOH, the DNA bound to the filter was hybridized with the radioactively labelled oligonucleotide probes poly(GA/CT) and poly(GT/CA) (Pharmacia) or a synthetic $GATA_{12}$ probe. After washing, membranes were placed between the film and intensifying screen and kept at -80°C for 2-3 days. Hybridizing plaques, identified by aligning the film with the original agar plate, were picked for further analysis.

In vivo excision

After performing secondary and tertiary screenings, purified positive single plaque phage clones were *in vivo* excised using the ExAssist helper phage/XLOLR system according to the *in vivo* excision protocol (Stratagene). The designed structure of the Lambda Zap Express vector allows efficient *in vivo* excision and recircularization of any insert cloned into the lambda vector to form a phagemid containing the cloned insert. Simultaneous infection of *E.coli* with the recombinant lambda vector and the helper phage results in single-stranded DNA synthesis through the cloned insert forming a circular DNA molecule - the pBK-CMV phagemid vector with the insert DNA. Newly created phagemids can be „packaged“ and secreted from the *E.coli* cells. The ExAssist helper phage contains an amber mutation that prevents replication of the phage genome in nonsuppressing *E.coli* XLOLR cells. This allows only the excised phagemid to replicate in the host, removing the possibility of co-infection with the ExAssist helper phage. For sequencing, DNA from excised recombinant pBK-CMV phagemid clones was extracted using Qiagen-tip 20 according to the plasmid mini protocol (QIAGEN).

Sequencing

Recombinant plasmid clones were sequenced with MVL (forward) and MVR (reverse) primers. Sequences containing microsatellites were used for primer design.

2.5.2 Construction and screening of plasmid libraries

Plasmid library construction

In order to enrich the library with single-copy sequences, total genomic DNA from tomato variety VFNTcherry was predigested with the methylation-sensitive restriction enzyme *Pst*I (GibcoBRL).

Digestion reaction:	VFNT cherry DNA	- 20μg
	10x React2	- 10μl
	40mM Spermidine	- 6μl
	10mM DTT	- 8μl
	*Pst*I (10u/μl)	- 3.5μl
	H_2O to a final volume	100μl
	at 37°C for 4h	

Restriction fragments in the size range from 2 and up to 9 kb were purified from an 1% agarose gel using the Geneclean kit (Dianova). These size-selected DNA fragments were further digested with *Sau*3AI (GibcoBRL), *Bam*HI (GibcoBRL), or *Acs*I (Boehringer Mannheim).

Digestion reaction:	2-9 kb fraction of VFNT cherry *Pst*I DNA fragments	- 35μl
	10x buffer	- 5μl
	40mM Spermidine	- 3μl
	10mM DTT	- 2μl
	*Sau*3AI, *Bam*HI, or *Acs*I (10u/μl)	- 1.8μl
	H_2O to a final volume	50μl
	at 37°C (for *Acs*I - 50°C) for 4h	

After digestion with *Sau*3AI or *Bam*HI, DNA fragments were precipitated with ethanol and ligated into the pUC18 *Bam*HI/BAP vector (Amersham Pharmacia) and after digestion with *Acs*I into the pUC18 *Eco*RI/BAP vector (Amersham Pharmacia).

Ligation reaction:	VFNTcherry DNA fragments (70ng/μl)	- 3.5μl
	pUC18 vector (0.5μg/μl)	- 0.5μl
	10x ligation buffer	- 0.5μl
	T4 DNA ligase (5u/μl, Boehringer Mannheim)	- 0.5μl
	at 14°C overnight	

The high-efficiency *E.coli* XL10-Gold ultracompetent cells (Stratagene) were used for transformation. Transformations were performed using 150µl aliquots of cells, 3µl of β-ME (Stratagene), and 2µl of the ligation mixture. Cells were heat-pulsed for 30 seconds in a 42°C water bath. After incubation at 37°C for one hour with shaking at 225-250 rpm, transformants were plated onto 22x22cm plates with LB-Amp-agar medium and incubated overnight at 37°C.
Using the automated colony picking system BioPick (BioRobotics Ltd.), single colonies were transfered into 384-well (24 rows x 16 columns) microtiter plates. Each well contained 40µl of the mixture of growth medium 2YT and 10x freezing medium HMFM in a ratio of 9:1.

LB-Amp-agar (per liter):
10g NaCl
10g bacto tryptone
5g yeast extract
20g agar
H_2O
pH 7.0
Autoclave, cool to 55°C and
add 1ml of 75mg/ml filter-sterilized Amp

2YT (per liter):
16g bacto tryptone
10g yeast extract
5g NaCl
H_2O
pH 7.0

HMFM (1x concentration):

36 mM	K_2HPO_4
13.2 mM	KH_2PO_4
0.4 mM	$MgSO_4$
1.7 mM	Na_3-citrate
6.8 mM	$(NH_4)_2SO_4$
4.4% (v/v) glycerol	

Freshly inoculated 384-well plates were incubated at 37°C overnight and then frozen at -80°C until further use. High-density colony filters for hybridization were prepared from thawed 384-well microtiter plates using the robotic gridding system BioGrid (BioRobotics Ltd.). Each of the 384-well plates was replicated twice onto a sterile LB-Amp-Agar plate covered with 22cm square nylon filter (Hybond N+, Amersham Buchler). Nylon filters were carefully labeled to indicate orientation, side, number, and date. Each clone was inoculated on two defined positions within the high-density arrays, because such double inoculations aid in distinguishing true positive clones from false positive hybridization signals and in correct identification of the precise well positions of positive clones. Plates were incubated at 37°C until the colonies were clearly visible (ca. 1mm in diameter) but did not merge together. Filters were removed from the agar surface and placed, colony-side-up, onto Whatman 3mm paper saturated with denaturing solution of 1.5M NaCl, 0.5M NaOH for two minutes, and then neutralized for five minutes in 1.5M NaCl, 0.5M Tris-HCl, pH 8.0, rinsed in 2xSSC (0.3M NaCl, 0.5M sodium citrate) for a few minutes and used for hybridization (as described below). One high-density nylon filter was containing the immobilized DNA from 18,432 clones spotted out of fourty-eight 384-well microtiter plates (4x4 grid).

Plasmid library screening

High-density filters were hybridized with $(GATA)_{10}$, $(GA)_n$ and $(GT)_n$ oligonucleotide probes at 65 °C and washed to a stringency of 0.5xSSC, 0.1%SDS. Clones that gave a signal were identified by the alignment of the autoradiograms with an appropriate grid indicating the position of all colonies. Positive clones were streaked onto LB-Amp-agar plates, and after incubation overnight at 37°C, a single colony was used for plasmid DNA isolation according to the plasmid mini protocol (QIAGEN).

2.6 Database searches

Computer searches were performed using the TIGR (The Institute for Genome Research) Tomato Gene Index (LGI, release version 1.2) software. TIGR Gene Indices (http://www.tigr.org/tdb/lgi/index.html) which included a database of tomato EST sequences, were searched with a minimum number of 10 for all types of dinucleotide, trinucleotide and GATA repeats using the WU-BLAST 2.0 search program.

2.7 DNA sequencing and sequence analysis

Plasmid clones were sequenced at the IPK Gatersleben on automated laser fluorescence (ALF) DNA sequencer (Pharmacia). DNA sequences were detected using fluorescein-labelled primers by the dideoxynucleotide chain termination method (Sanger et al., 1977) and the Autoread Sequencing kit (Pharmacia).

Sequences were compared for identity using DNASIS v2.5 software. Homology search against sequences in the nucleic acid database of the GenBank+EMBL+DDBJ+PDB sequences were made using BLASTN 2.0.9 (Altschul et al., 1997).

2.8 Primer design

Oligonucleotide primer pairs were designed from unique sequences flanking the microsatellite motifs by using the computer program Primer 0.5 (provided by E. Lander, Whitehead Institute, USA).

To obtain stronger binding to the template, primers were selected to have a G+C content of approximately 50% (melting temperature T_m of 60°C) and to be 18-23 nucleotides long. The T_m was limited by 50 and 65°C, and the difference in T_m between two primers within a primer pair was not more than 3°C.

Primers were synthesized by Pharmacia Biotech or by MWG Biotech. One primer was labelled with fluorescein for fragment analysis on an ALF sequencer.

Isolated microsatellite markers were designated as TMS (tomato microsatellite) and numbered consecutively.

2.9 PCR analysis

PCR reactions were performed in a volume of 25µl in Perkin-Elmer (Norwalk, CT) thermocyclers. The reaction mixture contained:

10 mM	Tris-HCl
50 mM	KCl
1.5mM	$MgCl_2$
0.2mM	of each deoxynucleotide
1u	*Taq* DNA polymerase
250 nM	of each primer
50-150 ng	template DNA.

PCR reactions on genomic, YAC and on plasmid DNA were performed with an initial denaturation step of 3 minutes at 94°C, followed by

	genomic and YAC profile			plasmid profile	
45 cycles	94°C	- 1 min.,	25 cycles	94°C	- 30 s.,
	50, 55, or 60°C	- 1 min.,		50°C	- 30 s.,
	72°C	- 2 min.,		72°C	- 1 min.,

and a final extension step at 72°C for 10 min., for 7 min.

2.10 Separation of PCR products on ALF

Fragment analysis was carried out on an ALF DNA sequencer (Pharmacia) using short gel cassettes. Denaturing 6% polyacrylamide gels, 0.35mm thick, were prepared using SequaGel XR solutions (Biozym).

Internal standards were prepared as follows: pBluescript II SK template DNA was amplified by PCR with different primer combinations. One primer in the primer pair was labeled with fluorescein. Plasmid DNA was denatured at 94°C for 3 min. and amplified in 25 cycles under the following conditions: denaturing at 94°C for 30s., annealing at 50°C for 30s., polymerizing at 72°C for 1 min. and a final extension step at 72°C for 7 min. Primer combinations and PCR product sizes are listed in Table 1.

Table 1. Standard pBluescript II SK DNA amplification by PCR.

Primer pair	Labeling	PCR product size, bp
KS II SK L	*	70
KS I SK S	 *	73
KS II T3	*	122
M13F L T3	*	196
M13F L M13R XL	*	231

25µl of two standard PCR products were mixed with 0.5ml of loading buffer (5mg/ml Blue dextran in deionized formamide). 3µl of loading buffer which contained 0,15µl of each of the two standard fragments and 1.0-1.5 µl of the PCR reaction which product should be detected were loaded onto the gel. Gels were run in 1xTBE (0.09M Tris-borate, 0.002M EDTA) buffer with 600V, 40mA and 50W with 2mW laser power with a sampling interval of 0.84s.

Fragment sizes were determined using the computer program Fragment manager version 1.2 (Pharmacia) and internal size standards (Fig. 2).

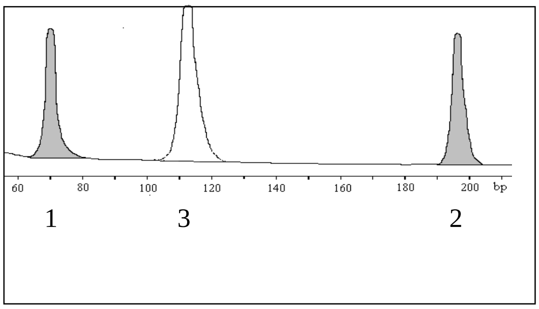


Figure 2. Detection of the PCR product size relative to two fragments with known sizes. 1- standard 1, 70 bp fragment amplified on pBluescript II SK plasmid DNA with KS II and SK L primer pair; 2- standard 2, 196 bp fragment amplified on pBluescript II SK plasmid DNA with M13F L and T3 primer pair; 3- the DNA fragment to be detected. An automatically calculated calibration curve based on the two size standards permits accurate size determination of the amplified product.

2.11 Allele detection and genetic mapping

SSR polymorphism is a result of the variation in the number of repeats in different individuals. It can be detected by PCR amplification of genomic DNA from different sources, and differences in the size of the amplified products can be analysed by gel electrophoresis. The polymorphism level of microsatellite markers in tomato was surveyed in a test set of twelve *L.esculentum* cultivars and *L.pennellii* LA716.

The map position of each SSR locus was determined on the existing high-density RFLP map of tomato (Tanksley et al., 1992) using the MAPMAKER v2.0 software (Lander et al., 1987). Markers detecting polymorphism between *L.esculentum* TA55 and *L.pennellii* LA716 were scored for parental alleles amplified from DNA of the segregating population and were assigned to the RFLP framework using the „place“ (chromosome linkage) and the „try“ (marker linkage) commands. The genetic linkage was detected with LOD score > 3.0. The Kosambi mapping function (Kosambi, 1944) was used for converting recombination frequency to map distances.

2.12 Genetic diversity estimations

Some of microsatellite markers were chosen to estimate genetic diversity within twelve tomato cultivars and *L.pennellii* LA716. The presence (1) or absence (0) of each amplified fragment was scored in a binary data matrix. Genetic distance was calculated for each pair of lines using the percentage difference in the program NCLAS of the computer package SYN-TAX IV (Podani, 1990), according to the equation: PD=1-2Nij/(Ni+Nj), where Nij is the number of fragments common to accessions i and j, and (Ni+Nj) is the total number of fragments in both accessions. This value is between 0 and 1 with a score of 0 indicating that all fragments are in common, and 1 indicates no common fragments. The unweighted pair-group method with arithmetic average (UPGMA) was chosen as a clustering method. The dendrogram was designed using DENDPLOT from the same computer package. Genetic diversity (GD) was calculated as GD=1-PD according to Nei and Li (1979).

2.13 Southern blot and hybridization analysis

Southern capillary DNA transfer to nylon membranes under alkaline conditions and hybridization of DNA with radiolabeled probes were performed according to a standard procedure described by Sambrook et al. (1989). DNA fragments longer than 10 kb were transfered under alkaline conditions after fragmentation with short-wavelength (260-280 nm) UV light for 5 min. After DNA transfer, membranes were rinsed in 2xSSC and stored at 4°C until use.

The hybridization buffer contained 0.5M sodium phosphate pH 7.2, 7% SDS, 1% BSA (Sigma).

DNA probes were labeled by random priming (Feinberg and Vogelstein, 1983). Filters were hybridized at 65°C with the α^{32}P-dCTP labeled denatured probes under standard hybridization conditions and washed at 65°C in 2xSSC/0.1%SDS, 1xSSC/0.1%SDS and 0.5xSSC/0.1%SDS consecutively. Positive signals were detected with X-ray films (Kodak) or with the Phosphor Imaging Analyser BAS 2000 (Fuji Photo Film Co. Ltd).

2.14 Tomato YAC library screening

A tomato YAC library (Martin et al., 1992) was screened to select YAC clones homologous to microsatellite markers. The YAC library is organized in 384 DNA pools. Each of the 384 pools contains the DNA from 96 recombinant YAC clones. Thus, the library comprises of 36,864 recombinant clones with approximately 250 kb inserts of tomato DNA representing five haploid genome equivalents.

The screening was performed in two steps. First, the 384 YAC DNA pools were probed with microsatellite markers by PCR. PCR reactions were performed in a volume of 25 µl with the initial template DNA concentration of 100 ng/µl. PCR conditions for YAC DNA amplification were described previously (see PCR analysis). As control, the genomic DNA from tomato variety Moneymaker was used for amplification with each primer pair. The PCR products were loaded onto 2% agarose gels. The DNA pools that amplified a fragment in the expected size range were considered positive.

One pool represents one 96-well microtiter plate of stock cell culture. In order to find the single positive clone a second step of screening was performed. Each 96-well (12x8 well) microtiter plate with YACs was replicated onto two plates with SD yeast minimal medium.

SD yeast minimal medium:	6.7 g/l	Bacto yeast nitrogen base (without aminoacids)
	20.0 g/l	Glucose (dextrose)
	0.8 g/l	CSM (minus uracil, minus tryptophane)
	20.0 g/l	agar
	pH 6.8	

Plates were incubated at 30°C for 48-60 hours. The colonies on the plate were grown in twelve columns and eight rows. Yeast colonies were harvested from each of the twelve columns and from each of the eight rows, and yeast DNA for PCR analysis was extracted using glass beads. Harvested yeast colonies were resuspended in 1ml of TE buffer. Glass beads (0.45-0.5 mm, Braun Biotech) and 400µl of DNA extraction buffer (200mM Tris-HCl pH 7.5, 250mM NaCl, 25mM EDTA, 0.5% SDS) were added to the cell suspension. After 5 min. of vigorous vortexing, samples were centrifuged at 13,000 rpm. The supernatant was transferred into a fresh tube and 320µl of cold isopropanol was added to precipitate the DNA. After 5 min. of centrifugation at 13,000 rpm, the DNA pellet was redissolved in 400µl of water. 1µl of this DNA solution was sufficient for PCR. Thus, each 96-well microtiter plate of YAC clones resulted in 12 row DNA pools and 8 column DNA pools. The 20 DNA pools were then tested individually by PCR. PCR products were separated on 2% agarose gels, and the position of the positive clone was found on intersection between the positive row pool and the positive column pool.

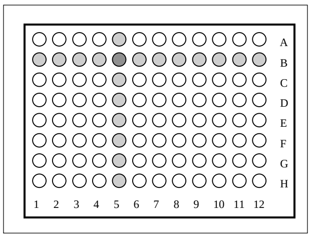


Figure 3. 96-well microtiter plate of the YAC clones and the position of the positive clone (B5) on intersection between the positive row pool (B) and the positive column pool (5).

For chromosome preparation, positive YAC clones were plated onto SD yeast minimal medium. Large scale yeast chromosome and YAC isolation was performed according to Carle and Olson (1985). A single red colony of yeast cells was transfered into 200ml of SD yeast minimal medium and incubated at 30°C for approximately 36-48 hours. Cells were harvested at 4°C by 5min. centrifugation at 5,000 rpm. The cell pellet was washed in 20 ml of 50mM EDTA solution pH 7.5. After another centrifugation, the pellet was resuspended in SCE to a concentration of 0.3 g/ml. 3 ml of yeast cell solution were mixed with 1 ml of solution 1 and with 5 ml of 1% low melting agarose in EDTA (125 mM, pH 7.5) at 40°C. The mixture was poured in a small petri dish (diameter 5cm) and solidified at room temperature. The agarose was overlayed with 5 ml of solution 2 and incubated overnight for the generation of spheroplasts at 37°C with gentle shaking. The liquid was removed and replaced with 5 ml of ESP solution. After incubation for 6-8 hours at 50°C, ESP was replaced then with 4-5 ml of EDTA (0.5M, pH 9) and the agarose-embedded HMW yeast DNA was used in PFGE or stored at 4°C.

SCE:

1.0M	sorbitol
0.1M	sodium citrate
0.06M	EDTA
pH 7.0	

Solution 1:

10 ml	SCE
0.5 ml	mercaptoethanol
10,000u	lyticase (Sigma)

Solution 2:

4.6 ml	EDTA 0.5M, pH 9
0.05 ml	Tris-HCl 1M, pH 8
0.4 ml	mercaptoethanol

ESP:

0.5 vol.	2% (w/v) sarcosyl
0.5 vol.	0.5M EDTA, pH 9
1 mg/ml	proteinase K (Boehringer Mannheim)

2.15 Pulsed field gel electrophoresis

This technique can resolve large DNA molecules in the size range of 50 to 5,000 kb in agarose gels by using alternating electric fields of different orientation. PFGE was carried out in a CHEF DR II-apparatus (contour-clamped homogeneous electric field, Chu et al. 1986, Bio Rad) at 14°C in 0.5xTBE buffer.

For the analysis of isolated YAC clones, PFGE was used to separate YACs and natural yeast chromosomes. The blocks of agarose containing yeast chromosome preparations were placed into the wells of a regular 1% agarose gel. Chromosome preparations from yeast strain AB1380 were included as DNA size standard. Analytical pulsed field gels were run at 148V with a pulse time 60s for 45 hours. Such conditions permit the optimal separation of fragments in the size range of 50 to 800 kb. If more effective band resolution was needed, the switch time and voltage were changed respectively. The switch time is the most important parameter in determining which size range of DNA molecules is resolved in a PFG (Fig.4). Preparative pulsed field gels for the isolation of large amounts of YAC DNA for subcloning were run at 160V, with a switch time gradient 20-60 s for 60 hours.

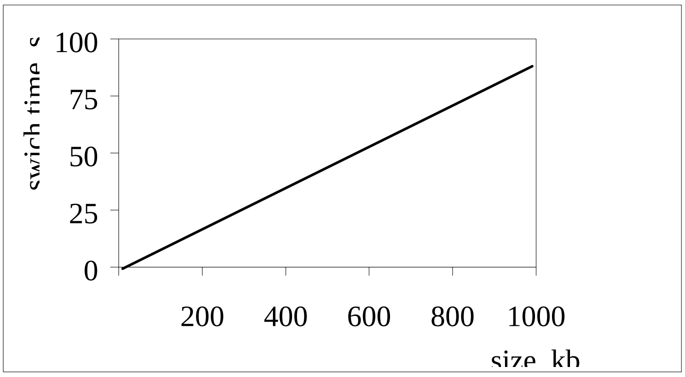


Figure 4. Determining the optimum constant swich time for separating different sized DNA molecules.

Before blotting of pulsed field gels, exposure of the ethidium bromide-stained gel to short-wavelength UV light for 5 min. for DNA fragmentation was preferable to depurination with acid. The procedures of Southern blotting and hybridization were performed as described for DNA in conventional agarose gels.

2.16 Random subcloning of centromere-associated sequences

YACs for subcloning were isolated by PFGE. Preparative pulsed field gels for the isolation of YAC DNA were run under optimized conditions which allowed the appropriate separation of all YAC clones and accurate excision of the correct bands without contamination with endogenous yeast DNA. The YACs were extracted from the gel using the Geneclean kit (Dianova). 1-2 μg of isolated YAC DNA was digested with *Sau*3AI (GibcoBRL):

DNA	- 1-2μg
REact4 (GibcoBRL)	- 2.5μl
40mM Spermidine	- 1.5μl
10mM DTT	- 1μl
*Sau*3AI (10u/μl)	- 0.75μl

H_2O to a final volume 25μl

at 37°C for 4h

*Sau*3A YAC DNA fragments were purified by precipitation with ethanol and ligated into pUC18 *Bam*HI/BAP vector (Amersham Pharmacia).

Ligation reaction:	*Sau*3AI DNA fragments (1μg)	- 3.5μl
	pUC18 vector (0.5μg/μl)	- 0.5μl
	10x ligation buffer	- 0.5μl
	T4 DNA ligase (5u/μl, Boehringer Mannheim)	- 0.5μl
	at 14°C overnight	

50μl aliquots of the high-efficiency *E.coli* XL10-Gold ultracompetent cells (Stratagene) were treated with 1.5μl of β-ME (Stratagene) and transformed with 0.75μl of the ligation mixture as described in section „Construction and screening of plasmid libraries“. Transformations were plated onto 22x22cm plates with LB-Amp-agar medium. 680μl of 3% X-gal and 170μl of 100mM IPTG were spread on the agar surface before plating for blue-white color selection to identify transformed cells carrying plasmids with inserts. Plates were incubated overnight at 37°C. Using the robotic system BioPick (BioRobotics Ltd.), single white colonies containing recombinant plasmids were transfered into 384-well microtiter plates. Medium content, cell growth parameters, preparing of high-density colony membranes and other procedures were as described in the plasmid library construction and screening section.

High-density filters were hybridized with total tomato DNA. Approximately 100ng of DNA was used per labelling reaction. Filters were hybridized at 65°C and washed to a stringency of 0.5xSSC, 0.1%SDS. Clones containing repetitive tomato DNA sequences gave a strong signal and were identified by the alignment of the autoradiograms with an appropriate grid indicating the position of all colonies. The size of the genomic inserts of the selected clones was estimated by PCR and agarose-gel electrophoresis. Single positive clones were used for plasmid DNA isolation and sequencing.

3. Results

3.1 Isolation of microsatellite containing sequences from the tomato genome and marker generation

3.1.1 Construction and screening of Lambda genomic libraries

For the isolation of microsatellite containing sequences from the tomato genome small-insert genomic libraries were constructed by cloning *Mbo*I fragments of tomato DNA into the bacteriophage λ vector. Partial digestion of genomic DNA with *Mbo*I that recognized the frequently occurring tetranucleotide sequence GATC yielded fragments of appropriate sizes (approximately 1kb) for direct insertion into the Lambda Zap Express vector with high efficiency. The highest titer of pfu ($1.5\text{x}10^3$ pfu/μl) was obtained by the ligation of 50-100 ng of digested genomic DNA to 1μg of Zap Express arms.

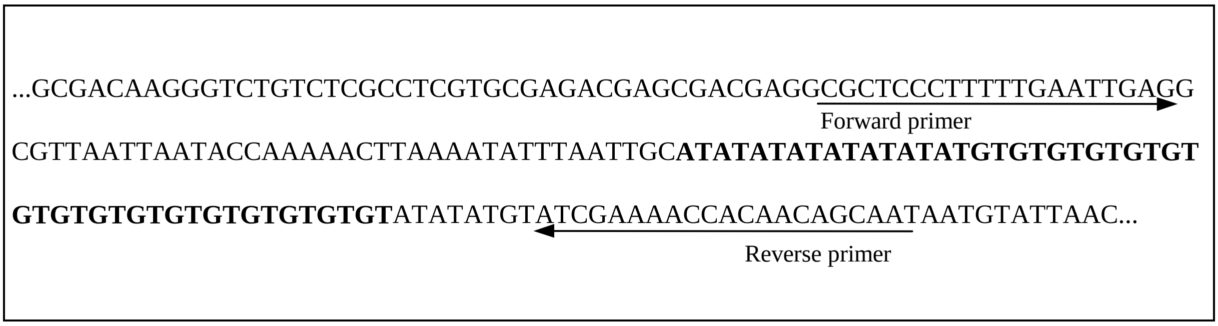
The library of tomato was screened for the following tetra- and dinucleotide simple sequence repeats. As tetranucleotide repeat, the GATA motif was chosen for library screening, because some of GATA-containing arrays were found near tomato centromeres previously (Arens et al., 1995; Broun and Tanksley, 1996). As dinucleotide probes, poly(GT) and poly(GA) were used as the most frequent in plant genomes next to AT simple sequence repeats. Because of instability of AT base pairing resulting in low stringency hybridization conditions and high background, the poly(AT) probe was not used in this study.

Approximately one-hundred thousand recombinant phage clones with an average insert size of one kilobase were screened with GATA, GT and GA oligonucleotide probes via plaque hybridization analysis. After carrying out three rounds of hybridizations, 400 positive clones (0.4% of total number) were purified. 50 clones hybridizing to GATA and 350 clones hybridizing to GA or GT probes were isolated. Comparing the number of hybridizing plaques to the total number of plated recombinant phages of known average insert size, the frequency of microsatellites in the tomato genome could be calculated. GT and GA repeats occur at comparable frequency in the tomato genome, while GATA repeats occur ten times less frequently than GT or GA repeats (Table 2).

Table 2. Screening of Lambda genomic libraries of tomato for positive clones with GATA, GT and GA oligonucleotide probes.

Total number of phage clones	Oligonucleotide probe	Number of positive clones	% of total number	Frequency in the tomato genome per Mb
100,000	GATA/CTAT	50	0.05	0.6
	GT/CA	200	0.2	5.6
	GA/CT	150	0.15	3.4

Four-hundred positive recombinant clones were converted into plasmids by *in vivo* excision and sequenced. 10 sequences which contained GATA microsatellites and 37 sequences contained GA or GT repeats were used for primer design (Fig. 5).



...GCGACAAGGGTCTGTCTCGCCTCGTGCGAGACGAGCGACGAGGCGCTCCCTTTTTGAATTGAGG
CGTTAATTAATACCAAAAACTTAAAATATTTAATTGC**ATATATATATATATATGTGTGTGTGTGT**
GTGTGTGTGTGTGTGTGTGTATATATGTATCGAAAACCACAACAGCAATAATGTATTAAC...

Figure 5. 5´-3´ nucleotide sequence containing a $(AT)_8(GT)_{16}$ microsatellite. The TMS1 primer pair was designed in the flanking regions using the Primer 0.5 program. Both primers are 20 bp long and have a G+C content 45% (T_m 59.5°C). The forward primer is labeled with fluorescein for detection of the PCR product on an ALF sequencer. The PCR-amplified fragment of VFNTcherry DNA is 134 bp.

For the remainder of the sequenced clones (87.5%) primer pairs could not be designed because of the following main reasons:

- There was no microsatellite in the sequenced region. Such inserts were often more than 1.2 kb in size, they were sequenced from both ends (500-700 bp from each side) but, nevertheless, the sequences remained incomplete. In many cases, clones contained short but multiple inserts.
- Some of the recombinant clones contained very long complex microsatellites (up to more than 800 bp long) and inserts with a size range of more than 1.5 kb. For such clones new

primers were designed close to the microsatellite sequence, but it was still not possible to sequence through the microsatellite and design the second primer.

- The sequences contained simple sequence units which were repeated less than ten times. Such short microsatellites were not taken into account according to the definition of the minimum length for a microsatellite (Morgante and Olivieri, 1993). In humans, dinucleotide sequences with ten or fewer repeats are known to yield little or now information (Weber, 1990).
- Microsatellite sequences were too close to the cloning site.

During the sequence analysis it was noted that the majority of the isolated microsatellites were complex (containing more than one microsatellite motif). Nine of ten isolated microsatellite clones harboring GATA motifs contained also dinucleotide motifs. Similarily, more than 75% of GT and 50% of GA microsatellites harbored arrays of AT and some other simple sequence repeat types. In addition, for three AT repeats that were found accidentaly during partial sequencing of the clones with extremely long inserts, primer pairs were designed.

The isolated microsatellite arrays contained a high number of repeats. The smallest microsatellite contained 12 repeats, while the longest completely sequenced dinucleotide microsatellite contained a total of more than 150 repeating units (Table 3).

Table 3. Characteristics of sequenced tomato microsatellites

Repeated motif	Total number of isolated microsatellites	Number of repeats		Type	
		min.	max.	simple	complex
GT	16	12	>65	2	14
GA	21	18	>150	7	14
GATA	10	12	>115	1	9

A total of 50 primer pairs were designed and after PCR amplification on VFNTcherry DNA 25 (50%) markers yielded a product of the expected size. The other 25 markers amplified

- several additional fragments with nearly equal intensity, and the fragment of the expected size was difficult to score,
- a single fragment of incorrect size,
- the primer pair resulted in a very weak amplification or no amplification at all.

Table 4. 25 microsatellite markers generated from the small insert phage library of tomato

Microsatellite marker	Repeated sequence	Fragment size, bp	Primer pair, 5´-3´	T_m, °C
TMS1	$(AT)_8(GT)_{16}$	134	CGCTCCCTTTTTGAATTGAG TTGCTGTTGTGGTTTTCGAT	60
TMS2	$(GT)_{41}(TA)_6(CT)_9$	387	TCTTTCATTTCATGTCACGA AGGAGACCTTATGATTCAAGG	55
TMS4	$(CT)_{12}(GATA)_{12}$ $ATAT(AC)_{10}$	230	CGATTAGAGAATGTCCCACAG TTACACATACAAATATACATAGTCTG	47
TMS6	$(GATA)_{45}$	335	CTCTCTCAATGTTTGTCTTTC GCAAGGTAGGTAGCTAGGGA	55
TMS7	$(TA)_{31}(GATA)_{13}$	170	ACAAACTCAAGATAAGTAAGAGC GTGAATTGTGTTTTAACATGG	50
TMS8	$(GATA)_{87}$	470	GCGCACCCAAAGTTGAAG CCTCATAGGGACGCACATAC	55
TMS9	$(GATA)_{24}$	360	TTGGTAATTTATGTTCGGGA TTGAGCCAATTGATTAATAAGTT	55
TMS17	$(GATA)_{24}(AT)_8\ (GT)_{25}$	250	AATGTAACAACGTGTCATGATTC AAGTCACAAACTAAGTTAGGG	50
TMS21	$C_{12}(CT)_{38}$	235	GTGTTTTTATGCAGGGTTTG CACACTTATACCTCACCCGT	55
TMS22	$(GT)_9(AT)_8(AC)_{13}(GA)_{12}$	164	TGTTGGTTGGAGAAACTCCC AGGCATTTAAACCAATAGGTAGC	55
TMS23	$(GT)_{32}(AT)_{67}$	412	GGATTGTAGAGGTGTTGTTGG TTTGTAATTGACTTTGTCGATG	55
TMS24	$(AT)_{10}(GT)_{19}$	375	AGGTTCAATGGACTTCTCGC AATTTGGATTTGTAAAAATTTGG	55
TMS26	$(GA)_{20}$	234	TTCGGTTTATTCTGCCAACC GCCTGTAGGATTTTCGCCTA	55
TMS27	$(GA)_{82}$	325	AATTTCGGACCCGCCGAG TTCAACGCCATCGATGC	60
TMS29	$(G)_{15}(GA)_{22}$	354	AGCCACCCATCACAAAGATT GTCGCACTATCGGTCACGTA	55
TMS33	$(GA)_{26}$	264	AGCATGGGAAGAAGACACGT TTGAGCAAAACATCGCAATC	60
TMS34	$(GA)_{19}$	208	TTCCTCACTATTTTGAATTGCG TGTACTTCTCTGCAGATTCCA	55
TMS35	$(GA)_{31}$	150	TTGTCGCTTCAGTTTTGGC TTCACCTTGCCACTGTGAAG	50
TMS37	$(GA)_{21}(TA)_{20}$	160	CCTTGCAGTTGAGGTGAATT TCAAGCACCTACAATCAATCA	55
TMS39	$(AT)_{29}$	120	CGGCGTATTCAAACTCTTGG GCGGACCTTTGTTTTGGTAA	60
TMS42	$(AT)_{17}(GT)_{18}$	279	AGAATTTTTTCATGAAATTGTCC TATTGCGTTCCACTCCCTCT	55
TMS43	$(GA)_{31}(GATA)_7$	323	TTGGCCTAATCCTTTGTCAT AACAATGTGACGTCTTATAAGGG	55
TMS44	$(GA)_{19}$	405	TTCAAGGTTTATTCGAAAATCC TTTGGGCCTATCACCTTGTC	55
TMS45	$(GA)_{17}(GT)_6$	249	CCGTCCAGAAGACGATGTAA CAAAGTCTTGCCAACAATCC	55
TMS48	$(GA)_{24}(TA)_{31}$	182	ATTGCTCATACATAACCCCC GGGACAAAATGGTAATCCAT	55

Amplification on genomic DNA was initially carried out under standardized PCR conditions as described in Materials and Methods. If necessary, the annealing temperature was increased by 5°C in the case of a high number of unspecific fragments or decreased by 5°C in the case of weak amplification. Special optimizations of PCR assay for each primer pair were not performed because, ideally, a single set of reaction conditions should be used for all assays to simplify the testing process of a high number of markers and to facilitate multiplexing easily reproducible in other laboratories.

3.1.2 Construction and screening of plasmid libraries enriched for single-copy sequences

Only 12.5% of microsatellite containing sequences isolated from the small-insert phage libraries of tomato were suitable for primer design. Only half of the synthesized primer pairs yielded functional markers.

In order to increase the efficiency of functional microsatellite markers, another type of tomato genomic libraries was constructed by small-insert cloning of genomic low-copy DNA into a plasmid vector. Total VFNTcherry genomic DNA was predigested with the methylation-sensitive restriction enzyme *Pst*I. *Pst*I does not cleave at its recognition sequence of 5´-CTGCAG-3´ if the 5´cytosine is methylated. 5´CNG3´ are the most common sites for methylation in plant genomes. Most of repetitive sequences are substantially methylated at this site, therefore, this enzyme cleaves preferentially in single- or low-copy DNA. Thus, cloning with restriction enzymes like *Pst*I yields libraries enriched for single-copy sequences.

Selected 2-9 kb fractions of *Pst*I restriction DNA fragments were futher digested with one of the following cutters *Sau3A*I, *BamH*I, or *Acs*I and inserted into the pUC18 plasmid. This approach permits to reduce the number of microsatellite clones derived from regions of highly repeated DNA. The average insert size in the pUC18 vector was approximately 250 bp.

About fourty thousand clones were analysed by colony hybridization. The number of clones hybridizing to poly(GT) and poly(GA) oligonucleotides in respect to the total number of clones that were analysed was less than 0.1%. This ratio was four times lower than the percentage obtained during Lambda library screening.

It was noticed during small-insert (approximately 1 kb) phage library screening, that microsatellites isolated from this library were containing extremely high numbers of repeating units (more than 30 on average) and were often complex (combination of different di- and tetranucleotide repeats). This led to difficulties with sequencing and primer design. Screening of plasmid libraries enriched for single- and low-copy DNA with an average insert size of 250 bp, led to four times lower numbers of positive clones. It is possible, that the low-copy DNA fraction contains smaller number of simple sequence repeats or 250 bp fragments failed to contain microsatellites. Increasing the size of the tomato DNA fragments before ligation into vector to approximately 750 bp by a second size selection increased the number of positive clones (Table 5 and 6).

Table 5. Screening data of plasmid genomic libraries with GT and GA oligonucleotide probes.

Type of the library	*Pst*I/*Sau3A*I	*Pst*I/*Acs*I	*Pst*I/*BamH*I	*Pst*I/*Sau3A*I
Average insert size, bp	250	200	250	750
Total number of clones	15,744	7,296	7,296	6,528
Number of positive clones	9	2	-	16
% of total number	0.06	0.03	-	0.25

Table 6. Increase of the number of SSR containing sequences is dependent on the insert size.

An average insert size of tomato DNA fragments in the library, bp	250	750	1,000 (phage library)
Number of positive clones of the total number of recombinant clones, %	<0.1	0.25	0.4

A total of 27 positive recombinant clones were sequenced from the plasmid library of tomato, of which 22 contained dinucleotide repeats. Five clones failed in detecting microsatellites mostly because of sequencing shortcomings. All isolated GT repeats and the majority of GA microsatellites were associated with AT repeats. The average number of GT motifs was 18, and GA motifs were on average 22 times repeated. Thus, microsatellites isolated from the plasmid libraries of tomato showed the same repeat length and complex structure as SSRs isolated from the Lambda libraries.

No microsatellites were isolated by hybridizations with tri- and GATA probes. Five (23%) of the isolated microsatellites were located near the restriction sites. For 16 microsatellites primer sets could be designed. 10 (62.5%) out of 16 synthesized primer pairs yielded functional markers (Table 7).

Table 7. Repeated motifs and primer sequences of 10 microsatellite markers isolated from tomato plasmid libraries.

Micro-satellite marker	Repeated sequence	Fragment size, bp	Primer pair, 5´-3´	T_m, °C
TMS52	$(AC)_{14}(AT)_{18}$	152	TTCTATCTCATTTGGCTTCTTC TTACCTTGAGAATGGCCTTG	55
TMS54	$(TA)_{14}(CA)_{15}$	226	TTGGTCTAGAACGATGAGCA GCCATGCATCACTGAATGAC	60
TMS56	$(CT)_{19}$	120	GATCTCAAAGGATGAACAATAC TCATTAGGAGATTCTTTGTATCA	55
TMS57	$(GT)_{13}(AT)_5$	260	ACATGACCGGTTGACGACTA AAATTGTCCACATGGTGGGT	60
TMS58	$(TA)_{15}\,(TG)_{17}$	225	CATTTGTTGTATGGCATCGC CAGTGACCTCTCGCACAAAA	60
TMS59	$(AT)_5(GT)_{11}(AT)_{11}$	100	TGAACGGGCCTTCTGTTATC ATCATCATTATAGTTCTTAAGTGAT	55
TMS60	$(AT)_{13}(CA)_{23}(AT)_{22}$	242	ATGCAGTTCCAAGCATCATT TTGCCACATTAATGTTGAAGT	60
TMS63	$(AT)_4(GT)_{18}(AT)_9$	150	GCAGGTACGCACGCATATAT GCTCCGTCAGGAATTCTCTC	60
TMS65	$(TA)_{25}(GA)_{20}$	308	AGCTTCATCCATTACGCCAC GTGCATCTGGCGTACCTACC	60
TMS66	$(GT)_{19}(AT)_4$	201	GGGTTAATAAAGCAATGTAGCG CTCTTCATTAAAGTTGCCGC	60

Under the standard PCR conditions, the remainder of the synthesized primer pairs (37.5%) amplified several fragments on tomato DNA, or a single fragment of incorrect size.

3.2 Generation of microsatellite markers from EST sequences

TIGR database is a collection of curated databases containing DNA and protein sequence, gene expression, cellular role, protein family, and taxonomic data for microorganisms, plants and humans. It consists of TIGR Gene Indices including the Tomato Gene Index (LGI) - integrating data from international EST sequencing and gene research projects and an analysis of all transcribed sequences represented in the world’s public EST data.

LGI release version 1.2 (August 5, 1999) contained a total of 26,362 tomato EST sequences. In order to isolate microsatellites from expressed parts of the tomato genome, LGI was searched with all possible simple sequence repeat sequences (repeated at least 10 times) using BLAST similarity searching. A total of 241 SSR containing sequences were identified, among them only 49 (0.2% of total number) were containing microsatellites with more than ten repeated motifs. AT dinucleotide repeats were by far the most frequent class of microsatellites with more than ten repeated motifs followed by GA repeats (Table 8). Six GT microsatellites were found which were repeated less than ten times and all were adjacent to AT repeats. AAT repeats were the most abundant among trinucleotide repeats followed by complementary ATT motifs which were repeated on average 7 times. About twenty GAT and GCT, and about ten GCA and GTA microsatellites were found in the database. All of them were repeated on average 7 times and were often adjacent to other types of repeats. No tetranucleotide microsatellites were found except one TTAA and one TTTA repeat which were repeated five times each.

Table 8. Relative abundancies of microsatellites in tomato EST database.

Repeated motif	Number of sequences	Mean n
$(AT)_n$	28	14
$(GA)_n$	22	23
$(GT)_n$	6	7
$(AAT)_n$	88	7

For many microsatellites within EST sequences primers could not be designed because they were located at the ends of ESTs in non-translated regions. Only twenty completely sequenced uninterrupted microsatellites were used for primer design. The majority of those sequences contained AT repeats. Eleven (55%) primer sets amplified a single fragment of the expected size (Table 9). The remainder resulted in the amplification of multiple fragments.

Table 9. List of 11 microsatellite markers generated from EST sequences.

Micro-satellite marker	Repeated sequence	Fragment size, bp	Primer pair, 5´-3´	T$_{m}$, °C
TC1843	(CAG)$_{8}$(AAT)$_{12}$	567	ATGGAGTTTCAGGACCACTT AGGATGATTCAATATATCCGC	60
TC1107	(AT)$_{11}$	97	TCCATCTCTCTCTAGACCTTTCT TTCTTAAATCCTCTCACTCA	58
TC948	(TA)$_{38}$	150	TTTTCGCGTTAAGAGATGTT CCGCCATACTGATACGATAG	60
TC11	(AG)$_{11}$	91	TCAACACAGAGAAAATAGGCA CAGCTTGCTCAGCCAGC	60
TC461	(TAT)$_{15}$	185	GGCTGCCTCGGACAATG TTATTGCCACGTAGTCATGA	60
EST245053	T$_{12}$(GT)$_{6}$(GA)$_{6}$	228	CCATTTAAATGACCCTATGCT AATCAAAAAGAATCTAAGCCCT	58
EST253712	(AT)$_{14}$	140	GAAATGAAGCTCTGACATCAAA TCATTGCTTGCATATGTTCATG	55
EST258529	(TA)$_{17}$	127	AACACCCTTTATTCAGATTCC GCATAAAAATGTTAAAGGGG	50
EST248494	(TA)$_{13}$T$_{20}$	211	CTGAAACGAGACAGAGGAAG AGCTGAGTACGTCTCCCATG	60
EST259379	(TA)$_{20}$	151	TTGGTCTCCCTCTTTATGCC GGCTTCATTGATGAACCCAT	55
EST268259	T$_{11}$(GT)$_{4}$(AT)$_{4}$(GT)$_{8}$	127	GCTGCTCCTATTGGTTACCA TCTCCTTATTTGGATTGGCC	55

3.3 Polymorphism of microsatellites in *L.esculentum* varieties

The microsatellite markers which amplified fragments in the expected size range were used to study the allelic diversity in the cultivated tomato. 12 lines were chosen for this experiment representing samples from processing, fresh market and cherry tomatoes and different geographic origins (US, Brazil, France, Russia). This number of genotypes represented a good sample of the tomato germplasm due to their clear differentiation with regard to geographic origin, morphological characteristics and agronomic traits. *L.pennellii* LA716 was also included as parent of the standard mapping population.

For the 46 analyzed microsatellites, a total of more than 125 alleles could be identified in this material ranging from one allele to a maximum of five different alleles. On average, approximately three alleles were detected in the investigated *L. esculentum* accessions for the 35 markers generated from genomic libraries. Two and one markers which were generated from Lambda and plasmid libraries respectively (8.6 % of total number) were monomorphic. All microsatellite markers that were generated from the EST sequences were polymorphic between *L.esculentum* and *L.pennellii* LA716. One of eleven markers detected 4 alleles, the others detected 2-3 alleles with an average of 2.55 (Table 10 and 11).

The level of polymorphism of microsatellite markers generated from EST sequences in the twelve *L.esculentum* lines was comparable with that observed for longer microsatellites isolated from genomic libraries (Table 10).

Table 10. Comparative analysis of polymorphisms detected with different microsatellite markers in twelve *L.esculentum* lines and *L.pennellii* LA716.

Microsatellite markers generated from	Phage libraries	Plasmid libraries	EST sequences
Total number of markers	25	10	11
Average allele size in VFNTcherry, bp	266.2	198.4	180.9
Average number of alleles in twelve *L.esculentum* lines and *L.pennellii* LA716	3.0	2.5	2.55

When surveying the markers for polymorphism between the two parents of *L.esculentum* x *L.pennellii* LA716 interspecific cross it was noted, that only half of the microsatellite markers yielded defined fragments in *L. pennellii* after PCR amplification. The other half did not amplify a product under the standard conditions. This was also observed for the markers generated from EST sequences. No amplification at all was scored as a null allele (Fig. 6).

Null alleles are especially pronounced for TMS43 and TMS44, where 9 of the 12 *L. esculentum* lines contained a null allele. From the mapping it is obvious that these markers are located in the region where the tobacco mosaic virus resistance gene is located. The *Tm-2a* gene conferring resistance of tomato to all known strains of tobacco mosaic virus (TMV) has been introgressed in two different alleles (*Tm-2* and *Tm-2a*) from the distantly related wild species *L. peruvianum* into the cultivated tomato. The polymorphism for the presence or absence of these sequences among *L. esculentum* cultivars reflects polymorphism for TMV resistance contained on the segment of introgressed DNA. Those two microsatellite markers yield amplification products only of VFNTcherry, TA55 and Momor tomato varieties carrying the *Tm-2a* gene.

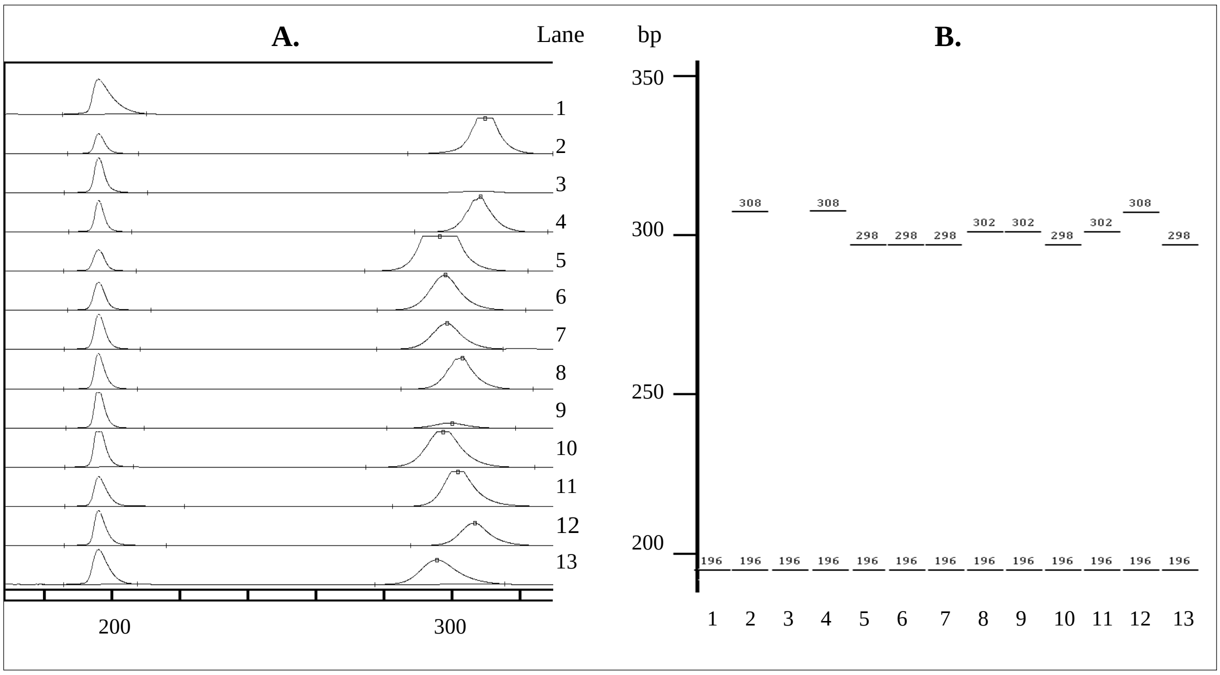


Figure 6. Detection of PCR-amplified fragments on an ALF DNA sequencer is shown as a printout in the **A. -** curve modus and **B. -** fluoregram modus of the computer program Fragment manager 1.2. DNA from twelve tomato lines was amplified with TMS65. Four alleles could be detected including the null allele in *L.pennellii* LA716. Two internal size standards of 70 and 196 bp were included (70 bp standard is not shown). The lanes are as follows: 1) standard as control, 2) VFNTcherry, 3) *L.pennellii* LA716, 4) TA55, 5) Moneymaker, 6) Momor, 7) RioGrande, 8) NewYorker, 9) Piline, 10) Fline, 11) Angela, 12) Puz11, 13) Monita.

Table 11. Number and distribution of alleles detected with each microsatellite marker in twelve *L.esculentum* lines and *L.pennellii* LA716.

Markers generated from phage libraries

Marker	Number of alleles	Comments
TMS1	3	134bp allele in VFNTcherry and Puz11; 138bp allele in all other *L.esculentum* lines; 118bp allele in *L.pennellii*
TMS2	3	387bp allele in VFNTcherry, Moneymaker, Momor, NewYorker, Angela, Monita, and TA205; 400bp allele in all other *L.esculentum* lines; 372bp allele in *L.pennellii*
TMS4	3	230bp allele in all *L.esculentum* lines except Fline, Puz11, and TA205 with 234bp allele; 226bp allele in *L.pennellii*
TMS6	3	two fragments of equal intencity of 335 and 344 bp in VFNTcherry, TA55, Moneymaker, Fline, and Puz11; two fragments of equal intensity of 317 and 344 bp in all other *L.esculentum* lines; null allele in *L.pennellii*

TMS7	4	170bp allele in VFNTcherry and Puz11; 152bp allele in RioGrande; 148bp allele in all other *L.esculentum* lines; null allele in *L.pennellii*
TMS8	2	470bp allele in all *L.esculentum* lines; two fragments in *L.pennellii* of 445 and 457bp
TMS9	3	360bp in VFNTcherry, TA55, Angela, Monita, and Puz11; 352bp allele in all other *L.esculentum* lines; null allele in *L.pennellii*
TMS17	4	250bp allele in VFNTcherry and RioGrande; 270bp allele in Fline and TA205; 260bp allele in TA205 and in all other *L.esculentum* lines; 245bp allele in *L.pennellii*
TMS21	1	235bp allele in all *L.esculentum* lines and in *L.pennellii*
TMS22	3	164bp allele in VFNTcherry and PuzII; 160bp in all other *L.esculentum* lines; 156bp allele in *L.pennellii*
TMS23	4	412bp allele in VFNTcherry, TA55, and PuzII; 426bp allele in NewYorker and Angela; 436bp allele in all other *L.esculentum* lines; *L.pennellii* and Piline - null allele
TMS24	3	375bp allele in all *L.esculentum* lines; two fragments in *L.pennellii* of 355 and 420bp; two fragments in VFNTcherry of 355 and 375bp;
TMS26	3	234bp allele in VFNTcherry; 228bp allele in all other *L.esculentum* lines; null allele in *L.pennellii*
TMS27	1	one 325bp allele in all *L.esculentum* lines and in *L.pennellii*
TMS29	3	354bp allele in VFNTcherry, RioGrande, NewYorker, Piline, Fline, Angela, Puz11, and TA205; 349bp allele in all other *L.esculentum* lines; 336bp allele in *L.pennellii*
TMS33	2	264bp allele in VFNTcherry, Moneymaker, Momor, Angela, Monita, and in *L.pennellii*; 258bp allele in TA55, RioGrande, NewYorker, Piline, Puz11, and TA205
TMS34	3	208bp allele in VFNTcherry and TA55; 186bp allele in all other *L.esculentum* lines; 192bp allele in *L.pennellii*
TMS35	2	null allele in *L.pennellii*, very weak amplification or null allele in Moneymaker, Piline, Fline, Puz11 and Monita; 150bp allele in VFNTcherry, TA55, and in all other *L.esculentum* lines
TMS37	5	160bp allele in VFNTcherry, TA55, Moneymaker, Momor, Piline, Monita, and TA205; 173bp allele in RioGrande; 153bp allele in NewYorker, Fline, and Puz11; 148bp allele in Angela; several cosegregating fragments in *L.pennellii* around 180bp
TMS39	4	120bp allele in VFNTcherry; 115bp allele in TA55, NewYorker, Piline, and TA205; 130bp allele in all other *L.esculentum* lines; 112bp allele in *L.pennellii*
TMS42	4	279bp allele in VFNTcherry; 283bp allele in Fline, TA205 and Angela; 272bp allele in all other *L.esculentum* lines; null allele in *L.pennellii*
TMS43	2	323bp allele in VFNTcherry, TA55, and Momor; null allele in *L.pennellii* and in all other *L.esculentum* lines
TMS44	2	405bp allele in VFNTcherry, TA55, and Momor; null allele in *L.pennellii* and in all other *L.esculentum* lines
TMS45	4	amplified two fragments in VFNTcherry, TA55, Momor, and Fline of 249 and 265bp; one 265bp allele in Moneymaker, Angela, and Monita; one 259bp allele in RioGrande, NewYorker, Piline, Puz11,

		and TA205; weak amplification in *L.pennellii* around 233bp
TMS48	4	182bp allele in VFNTcherry, TA55, and PuzII; 206bp allele in Angela; 160bp in all other *L.esculentum* lines; 144bp allele in *L.pennellii*

Markers generated from plasmid libraries

TMS52	3	152bp allele in VFNTcherry and TA55; 146bp allele in all other *L.esculentum* lines; null allele in *L.pennellii*
TMS54	2	226bp allele in all *L.esculentum* lines; null allele in *L.pennellii*
TMS56	4	120bp allele in VFNTcherry, Moneymaker, Momor, NewYorker, Angela, Monita; 122bp allele in TA55; 124bp allele in all other *L.esculentum* lines; 115bp allele in *L.pennellii*
TMS57	1	260bp allele in all *L.esculentum* lines and in *L.pennellii*
TMS58	2	225bp allele in all *L.esculentum* lines; null allele in *L.pennellii*
TMS59	2	100bp allele in all *L.esculentum* lines; 113bp allele in *L.pennellii*
TMS60	2	242bp allele in all *L.esculentum* lines, null allele in *L.pennellii*
TMS63	3	150bp allele in VFNTcherry, Moneymaker, Momor, Angela, Monita; 175bp allele in TA55, RioGrande, NewYorker, Piline, Fline, Puz11, TA205; null allele in *L.pennellii*
TMS65	4	308bp allele in VFNTcherry, TA55 and Puz11; 302bp allele in NewYorker, Piline and Angela; 298bp allele in other *L.esculentum* lines, null allele in *L.pennellii*
TMS66	2	201bp allele in all *L.esculentum* lines except Piline; Piline and *L.pennellii*-null allele

Markers generated from EST sequences

TC1843	3	567bp allele in all *L.esculentum* lines except NewYorker-570bp; 562bp allele in *L.pennellii*
TC1107	2	97bp allele in all *L.esculentum* lines; 85bp allele in *L.pennellii*
TC948	2	150bp allele in VFNTcherry, TA55, Moneymaker, Momor, RioGrande, Fline, Angela, Monita, TA205; null allele in *L.pennellii*, NewYorker, Piline, Puz11
TC11	3	91bp allele in TA55, Moneymaker, Momor, NewYorker, Angela, Puz11, Monita; 99bp allele in VFNTcherry, RioGrande, Piline, Fline, TA205; and 104bp allele in *L.pennellii*
TC461	3	185bp allele in VFNTcherry and RioGrande; 180bp allele in Fline, Angela, and TA205; 182bp allele in other *L.esculentum* lines; two fragments of 182 and 171bp in *L.pennellii*
EST245053	2	128bp allele in all *L.esculentum* lines; 116bp specific allele in *L.pennellii*
EST253712	4	140bp allele in all *L.esculentum* lines except TA205-138bp and VFNTcherry-148bp; null allele in *L.pennellii*
EST258529	2	127bp allele in all *L.esculentum* lines; 105bp specific allele in *L.pennellii*
EST248494	2	211bp allele in all *L.esculentum* lines; specific 180bp allele in *L.pennellii*
EST259379	3	151bp allele in all *L.esculentum* lines except Puz11-147bp, and specific 130bp allele in *L.pennellii*
EST268259	2	127bp allele in all *L.esculentum* lines; specific 104bp allele in *L.pennellii*

3.4 Estimation of genetic distances

Genetic distances between twelve tomato cultivars and *L.pennellii* LA716 were estimated using fourteen microsatellite markers which detected the highest number of alleles (Table 12). These markers revealed a total of 51 alleles, an average 3.6 alleles per TMS.

Table 12. Microsatellite markers used to estimate genetic distances between twelve tomato cultivars and *L.pennellii* LA716.

Marker	TMS4 TMS6 TMS9 TMS26 TMS29 TMS34	TMS23 TMS39 TMS42 TMS45 TMS48 TMS56 TMS65	TMS37
Number of alleles	3	4	5

The high level of polymorphism allowed efficient discrimination of tomato varieties. Fragments amplified by these microsatellite markers of each tomato line were scored as presence (1) or absence (0). Genetic distances were estimated according to Nei and Li (1979) based on the probability that the amplified fragment from one genotype would be present in another genotype. This value range between 1 (no common fragments) and 0 (all fragments are in common).

Among the cultivars, genetic distances ranged from 0.21 to 0.74 and on average 0.42. Cluster analysis is reflected in the phenogram (Fig. 7). The phenogram discriminates all varieties and shows close genetic similarity of Moneymaker, Momor and Monita, which is expected from their related descent. VFNTcherry and TA55 form a cluster. As expected, the distantly related *L.pennellii* LA716 is separated from *L.esculentum* lines.

Based on these data, tomato microsatellites displayed a much higher level of allelic variation in cultivated tomatoes than RFLP markers (Miller and Tanksley, 1990). Thus, the markers could be used for variety and hybrid identification.

Ten microsatellites were tested for the ability to amplify fragment of the expected size in related solanaceous species (*Solanum*, *Capsicum*, *Datura*, *Petunia* and *Nicotiana*). Only two of ten tested primer sets yielded a clear fragment in potato and pepper DNA. This indicated that the majority of isolated markers did not amplify in related genomes and were tomato genome-specific.

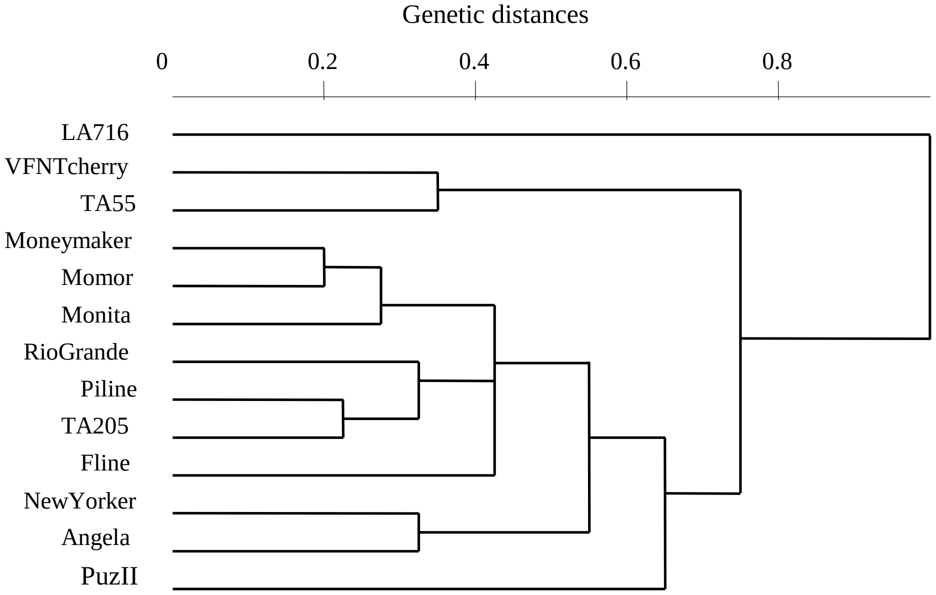


Figure 7. Dendrogram of genetic relationship between twelve tomato cultivars and *L.pennellii* LA716 estimated using fourteen microsatellite markers.

3.5 Genetic mapping of tomato microsatellites

Genetic mapping of the isolated microsatellite markers was performed using an F2 population consisting of 43 plants derived from the *L.esculentum* TA55 x *L.pennellii* LA716 interspecific cross. This segregating population was used for the construction of a saturated RFLP map (Tanksley et al., 1992). The current high-density molecular linkage map of tomato contains more than 1,000 markers which are distributed over 1270 cM with an average spacing between markers of approximately 1cM. The map position of each microsatellite marker was determined on the existing high-density linkage map using the MAPMAKER software (Lander et al., 1987).

A total of 41 polymorphic microsatellite markers which revealed 43 independent loci were mapable in this population. All markers were placed with a LOD score greater than 3 to the respective region. The mapping results are shown in Table 13 and Fig. 8. If no PCR product was observed in the *L.pennellii* parent (null allele), the microsatellite was scored as dominant marker (presence/absence of the *L.esculentum* PCR product).

TMS24 amplified two fragments of 355 and 375 bp on VFNTcherry DNA, two fragments of 355 and 420 bp on *L.pennellii* and a 375 bp fragment on TA55. The three PCR-amplified fragments on the segregating population were scored as two loci. TMS24A and TMS24B were mapped in centromeric region of chromosomes 2 and 1 respectively, where 45S ribosomal and 5S rRNA genes are located. Homology search against database sequences using BLAST showed 99% identity of a 100 bases sequence flanking one side of the microsatellite with nucleotide sequences of the ribosomal RNA genes and intergenic spacers from various plant species including *L.esculentum*.

TMS45 which amplified on TA55 DNA two fragments of nearly equal intensity of 249 and 265 bp respectively and a 233 bp PCR product in *L.pennellii*, was scored for two different loci TMS45A and TMS45B. TMS45A was mapped in the region of centromere on chromosome 1 and TMS45B mapped to the centromeric region of chromosome 9.

The 22 markers isolated from tomato phage libraries and 9 markers isolated from the genomic plasmid libraries did not map in a random fashion onto the tomato chromosomes. Independently of the type and the number of repeated motifs, the map position of all microsatellite markers was at or close to the position of the presumed centromere of the respective chromosome. The largest tomato chromosome 1 with 7 different microsatellite markers contained a higher number of microsatellites than it would be expected from a random distribution. Being one of the smallest tomato chromosomes, chromosome 12 harboured 6 microsatellite markers that were all linked to the centromeric region. Interestingly, mapping of microsatellites isolated from tomato plasmid libraries that had been enriched for single-copy sequences revealed that these markers were also located in the centromeric heterochromatin.

Compared to the markers isolated from genomic clones, significantly different map positions were observed for microsatellites generated from sequences derived from expressed parts of the tomato genome. 10 markers were mapped onto the genetic map. Two of them (EST253712 and EST259379) mapped in the centromeric regions of chromosome 6 and 4 respectively, but 8 were mapped in the regions distant from the centromeres. For example, seven markers isolated from genomic libraries were clustered in the region of centromere on chromosome 1, but two EST markers TC1107 and EST268259 were cosegregating with TG375, and EST245053 was mapped to the end of the long arm of this chromosome.

Table 13.

A. Genetic mapping data of 22 microsatellite markers generated from the small-insert phage libraries of tomato

Locus	Location on chromosome	Nearest or cosegregating marker
TMS1	2	TG33
TMS2	3	TG247, TG 414
TMS4	10	CT154, TG52
TMS6	1	CT98
TMS7	12	TG283
TMS8	3	TG525
TMS9	12	TG283
TMS17	1	CT98
TMS22	4	CT157
TMS23	12	TG283
TMS24A	2	CT106A
TMS24B	1	TG51,CT98
TMS26	4	TG652
TMS29	8	CT187A
TMS34	9	TG390
TMS35	2	TG33
TMS37	5	TG96
TMS39	5	TG503
TMS42	11	TG47
TMS43	9	TG390, CD3
TMS44	9	CD3
TMS45A	1	TG51
TMS45B	9	CD3
TMS48	12	TG283

B. Genetic map position of 9 microsatellite markers isolated from plasmid libraries of tomato.

Marker	Location on chromosome	Nearest or cosegregating marker
TMS52	12	TG394
TMS54	3	CP116A
TMS56	1	CT98
TMS58	3	TG366
TMS59	8	CT187A
TMS60	7	TG166
TMS63	1	CT98
TMS65	12	TG283
TMS66	1	TG51,CT98

C. Genetic mapping of 10 microsatellite markers generated from EST sequences

TC1843	7	TG418A
TC1107	1	TG375
TC11	4	TG62
EST245053	1	TG158
TC948	8	CT69, CT88
EST253712	6	TG178
EST258529	5	CT167
EST248494	8	CT69, CT88
EST259379	4	TG182, CT157
EST268259	1	TG375

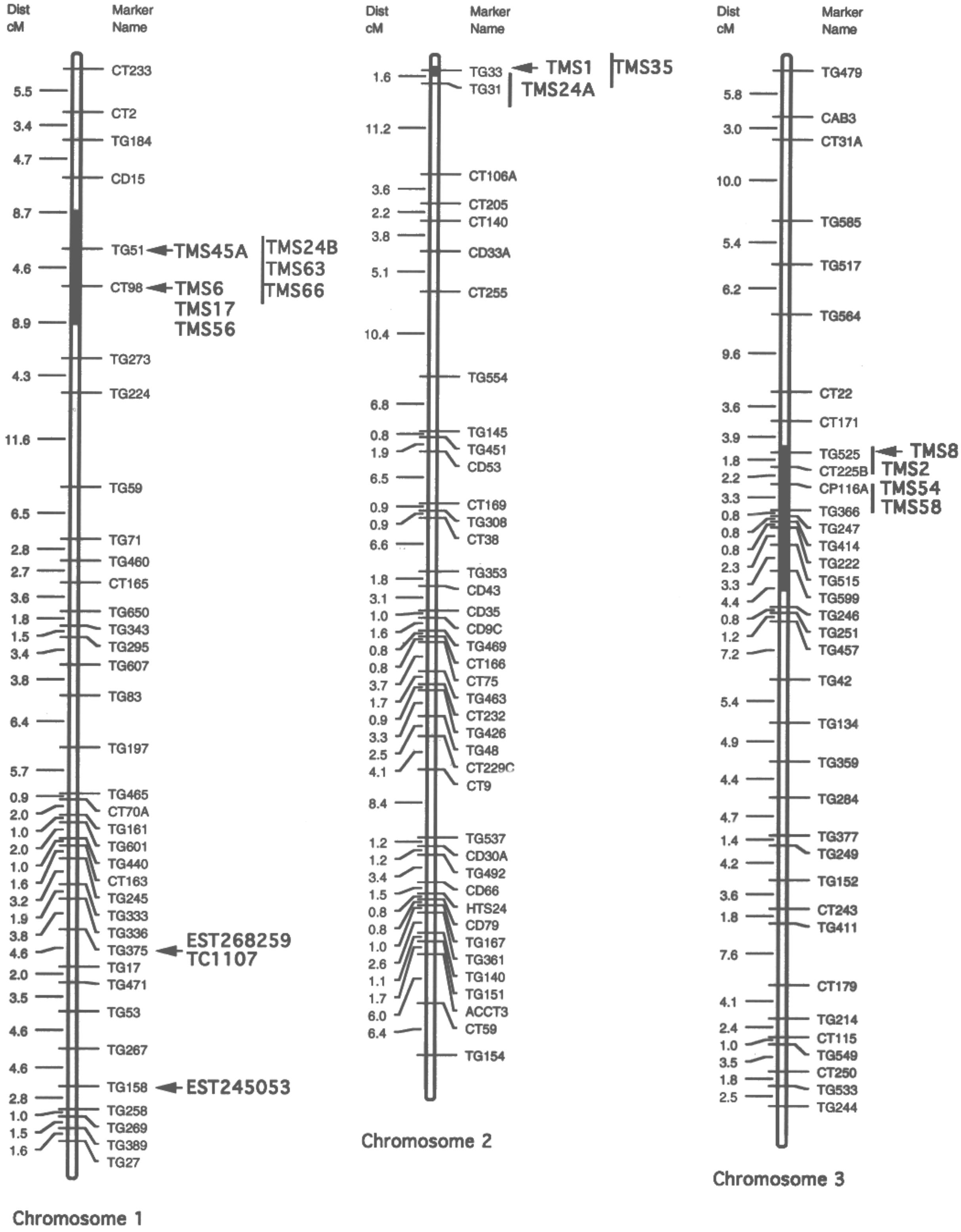


Figure 8. Positions of microsatellite loci on molecular linkage map of tomato. Arrows indicate that a TMS marker is cosegregating with a given locus. Vertical lines indicate the range into which a marker could be placed with a LOD score of 3. The position of the centromere for each chromosome is indicated by a black bar in the chromosome picture.

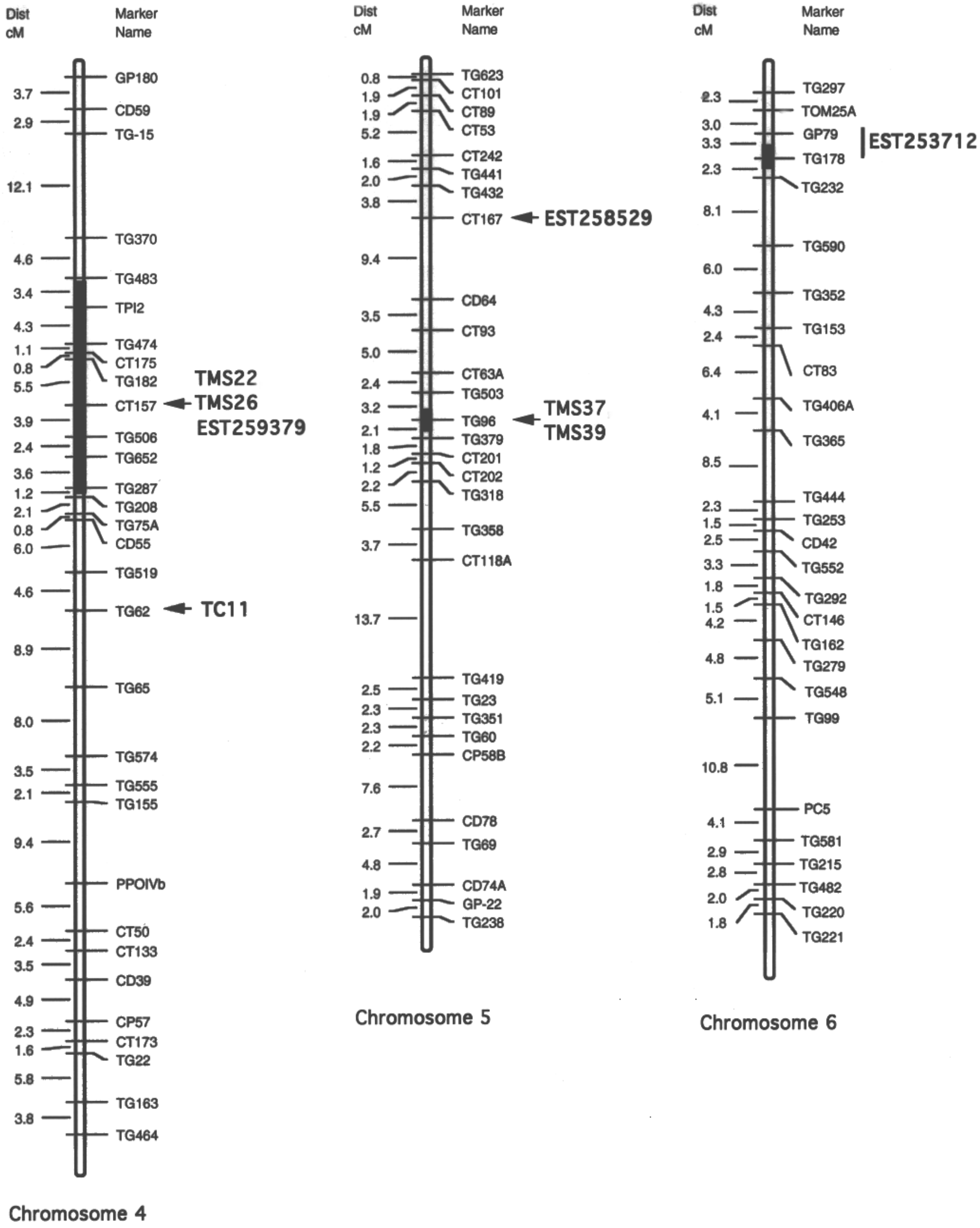


Figure 8. (Continued).

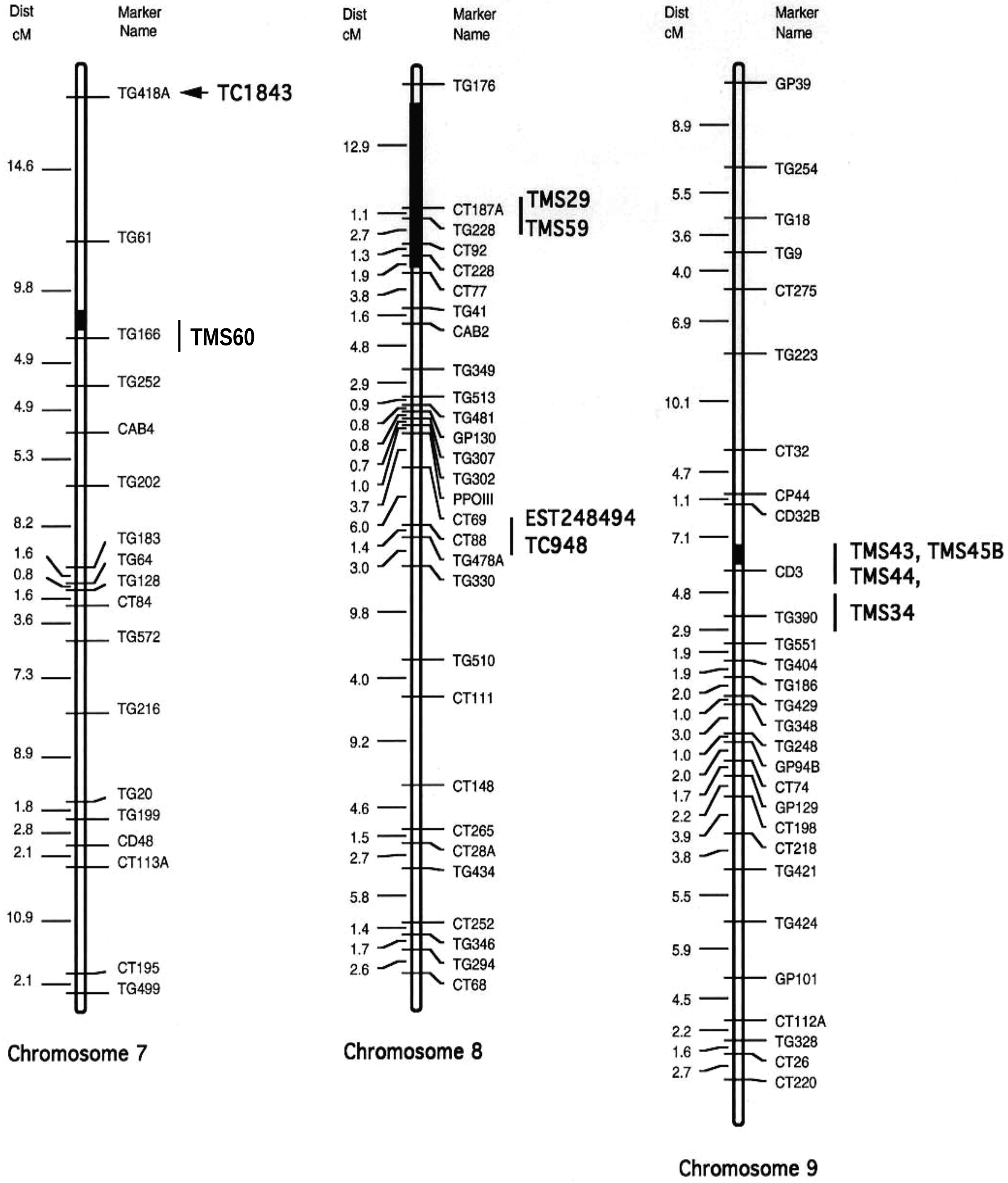


Figure 8. (Continued).

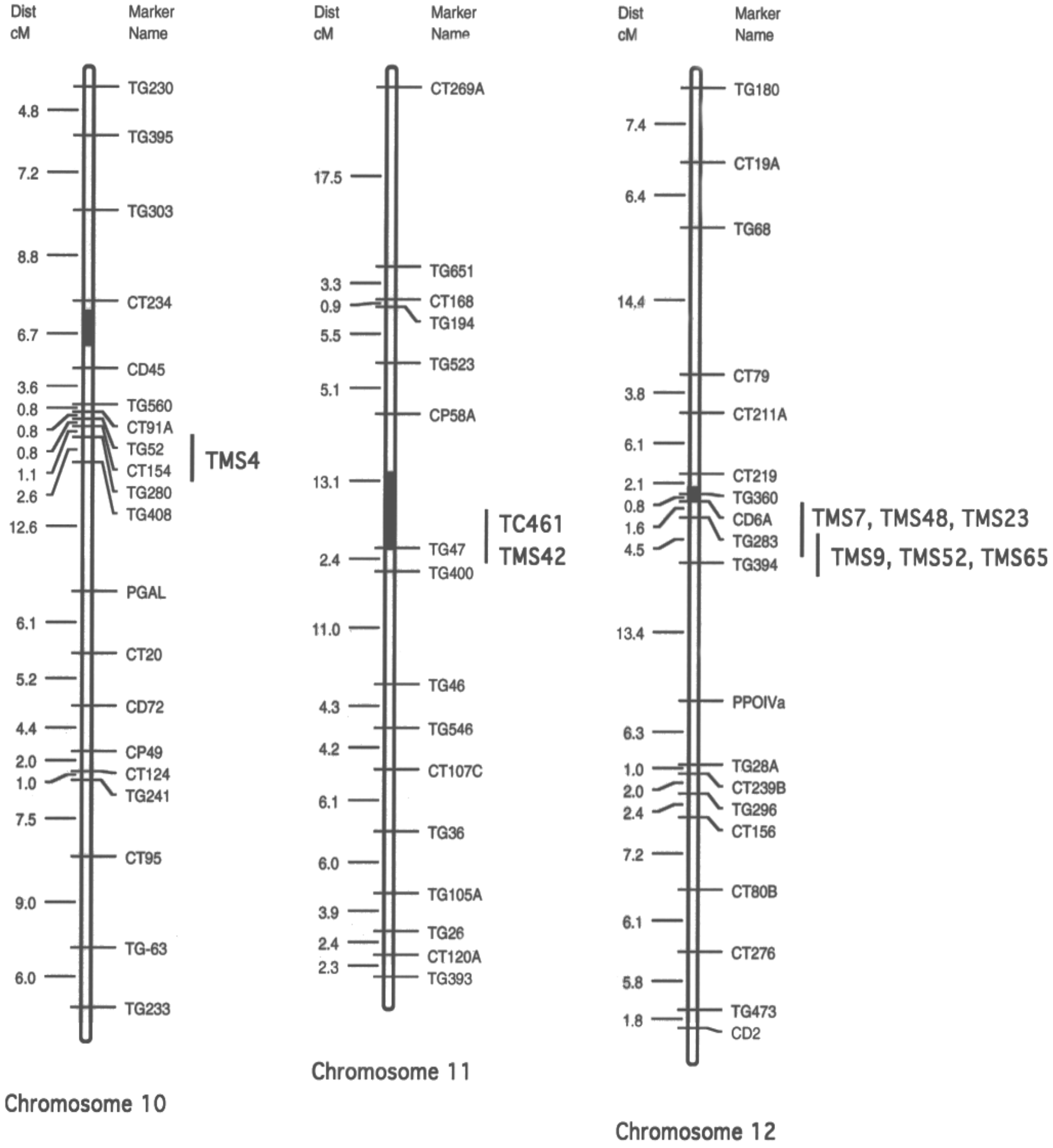


Figure 8. (Concluded).

3.6 Isolation and characterization of tomato centromere-associated sequences in YACs

3.6.1 YAC isolation and characterization

Most of the generated microsatellite markers representing different types and numbers of repeats did not map in a random fashion but were highly clustered near the centromeres of all twelve tomato chromosomes.

The highest number of microsatellites were mapped in the centromeric regions of chromosomes 1 and 12. Chromosome 1 is the largest one. The centromere on chromosome 1 has been identified by RFLP mapping and by *in situ* hybridization with 5S rDNA (Lapitan et al., 1991). Chromosome 12 is one of the smallest tomato chromosomes but it has nearly the same number of mapped microsatellites as chromosome 1. Some of the markers located on chromosomes 1 and 12 were used to isolate YAC clones for the characterization of these clones with respect to their DNA structure.

A tomato genomic library in yeast artificial chromosomes (Martin et al., 1992) comprising of 36,864 recombinant clones with inserts of tomato DNA of on average 250 kb (the equivalent of five haploid genomes) was screened by PCR with markers located on chromosomes 1 and 12. cDNA marker CT98 together with three cosegregating microsatellite markers TMS6, TMS17 and TMS45 located in centromeric region of chromosome 1, and RFLP marker TG283 together with cosegregating microsatellites TMS7, TMS9, TMS23 and TMS48 linked to the centromere on chromosome 12, were used for library screening.

Thirteen YAC clones containing centromeric sequences from chromosome 1, and twelve YACs harboring centromeric sequences from chromosome 12 were isolated. After selection of single colonies, followed by the large-scale yeast chromosome and YAC isolation, YACs and yeast chromosomes were separated by PFGE. Chromosome preparations of the host strain *Saccharomyces cerevisiae* AB1380 were used as a marker for size determination. The size of isolated YAC clones ranged from 150 and to 650 kb (Fig. 9A).

Blotted YAC gels were hybridized with a YAC vector-specific probe pBR322 DNA to characterize YAC size, number and relative stability (Fig. 9B). Use of purified YAC vector as a probe would result in crosshybridization because it contains sequences that are also present on endogenous yeast chromosomes. After initial analysis with a vector-specific probe, YACs were subsequently hybridized with an insert-specific GATA (Fig. 9C), GT and GA oligonucleotide probes.

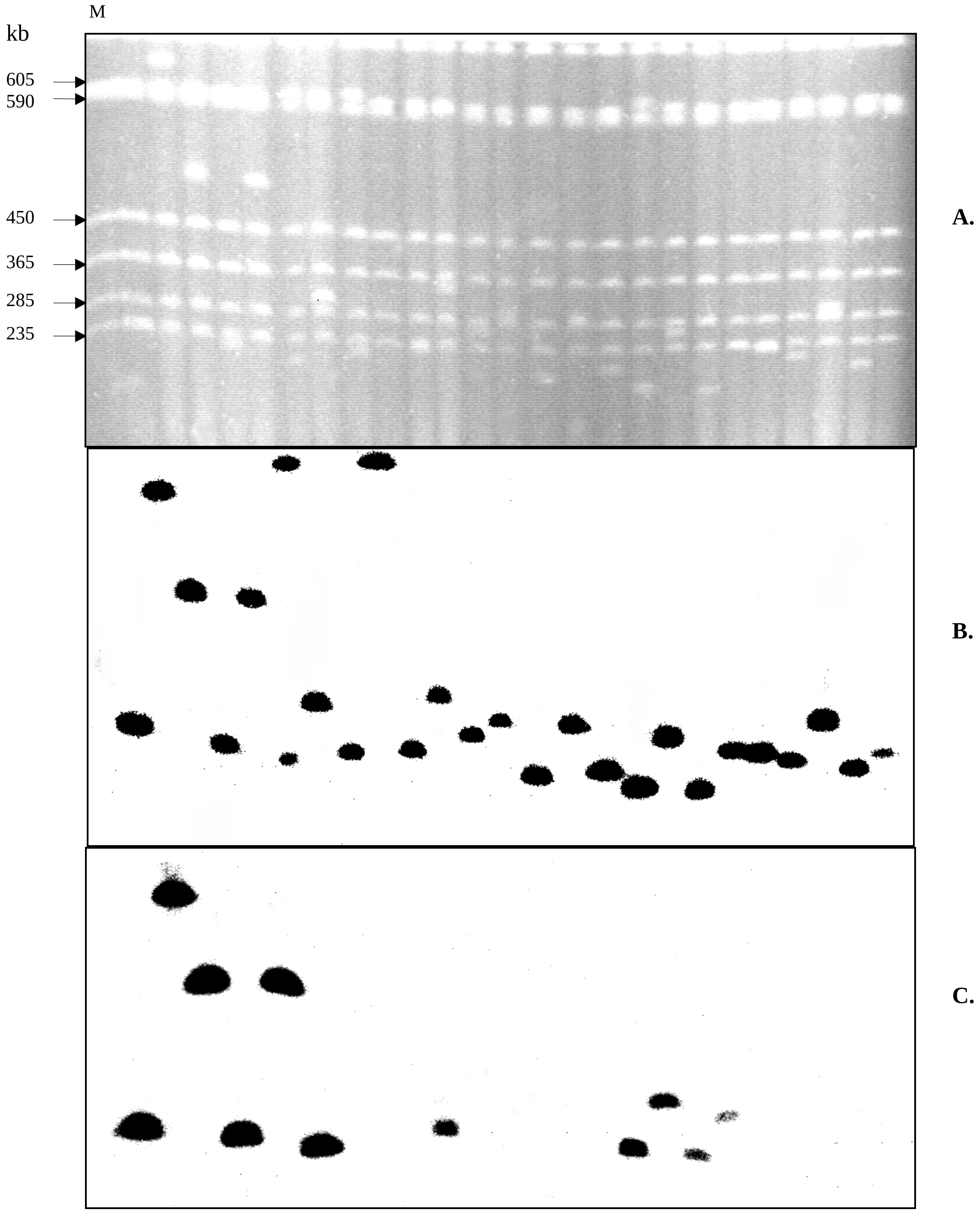


Figure 9. A.- PFG of tomato YAC clones was running at 148V, with a switch time of 60s for 48 hours and blotted; **B.-** hybridization analysis with YAC vector-specific probe pBR322; and **C.-** hybridization with insert-specific probe GATA; M - chromosome preparation of yeast strain *Saccharomyces cerevisiae* AB1380 as marker for size detection.

Hybridization with poly(GT) probe reproduced the gel picture and revealed, that all YAC clones as well as yeast chromosomes gave positive signals. This indicates that GT simple sequence repeats are present in the yeast genome as well. The poly(GA) probe hybridized only to two YAC clones that were isolated with marker TMS48.

The YAC clones which were isolated with different microsatellite markers but for the same region of chromosome did not show obvious overlaps. This indicates that even genetically cosegregating or tightly linked microsatellites are physically spread over a large region of heterochromatin, and that the clustering is most likely due to the highly suppressed recombination near the centromere.

3.6.2 Random subcloning

Two YAC clones derived from tomato centromere of chromosome 1 and three YAC clones harbouring centromeric sequences from chromosome 12 were chosen for subcloning (Table 14). For better separation of YACs from closely migrating endogenous chromosomes preparative PFG was run at 160V, with a switch time gradient of 20-60 s for 60 hours (Fig. 10). Under those conditions all five YACs of different sizes could be isolated from the gel with minimal contamination of endogenous yeast DNA.

Table 14. YAC clones containing tomato centromeric sequences from chromosome 1 and 12 that were chosen for subcloning.

No.	YAC clone	Size, kb	Isolated with marker	Located on chromosome
1	189F7	480	TMS6	1
2	666A3	220	TMS17	1
3	94F7	180	TMS9	12
4	187H2	260	TMS9	12
5	273G4	310	TMS48	12

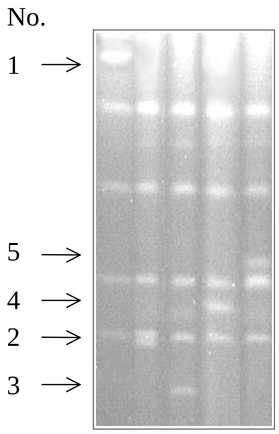


Figure 10. Preparative PFGE of these YAC clones. The gel run at 160V, with a switch time gradient 20-60 s for 60 hours for the isolation of YAC DNA.

Isolated YAC DNA was digested with the frequent-cutting restriction enzyme *Sau*3AI to generate relatively small fragments that would contain individual repetitive elements. Restriction fragments were purified and subcloned into plasmid vectors. An average of 2,000 recombinant clones were collected for each of the five YACs. PCR amplification on random chosen clones detected insert sizes from 100 to 2,000 bp with an average of approximately 300 bp.

To identify subclones containing highly repetitive sequences, high-density colony membranes were hybridized with total tomato genomic DNA. Only subclones containing sequences which were present in multiple copies in the tomato genome gave a strong signal and were selected. About 3% of the total number of subclones hybridized at high stringency to total genomic DNA.

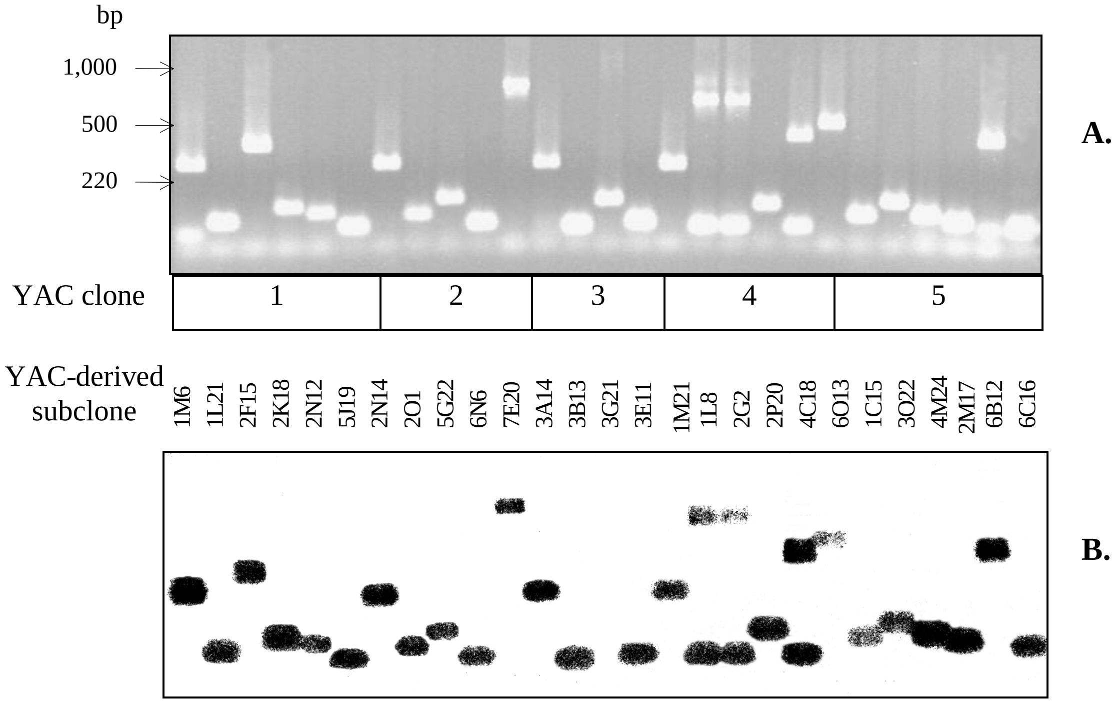


Figure 11. A. - PCR amplification of tomato centromeric DNA fragments in plasmids. **B.** - Hybridization with total tomato genomic DNA confirmed that the isolated clones were containing inserts with repetitive DNA.

10 to 15 subclones derived from each YAC clone that produced strong signals were purified and analysed by PCR. 4 to 6 subclones from each of the five YAC clones were chosen for

further characterization. PCR products of the selected subclones containing repetitive DNA were loaded onto an agarose gel and blotted (Fig. 11A). Hybridization with total tomato DNA showed that the selected clones harbored inserts with repetitive tomato DNA (Fig. 11B).

3.6.3 Hybridization of subcloned tomato centromeric sequences to genomic DNA

To study genomic organization and copy number of the isolated repeats, purified PCR-amplified centromeric sequences were surveyed by Southern hybridization to digested tomato DNA. The DNA from three *Lycopersicon* and one *Solanum tuberosum* lines was digested with different restriction enzymes, electrophoresed and blotted. Hybridizations with each probe under high-stringency conditions showed an interspersed repeat pattern, characterized as a smear superimposed over very weak defined fragments. Each probe hybridized to the genomic DNA from all tomato species but not to the potato DNA (Fig. 12).

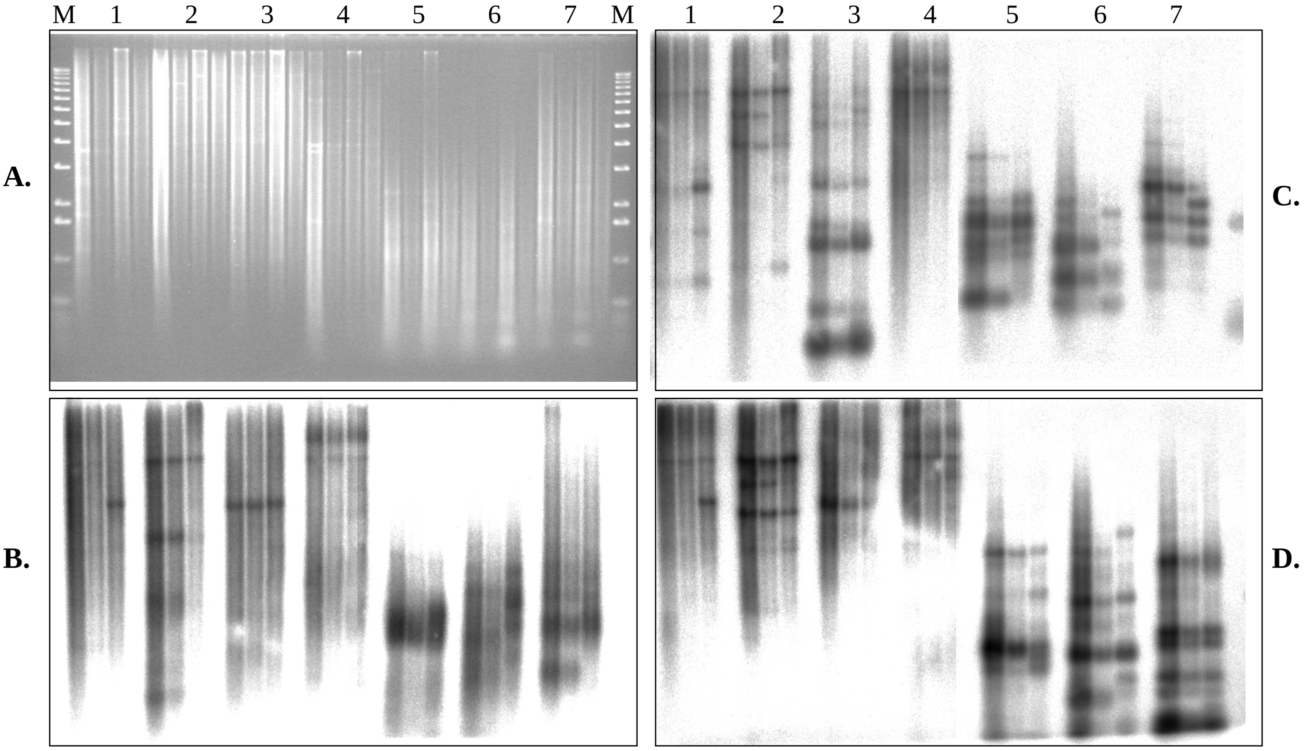


Figure 12. Southern analysis of centromere-derived sequences in the tomato genome. **A.** Genomic DNA of *Lycopersicon esculentum* VFNTcherry, *L. esculentum* TA55, *L. pennellii* LA716 and *Solanum tuberosum L.* was digested with (1) *Eco*RI; (2) *Eco*RV; (3) *Hind*III; (4) *Dra*I; (5) *Rsa*I; (6) *Taq*I; (7) *Hae*III, and probed subsequently with subcloned sequences. Hybridization pattern for some of them are shown: **B.** 41M21 as well as 56B12; **C.** 12F15 as well as 51C15; **D.** 33A14 as well as 53O22; M - 1kb-size standard.

All of isolated tomato centromeric DNA elements were present in *Lycopersicon* species but were completely absent in the potato genome, idicating the rapid evolution of these sequences that is typical of repetitive DNA (Zamir and Tanksley, 1988; Presting et al., 1996).

For *Eco*RI, *Eco*RV and *Dra*I restriction enzymes, which recognized sequence of 6 nucleotides, the fragments that hybridized to isolated repeats were often larger than 2 kb. In contrary, for *Rsa*I, *Taq*I and *Hae*III, which cleaved in 4 nucleotide recognition sequence, most of the DNA fragments containing isolated repeats were often smaller than 2 kb. All probes were polymorphic between *L.esculentum* and *L.pennellii* at least with half of the surveyed restriction enzymes.

3.6.4 Sequence analysis for the identity and homology search

After plasmid DNA preparation subcloned repetitive DNA fragments were completely sequenced from both directions. 22 obtained sequences were compared for the identity with the DNASIS computer program. Some of the sequences derived from the different YAC clones and different chromosomes showed significant homology to each other. For example, two sequences 22N14 and 33A14 both 450 bp long showed 100% identity (Fig. 13). The difference between two 90 bp sequences 26N6 and 33E11 was in four nucleotides.

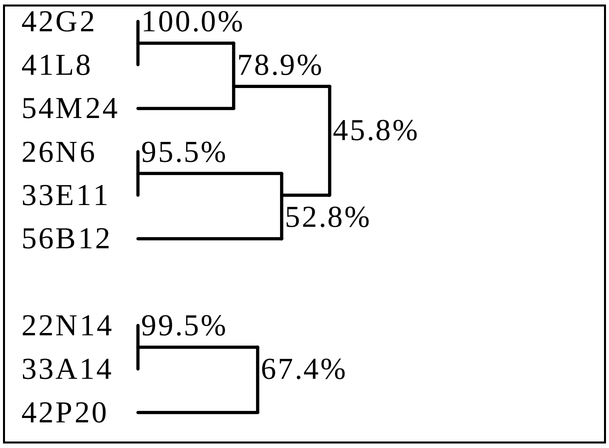
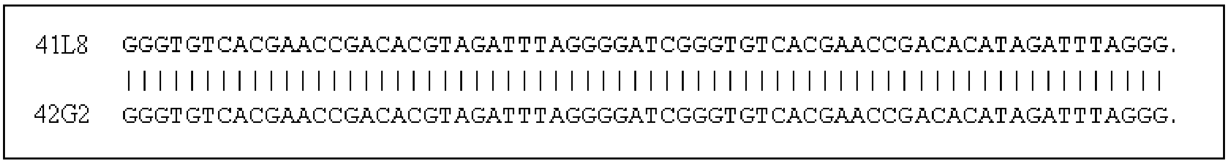


Figure 13. Percent of nucleotide similarities among sequences derived from tomato centromeric regions was estimated using the DNASIS v2.5 program.

Subclones 41L8 and 42G2 derived from the same YAC clone contained a 100% similar sequence of 66 bp consisting of two identical elements (Fig.14).



```
41L8  GGGTGTCACGAACCGACACGTAGATTTAGGGGATCGGGTGTCACGAACCGACACATAGATTTAGGG.
      ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
42G2  GGGTGTCACGAACCGACACGTAGATTTAGGGGATCGGGTGTCACGAACCGACACATAGATTTAGGG.
```

Figure 14. Two 100% homologous sequences 41L8 and 42G2, derived from 187H2 YAC clone, consist of two almost identical elements which were also found in several other subclones.

As well as in 41L8 and 42G2, this GGGTGTCACG(TA/AT)CCGACAC (GTAGATTTAGGG) consensus motif with slight nucleotide differences was also present in 15J19 and 26N6 sequences two times, and four times in 54M24 and 56B12, although these subclones were obtained from different YACs and different chromosomes. Alignment of this DNA element showed no homology to the database sequences.

By similarities in hybridization pattern to restricted genomic DNA and sequence analysis the subclones formed specific groups listed in Table 15. The A/T base content of the cloned sequences varied from 43 to 62%.

This table illustrates also the results of the alignment of these sequences to the GenBank+EMBL+DDBJ+PDB sequences using BLASTN. The majority of the analysed sequences showed no significant homology to the database sequences.

The major families of repeated DNA in the genome of *Lycopersicon esculentum* have been characterized (Ganal et al., 1988), among them two classes of interspersed tomato genome repeats TGRII and TGRIII. Isolated repetitive sequences did not reveal homology with known repeats in tomato. These preliminary data suggested that the isolated centromere-associated sequences represented new families of repeats in the tomato genome.

Table 15. List of subclones derived from five centromere-associated YAC clones presented in connection groups according to sequence homology and results of the alignment to the GenBank+EMBL+DDBJ+PDB sequence databases.

Subclone	Insert size, bp	YAC clone	Tomato chromosome	Alignment using BLASTN database search	Homology	
					bp	%
26N6	90	2	1	-		
33E11	90	3	12	-		
54M24	140	5	12	-		
52M17	102	5	12	-		
42G2	66	4	12	-		
41L8	66	4	12	-		
56B12	680	5	12	-		
42P20	200	4	12	-		
33A14	450	3	12	-		
22N14	450	2	1	-		
15J19	80	1	1	-		
12N12	120	1	1	-		
11M6	455	1	1	-		
12F15	605	1	1	*Gypsy*-like retrotransposon	396	92
25G22	225	2	1	-		
27E20	1,300	2	1	-		
33B13	65	3	12	-		
51C15	135	5	12	*T3*-type retrotransposon	79	83
53O22	225	5	12	-		
41M21	435	4	12	Tomato microsatellite region	219	86
46O13	810	4	12	*L.esculentum* RAPD band	221	85
46N17	1,800	4	12	-		

4. Discussion

PCR-based molecular markers such as microsatellites have been proven to be very useful in plant genome analysis because they detect higher levels of DNA polymorphisms than other marker assays and are amenable to automation. Large scale isolation of functional microsatellite markers has been undertaken for different plant species such as rice (McCouch et al., 1997); wheat (Röder et al., 1998b); potato (Milbourne et al., 1998); soybean (Cregan et al., 1999). Availability of highly polymorphic markers is of great importance especially for organisms with a narrow genetic base. Cultivated tomato is known for its low level of DNA polymorphism, that is why highly polymorphic markers are necessary for genome analysis in *L. esculentum*. So far, only a limited number of microsatellite markers has been developed for tomato. The majority of these markers were isolated from database sequences and used to identify tomato cultivars and accessions (Smulders et al., 1997; Bredemeijer et al., 1998). This research work was aimed mainly at the characterization of simple sequence repeats in the tomato genome, establishment of a set of functional markers for tomato and evaluating the utility of microsatellites in studies of genetic diversity within *L. esculentum*. Until now, only two microsatellites GA and ATT and several hypervariable fragments containing GATA repeats had been mapped onto the tomato genetic map and they were adjacent to centromeric regions (Arens et al., 1995; Broun and Tanksley, 1996), therefore, some clustering of markers was expected.

4.1 Long tomato microsatellites are predominantly associated with centromeric regions

4.1.1 Frequency of microsatellites in the tomato genome

The frequencies of microsatellites in the tomato genome were estimated based on hybridization analysis of phage libraries. The phage libraries used in this study were constructed from total genomic *Mbo*I DNA fragments with an average insert size of 1 kb. More than one-hundred thousand cloned sequences were analysed by plaque hybridization under stringent conditions (eliminating most repeats of n<10 from detection) with GATA and also with GA and GT oligonucleotide probes.

The frequencies of microsatellites in the tomato genome were estimated to be one GATA repeat every 1.7 Mp, one GA repeat every 290 kb, and one GT repeat every 180 kb. These numbers were about five times higher than described earlier by Broun and Tanksley (1996). Verifying the size of the tomato genomic libraries and using the same high stringency hybridization conditions, they analysed approximately 15,000 clones with different di-, tri- and tetranucleotide repeats and reported equal abundance of GA and GT repeats - one every 1.2 Mb and one GATA repeat every 6.3 Mb.

The frequencies of GA and GT microsatellites based on plaque or colony hybridization analysis reported for other higher plant species range widely from every 17 to 500 kb and from 86 to 800 kb respectively. For instance, GA microsatellites in *Arabidopsis* occur every 244 kb, GT every 430 kb (Bell and Ecker, 1994), in the wheat genome they are less abundant, one GA repeat occurs every 440 kb, and one GT repeat every 704 kb (Röder et al., 1995). The barley genome containes one GA repeat every 330 kb and one GT repeat every 620 kb (Liu et al., 1996). Prevalence of GA over the GT repeats seem to be a general feature of plant genomes (Morgante and Olivieri, 1993; Lagercrantz et al., 1993).

In the tomato genomic libraries, GT microsatellites are slightly more abundant than GA repeats. In contrary, GA repeats are more abundant than GT repeats in the tomato EST database. Only few GT microsatellites were found during database searches and all of them were adjacent to other types of repeats. AAT and complementary TTA trinucleotide repeats were the most abundant microsatellites in the EST database but they were often repeated less than ten times. GATA microsatellites were found neither in the sequence databases, nor in the plasmid libraries but were frequent in phage libraries.

Essential distinctions in frequencies of different motifs estimated in different genomic regions are a direct evidence that some simple sequence repeats are not randomly distributed in the tomato genome but tend to cluster at specific regions or cosegregate with other repetitive sequences. Therefore, the abundance of different microsatellites in the entire tomato genome can not be estimated precisely and defined genomic regions should be considered.

4.1.2 Repeat length and complex structure

DNA sequencing of clones containing SSRs revealed the unexpected prevalence of repeats that were complex in their structure. Approximately 90% of GATA repeats were associated with dinucleotide repeated motifs. Similarily, 75% of GT and 50% of GA microsatellites harbored arrays of AT and/or other simple sequence repeat types. A comparable complex structure was also observed for the majority of microsatellite sequences in the potato genome, where SSRs were represented by compositions of various repeat types (Milbourne et al., 1998). Many of the dinucleotide microsatellites identified in soybean genomic libraries (Powell et al., 1996) were compound in nature containing (GA)n(GT)n or (GA)n(AT)n motifs.

Moreover, the isolated tomato microsatellites contained a high numbers of repeats. The shortest dinucleotide repeat contained 12 repeated units. Several of the sequenced dinucleotide and GATA arrays contained more than 100 repeats. With an average of more than 30 repeating units, they represent a class of microsatellite markers in tomato that has not yet been investigated in detail. High repeat numbers were not a consequence of hybridization conditions, because wheat microsatellites that were isolated in this laboratory using the same hybridization conditions had on average a much lower number of repeated dinucleotide units (Röder et al., 1995).

This predominance of compound repeats was unusual when compared to microsatellite repeat structures observed in other plant species where the majority of the isolated microsatellites contained a single repeat type and a lower number of repeating motifs. In wheat, barley, rice and maize such long and complex microsatellites represent much less than 50% of the total number.

Length and complexity of the tomato microsatellites were associated with difficulties during sequencing, primer designing and PCR amplification. It was often a problem to obtain the complete sequence of complex microsatellites especially of those containing long stretches of AT repeats. Up to 80% of the tetranucleotide GATA repeats could not completely be sequenced because of extremely large numbers of motifs and the total length of microsatellite exceeding 100-150 bp. Moreover, microsatellite markers with long and complex repeats often required lower annealing temperature, and as a result, the background in PCR amplification increased.

Screening tomato small-insert genomic libraries with GA, GT and GATA oligonucleotides showed that the number of positive clones did depend on the insert size of the library. The larger the insert size of the library, the higher the number of microsatellite containing clones. Tomato plasmid libraries with an average insert size of 200-250 bp revealed less than 0.1% of microsatellite containing clones. Nearly the same number of microsatellites was found in EST sequences. No GATA and trinucleotide microsatellites were isolated from the plasmid libraries, mainly because of two possible reasons: 200-250 bp genomic sequences cloned in plasmids were too short to contain complex microsatellite repeats which are usual for tomato, or such repeats were not present in single copy DNA.

The 1 kb insert size in phage libraries resulted in 0.4% positive clones compared to the total number of recombinant clones. Thus, extending the length of genomic DNA fragments would increase the frequency of microsatellite containing clones because the larger the insert the more likely it is to contain a microsatellite. The limitation is that inserts of more than 1.5 kb can not be sequenced without additional primer designing which increases costs of marker development.

4.1.3 Allelic variability

Newly developed tomato microsatellite markers regardless of the method of isolation were found to be highly polymorphic. The variability of the isolated microsatellites was much higher than that was found previously in tomato. More than 90% of the microsatellite markers isolated from genomic libraries and all markers generated from EST sequences were polymorphic between *L. esculentum* and *L. pennellii*. This is nearly twice as much as the percentage described by Smulders et al. (1997) for tomato microsatellites with shorter number of repeating units extracted from databases and four times the rate described by Broun and Tanksley (1996). Within the *L. esculentum* gene pool, the number of alleles was higher than described previously for tomato. Not taking presence/absence of amplification into account, approximately 75% of the microsatellites display more than one amplified allele in the twelve investigated tomato varieties. This extends the results of Smulders et al. (1997), where only one third of analysed microsatellite markers with repeat numbers up to twenty generated polymorphic fragments among tomato cultivars.

The difference in reported variabilities for tomato microsatellites was not directly dependent on the length of the microsatellite array. Several examples support this assumption. Two markers, TMS2 and TMS37, were isolated from tomato phage libraries and detected three and five alleles respectively in the same set of tomato lines. The TMS2 primer pair flanking the complex repeat $(GT)_{41}(TA)_6(CT)_9$ amplified long fragments of 387 and 400 bp in *L. esculentum* varieties and a 372 bp allele in *L. pennellii*. In contrast, marker TMS37 flanking the significantly shorter microsatellite array $(GA)_{21}(TA)_{20}$ detected four alleles in the size range from 148 to 173 bp in cultivated tomatoes and an 180 bp fragment in *L. pennellii*. In addition, microsatellite markers generated from EST sequences which were containing relatively short simple sequence repeats detected comparable level of polymorphism to those of longer microsatellites isolated from genomic libraries.

No obvious correlation between repeat type, complexity of structure or repeat length with level of polymorphism could be observed in tomato. This was comparable to barley (Liu et al., 1996) and potato (Milbourne et al., 1998). In soybean, GT microsatellites were more informative than GA (Powell et al., 1996). In contrast in *Arabidopsis,* it was found that GT repeats were less polymorphic than GA and the complexity of microsatellite structure decreased the level of variability (Bell and Ecker, 1994). It would be of interest to establish, which motif of the complex repeats is responsible for the length polymorphism.

It is important to mention that approximately half of the microsatellites did not amplify a product in *L. pennellii. L. esculentum* and *L. pennellii* are distantly related and show considerable level of DNA variation. Two microsatellite markers, TMS43 and TMS44, which are linked to the TMV resistance gene *Tm-2a,* specifically amplify only fragments from DNA introgressed from *L. peruvianum* in cultivated tomato. These markers displayed two alleles and amplified fragment in three *L. esculentum* varieties carrying TMV resistance gene *Tm-2a* and have a null allele in all other tomato lines including *L. pennellii.* In addition, only a very small percentage of markers were able to produce fragments of the expected size in related solanaceous species.

The majority (90%) of the random tomato genomic sequences, both single copy and repetitive, at high stringency hybridization conditions failed to detect homologous sequences even in closely related species, and only 50% of single copy cDNA clones were well conserved (Zamir and Tanksley, 1988). This is a fact which has similarily been observed for other plant families. Especially in wheat, microsatellite markers do only amplify loci from one of the three A, B, or D genomes and are usually unable to produce fragments in the closely related species barley and rye (Röder et al., 1995). It suggests that not only the array length of microsatellites is highly variable but also flanking regions of microsatellite repeats, and, on the other hand, that the majority of microsatellite containing sequences are more likely correspond to noncoding DNA. As a matter of fact, the data obtained in this study suggest that large numbers of alleles in *L. esculentum* are frequently associated with null alleles in

L. pennellii and that microsatellites with low levels of variability do more frequently amplify fragments from *L. pennellii*.

Previous attempts to identify polymorphisms between modern tomato cultivars on the basis of protein profiles or isozymes had not been entitely successful due to the lack of sufficient variation. It was possible to distinguish cultivars on the basis of one or more unique RFLP with an average distance value of 0.01 to 0.04 between *L. esculentum* lines including VFNTcherry (Miller and Tanksley, 1990; Broun et al., 1992). Genetic distances between tomato cultivars estimated using microsatellite markers ranged from 0.21 to 0.74 with an average value of 0.42. Because of the high level of polymorphism in *L. esculentum* detected with microsatellites, the markers described here are highly suitable for the variety and hybrid identification and mapping of agronomically important genes or QTLs within *L. esculentum.*

4.1.4 Genomic distribution

A total of 41 microsatellite markers detecting 43 independent loci were mapped onto the high-density molecular linkage map of tomato. 22 markers which were isolated from the tomato phage libraries were highly centromere-linked on the respective chromosome. This was expected for large arrays of the GATA repeats since this was shown previously by direct genetic mapping using Southern hybridization (Arens et al., 1995; Grandillo and Tanksley, 1996). The mapping data of the relatively short clusters of GATA repeats which have been isolated in this study and can not be detected by hybridization reveal that they are also located near the centromeres. The genetic mapping of dinucleotide repeats has confirmed that not only GATA repeats but also other simple sequence repeats are highly associated with the tomato centromeric regions.

In order to find markers in genomic regions distant from the centromeres, *Pst*I libraries in plasmids were constructed and screened for microsatellite repeats. This methylation-sensitive restriction enzyme yields genomic libraries enriched for single and low copy DNA because it does not cleave highly methylated repetitive DNA which is known to be located in heterochromatic pericentromeric and telomeric regions. It was expected that markers isolated from the libraries enriched for single copy sequences should yield a more random distribution in the genomic euchromatin. Nine markers isolated from plasmid libraries harbored long arrays of GA or GT motifs and eight of nine were associated with AT repeats. Genetic mapping data of all microsatellites have revealed also a centromeric location.

To answer the question whether the tomato genome contains microsatellites in its expressed regions, the screening of the tomato EST database for the presence of microsatellites was performed. Less than 0.2% of database sequences harbored microsatellites, and short trinucleotide repeats AAT and TTA prevailed among them. Among dinucleotide repeats, AT motifs were the most frequent followed by GA repeats. Eleven markers were generated and ten could be mapped onto the genetic map. Two markers containing only TA motifs repeated 14 and 20 times were mapped in the centromeric regions of chromosome 6 and 4. The other eight markers containing AT, GA, GT and some other repeat types and combinations thereof were localized distant from the centromeres on different chromosomes.

The mapping results demonstrate that the majority of the tomato microsatellites are not randomly distributed along the chromosomes but tend to cluster in centromeric regions. 35 of 43 microsatellite loci were mapped near the centromeres of the tomato chromosomes, and the number of markers per chromosome ranged from one on chromosomes 6 and 10 to six on chromosome 12 and ten on chromosome 1 showing no apportionment to the physical length of these chromosomes. This deviation may be due to limited number of markers investigated to date, and differences in SSR frequency per chromosome can not be ruled out at this time.

All microsatellites that have been mapped near the tomato centromeres had different types and number of repeated motifs, therefore, centromeric location is not predicted by the repeat type and length. Such a clustering of microsatellite repeats has not yet been shown for other plant species by genetic mapping. In barley, maize and rice, as well as in Arabidopsis, soybean, and potato genetic mapping of a number of microsatellite sequences clearly indicated that microsatellite markers are nearly random distributed in the genome. In wheat, it has been shown that microsatellite markers are not only randomly distributed on the genetic map but also along physical length of the chromosomes (Röder et al., 1998a,b).

A recently created high-density AFLP map of tomato spanned 1482 cM and contained 1078 AFLP markers obtained with *Eco*RI+*Mse*I and *Pst*I+*Mse*I primer combinations (Haanstra et al., 1999). The *Eco*RI/*Mse*I AFLP markers were not evenly distributed along the chromosomes. Around the centromeric region, the vast majority (848) of *Eco*RI/*Mse*I AFLP markers were clustered and covered a genetic distance of 199 cM, corresponding to one *Eco*RI/*Mse*I AFLP marker per 0.23 cM. In the distal parts 1283 cM were covered by 230 *Eco*RI/*Mse*I AFLP markers, corresponding to one marker per 5.6 cM. As it was expected that the methylation-sensitive restriction enzyme *Pst*I would recognise restriction sites in nonmethylated euchromatin, the *Pst*I/*Mse*I AFLP markers showed a more even distribution with 16 *Pst*I/*Mse*I AFLP markers covering a genetic distance of 199 cM around the centromeric regions and 81 *Pst*I/*Mse*I AFLP markers covering a genetic distance of 1283 cM on the more distal parts of the chromosomes, corresponding to one marker per 12 and 16 cM respectively. The clustering of 78.7% of the mapped *Eco*RI/*Mse*I AFLP markers was explained as a result of a much lower frequency of recombination nodules in heterochromatin compared to euchromatin observed by Sherman and Stack (1995) rather than to a nonrandom distribution of markers on the chromosomes. In contrast to the mapped *Pst*I/*Mse*I AFLP markers, no microsatellites isolated from tomato genomic libraries were found in distal euchromatic regions even with the use of the methylation-sensitive restriction enzyme *Pst*I. It is not known whether the clustering of simple sequence repeats reflects a functional role of these sequences or is merely the result of reduced recombination at the centromeres. Due to the suppressed recombination in this area, approximately 35% of RFLP markers are also clustered in centromeric regions.

The five microsatellites isolated from genomic libraries are all tightly linked with TG283 in the centromeric region on chromosome 12, but could not be linked by the isolated YAC clones spanning hundreds of kilobases. For example, in the centromeric region of tomato chromosome 9, the ratio of genetic to physical distance is 5 Mb per 1 cM (Pillen et al., 1996) indicating extremely suppressed recombination in this area. At the moment it is not clear, how large the regions are which contain the centromeric microsatellite sequences in tomato. The maize B chromosome centromeric region, where the 1.4 kb B centric repeat element is arranged in a degenerate tandem array with a minimum of interspersed sequences, is estimated to be at least 9 Mb in size (Alfenito and Birchler, 1993; Kaszás and Birchler, 1996). In *Arabidopsis*, the regions of centromere occupy approximately 1-3 Mb per chromosome (Round et al., 1997; Jackson et al., 1998; Fransz et al., 2000). It is also not clear how close the association of these simple sequence repeats is with the actual tomato centromeres. The fact that tomato centromeres are associated with simple sequence repeats requires further investigation since this has been observed also in sugar beet and some other plants by *in situ* hybridization (Schmidt and Heslop-Harrison, 1996). To answer these questions, it will be necessary to delimit the position of the centromeres relative to the associated microsatellite markers by the use of telotrisomics (Frary et al., 1996) or induced deletion stocks (Liharska et al., 1997; Weide et al., 1998). Only then, the question can be answered whether simple sequence repeats are associated with or an intrinsic feature of tomato centromeres.

4.2 Use of microsatellite markers for isolation and characterization of tomato centromeric sequences

The predominant centromeric clustering of tomato microsatellites reduces their value as molecular markers in genome mapping experiments, but such markers can be extremely useful in investigation of the fine molecular organization of centromeric regions in the tomato genome. To answer the question, what other repetitive sequences except microsatellites constitute tomato heterochromatic regions around the centromeres, the centromere-associated microsatellite markers were used for isolation of homologous YAC clones and characterization of these clones with respect to their molecular structure.

Twenty-five YAC clones were isolated using microsatellite markers cosegregating in centromeric regions of chromosomes 1 and 12. Hybridization analysis revealed that isolated YACs were containing microsatellite repeats with respect to the marker which was used for their isolation. Interestingly, all yeast chromosomes give positive signals in hybridization with polyGT oligonucleotide probe, indicating that GT microsatellites are frequent in yeast genome.

Despite of isolation with genetically tightly linked markers, YAC clones did not overlap assuming the total length of cloned DNA of approximately 6 Mb. The largest YACs containing centromeric sequences from chromosomes 1 and 12 were selected for subcloning. Centromere-derived *Sau*3A fragments cloned in plasmid vectors were analysed by hybridization with genomic DNA. Only 3% of the cloned sequences corresponded to highly repetitive DNA. According to findings of Zamir and Tanksley (1988) and Peterson et al. (1996), the amount of repetitive DNA in centromeric heterochromatin should be at least 20%. It is possible, that cloned genomic DNA fragments are unlikely to contain portion of the large uninterrupted blocks of centromeric tandem repeats. On the other hand, no tandemly repeated DNA sequences except the telomeric satellite TGRI at centromeric sites of two chromosomes and microsatellites have been found to date to be associated with centromeres in tomato.

The major families of repeated DNA in the tomato genome have been characterized (Ganal et al., 1988). Tomato genome repeat TGRI is the most highly repeated 162 bp satellite DNA with 77,000 copies representing approximately 12.5 Mb or 1.75% of the total DNA (calculated based on the genome size of 7.14x10^5 kb (Galbraith et al., 1983)). *In situ* hybridization showed that TGRI is clustered at or near the telomeres of most of the chromosomes, as well as at some interstitial sites and was also observed at the centromeres of chromosomes 2 and 3. As it was noted, despite its prominence in the genome, this tandem repeat was extremely underrepresented or even absent in some genomic libraries. Another family of tandem repeats consisting of the genes coding for the 45S ribosomal RNA represents 3% of the genome, is located at a single locus at the end of chromosome 2 and in contrast to other families of repeats, is conserved in *Lycopersicon* and other *Solanaceae* species. TGRII is a highly interspersed repeated DNA sequence with 4,200 copies per genome labeling the entire length of most chromosomes with no evidence for clustering as confirmed by *in situ* hybridization. TGRIII with approximately 2,100 copies is an interspersed repeat with some clustering at or near the centromeres in several chromosomes. Altogether, the four repeat families account for approximately 5% of the total nuclear DNA in tomato. Hybridization analysis of TGRI satellite repeats to digested genomic DNA provides evidence that there are no other high copy number, tandemly-arranged sequences in tomato.

Only a limited number of sequences were characterized at tomato centromeric regions including the TGRIII interspersed repeat, telomere-homologous low copy sequences (Presting et al., 1996), paracentromeric low copy and middle repetitive sequences on tomato chromosome 6 (Weide et al., 1998) and microsatellites (Arens et al., 1995; Broun and Tanksley, 1996). In addition, analysis of the sequence flanking one side of a long centromere-associated GACA repeat showed high homology to a *Lilium henryi* retrotransposon (Vosman and Arens, 1997). However, other middle and highly repetitive sequences may also contribute to this heterochromatic area.

YAC-derived subclones which contained segments of repeated DNA were randomly selected and studied in respect of genome organization and sequence similarities. Characterization of new families of repetitive DNA in the tomato genome may lead to additional insights on the events that have occurred during the evolution of *Lycopersicon*.

4.2.1 Genome organization and specificity of centromere-derived sequences

Southern blot hybridization of cloned repetitive sequences to digested genomic DNA was used to study the organization of these repeats in the genome of tomato. Digested potato DNA was also included to determine the conservation of repeats in related species. As expected, all 27 cloned tomato centromeric sequences, like TGR sequences and repetitive DNA in general, were present in *L. esculentum* and *L. pennellii* but were absent in the potato genome indicating their rapid divergence since the evolution of *Lycopersicon* from a section of *Solanum*. The rapid evolution of these sequences may be also an indication of no functional importance.

All cloned sequences showed an interspersed repeat pattern with the surveyed restriction enzymes and no defined pattern in the form of a ladder structure was observed. A series of hybridizing fragments were polymorphic between *L. esculentum* and *L. pennellii* for each of the probed sequences providing the possibility for genetic mapping. It is possible, that fragment separation by PFGE together with additional restriction enzymes could result in a clearer picture of organization of the isolated repetitive sequences.

4.2.2 DNA sequence analysis of isolated repeats

Sequence analysis provides direct information on sequence organization and allows the study of homogeneity or heterogeneity of different members of a repeat family.

Twenty-two sequenced clones contained fragments in the range of 65 to 1,300 bp. Multiple alignment showed that 59% of tomato centromere-associated sequences detected nucleotide similarities with each other. For example, 90 bp clones 26N6 and 33E11, as well as 450 bp clones 22N14 and 33A14, derived from different chromosomes showed 100% identity. According to sequence similarities and hybridization patterns to restricted genomic DNA, clones formed connection groups. Up to 50% identity could be detected between connection groups. The repeated motif GGTGTCACGT(A/T)CCGACAC was found to be responsible for these identities.

Two sequences 12F15 and 51C15 of 605 and 135 bp, 57.6 and 56.7% AT-rich, isolated from centromeric region of chromosome 1 and 12 respectively showed high homology to tomato *T3/gypsy*-like retrotransposons and a resembling hybridization pattern but inessential (23%) sequence similarity indicating that they were derived from different regions of the retrotransposon. In addition, these sequences were not found in the potato genome. *T3/gypsy*-like retrotransposons are abundant in the tomato genome (over 6,500 copies of an 8 kb LTR retrotransposon). Sixteen *T3/gypsy*-like retrotransposon sequences characterized in the genome of tomato cultivars *L. esculentum* and wild *Lycopersicon* species previously (Su and Brown, 1997) showed high sequence heterogeneity.

4.2.3 Molecular structure of centromeric regions in eukaryotes

The well characterized fission yeast *Schizosaccharomyces pombe* centromeric regions are 50-150 kb consisting of a 4-7 kb central single copy core surrounded by middle repetitive elements organized in chromosome specific arrays which are essential for centromere function (Baum et al., 1994; Smith et al., 1995). A number of plant centromeric sequences have been described that may contribute to centromere function. The 180 bp tandem repeat in *Arabidopsis* with a genomic organization that resembles the 170 bp alphoid repeat arrays at primate centromeres (Willard, 1990) is positioned around the centromere on each chromosome (Maluszynska and Heslop-Harrison, 1991). A number of middle and low copy nontandemly repeated sequences and the repeats which are a diverged copy of the LTR of the *Arabidopsis* retroelement *Athila* (Pelissier et al., 1995) have been found to be associated with the 180 bp repeat (Thompson et al., 1996a,b). Complete sequencing within the genetically defined centromeres on chromosomes 2 and 4 has revealed the presence, distribution and diversity of specific repetitive elements on individual chromosome (Lin et al., 1999). In addition, representatives of previously described intermediate copy repeats as well as some new repeats were identified. Not only the accumulation of retroelements including members of the *LINE*-like Ta11 elements, LTR elements of both the *Ty3/gypsy* and *Ty1-copia* family, and the *Athila* family, and other transposons in the pericentromeric regions was detected but also a number of intact genes some of which appear to be expressed.

The maize B chromosome centromeric region contains the 1.4 kb B centric repeat which is arranged in a degenerated tandem array with a minimum of interspersed sequences and shows high homology to the maize knob sequences which can act as neocentromeres in certain genetic backgrounds (Alfenito and Bichler, 1993). Analysis of cloned centromeric segments from the maize chromosome 9 revealed two major classes of repeated sequences: CentC and CentA (Ananiev et al., 1998). CentC is a tandem repeat which forms clusters of different sizes at centromeric sites of all maize chromosomes with any obvious homology to the maize knob-associated tandem repeat. CentA has a structural similarity to retrotransposable elements because it is composed of a segment of DNA flanked by two LTRs and a short segment of homology to the sorghum pSau3A9 element. Thus, centromeric regions of maize chromosomes contain blocks of tandemly repeated elements interspersed with blocks of retrotransposable elements.

The two repetitive DNA elements, pSau3A9 and CCS1, were isolated from sorghum (Jiang et al., 1996) and *Brachypodium sylvaticum* (Aragón-Alcaide et al., 1996) centromeres respectively. High-resolution *in situ* hybridization experiments localize these repeats onto the primary constrictions of various grass species, showing sequences conservation across a variety of monocotyledonous species. They are the only plant centromeric repetitive DNA elements described to date which are conserved in distantly related species suggesting that these elements may play a role in centromere function. By using pSau3A9 element as a probe, rice BAC clones were isolated (Dong et al., 1998). Seven distinct families of repeats were identified representing interspersed middle repetitive elements and one tandem repeat. All of them are present on every rice centromere as demonstrated by *in situ* hybridization. The same sorghum pSau3A9 element was used for the isolation of a barley homolog (Presting et al., 1998). This barley homolog had high similarity to the integrase region of the polyprotein gene of *Ty3/gypsy* group retrotransposons. It was shown by FISH, that the entire polyprotein gene and flanking sequence including the presumed LTR, were present at barley centromeres.

The recent identification of plant centromeric sequences shows complex and heterogeneous sequence arrangement of plant centromeres, accumulation of different retrotransposable elements and often tandem repeats which are highly interspersed with other repetitive sequences including also single copy genes. It has been known, that eukaryotic centromeric heterochromatin is associated with highly suppressed recombination, and tends to accumulate repetitive sequences. The amount of repetitive DNA families necessary for centromere function is unknown. The initiation of kinetochore formation may simply require the presence of a minimal amount of appropriately spaced binding motifs (e.g. the CENP-B box). It is also possible, that conserved middle repetitive elements found in centromeric regions of several different plant species play a key role in centromere function while arrays of tandem repeats play a supporting role. Also emerging is the idea, that centromere function is not strictly tied to a particular DNA sequence but can be controlled epigenetically by processes that operate above the level of the primary nucleotide sequence.

The centromere of the *Drosophila* minichromosome Dp1187 is a 420 kb region which is necessary and sufficient for fully stable chromosome inheritance and is primarily composed of pentanucleotide AATAT and AAGAG satellites interspersed with transposable elements (Sun et al., 1997). None of these sequences have been found to be centromere-specific or present at all centromeres. The data obtained for *Drosophila* suggest that the majority of the centromeric sequences are not specific to centromeres. If there are specific sequences involved in centromere function, they must comprise a minor portion of the regions required for full function and may even differ among individual centromeres. Alternatively, centromere function may be provided by a specific three-dimentional high order structure (HOS) which may be under the control of epigenetic mechanisms (Copenhaver and Preuss, 1999).

Various sequences were characterized in the tomato centromeric regions among them transposable elements, low copy sequences with homology to telomeric repeats and different types of interspersed repeats which could be transposable elements as well. This study revealed the predominant centromeric location for different microsatellite sequences and a new class of interspersed repeats. It might be possible that complex tomato microsatellites could play a role at centromeres of tomato similar to satellite DNA in centromeric regions of other plant organisms.

Further studies would be desirable for a more precise characterization of the isolated repetitive tomato centromeric sequences. The chromosomal location of interspersed repeats can be identified using *in situ* hybridization or RFLP mapping by PFGE. It is possible, that a number of the isolated sequences are interspersed with each other in the centromeric regions of all chromosomes or specific for a single centromere. Reconstruction experiments are necessary to determine the precise copy numbers of newly isolated repeats in the tomato genome as well as identification of the complete repeat unit by screening of phage genomic libraries. Genome/species specificity experiments of these repeats could help to understand the evolutionary events in the genome of tomato. Furthermore, since there is evidence that tomato centromeres comprise a large portion of single copy sequences, an analysis of single and low copy clones could yield interesting results.

Altogether these data would extend our knowledge about the sequence organization of tomato centromeric regions providing possibility of comparative analysis with other plant centromeres. Knowledge of the fine structure of a fully functional centromere will help elucidate the biochemical nature of DNA architectures and DNA-protein interactions at the centromere which will be critical to understanding of centromere function and the efficient construction of artificial chromosomes.

5.1 Abstract

Availability of highly polymorphic molecular markers is of great importance in genetic mapping and genome analysis in organisms with low levels of variation. Microsatellites are becoming the marker assay of choice, because they reveal high levels of polymorphism and are amenable to automation.

This research work was aimed at isolation, characterization and mapping of microsatellites in tomato. A considerable number of markers was isolated from different genomic regions by screening two types of genomic libraries and the tomato EST database. Sequence analysis revealed that the majority of microsatellites were complex, representing combinations of different types of repeated motifs. Moreover, tomato microsatellites were characterized by a high number of repeats that was unusual compared to other plant species. Such features of tomato microsatellites influenced negatively the efficiency of marker isolation.

Despite the complex structure, microsatellites display high variability in *Lycopersicon esculentum* varieties. Newly isolated markers detected up to five alleles in a set of twelve *L. esculentum* lines and the majority were polymorphic between *L. esculentum* and *L. pennellii*. Genetic distances between tomato cultivars estimated by using fourteen markers ranged from 0.21 to 0.74 with an average of 0.42. Due to the large number of alleles in *L. esculentum* gene pool, the markers can be used in genotyping tomato cultivars and accessions.

A total of 41 markers detecting 43 independent loci were mapped onto the tomato high-density molecular linkage map. All 31 markers isolated from genomic libraries mapped exclusively near the centromeres of different chromosomes. Such clustering of genomic microsatellite repeats has not been described for other plant species. In contrast, only two markers of ten generated from EST sequences were centromere-linked. The other eight were randomly distributed in euchromatin.

The fact of predominant centromeric clustering of tomato microsatellites isolated in this study decreases their value in genetic mapping experiments but these markers provide a unique opportunity for the molecular characterization of centromeric regions on individual tomato chromosomes. By the isolation of YAC clones homologous to the centromere-associated microsatellites, it was possible to characterize a number of new repeated DNA sequences from the tomato genome.

5.2 Zusammenfassung

Die Verfügbarkeit von hochgradig polymorphen molekularen Markern ist für die genetische Kartierung und Genomanalyse von Organismen mit geringer Variabilität von großer Bedeutung. Mikrosatelliten werden zunehmend zum Markersystem der Wahl, weil sie einen hohen Grad von Polymorphismus zeigen und ihre Anwendung automatisierbar ist.

Ziel der vorliegenden Forschungsarbeit war die Isolierung, Charakterisierung und Kartierung von Mikrosatelliten für Tomate. Durch die Sichtung von zwei Typen von genomischen Bibliotheken und der Tomaten-EST-Datenbank konnte eine beträchtliche Anzahl von Markern aus verschiedenen Regionen des Genoms isoliert werden. Die Sequenzanalyse zeigte, daß diese Mikrosatelliten überwiegen komplex aufgebaut, also aus verschiedenen repetitiven Motiven bestehen. Darüber hinaus zeichnen sich die Mikrosatelliten aus der Tomate durch eine hohe Zahl von Wiederholungen aus, was im Vergleich zu anderen Arten ungewöhnlich ist. Diese besonderen Merkmale der Tomatenmikrosatelliten haben einen negativen Einfluß auf die Effizienz der Markerisolation.

Trotz ihres komplexen Aufbaus zeigen die Mikrosatelliten eine große Variabilität zwischen verschiedenen Sorten von *Lycopersicon esculentum*. Die isolierten Marker zeigten bis zu fünf Allele in zwölf Linien von L. *esculentum*, und die meisten waren zwischen *L. esculentum* und *L. pennellii* polymorph. Die anhand von 14 Markern geschätzte genetische Distanz von Tomatensorten liegen zwischen 0.21 und 0.74 mit einem durchschnittlichen Wert von 0.42. Wegen der großen Zahl von Allelen im *L. esculentum*-Genpool eignen sich die Marker für die Genotypisierung von Tomatensorten und -akzessionen.

Insgesamt wurden 41 Marker, die 43 unabhängige Loci detektieren, in die dicht besetzte Kopplungskarte von Tomate eingefügt. Alle 31 aus genomischen Bibliotheken stammenden Marker kartierten ausschließlich in der Nähe der Zentromere verschiedener Chromosomen. Eine solche Häufung von genomischen Mikrosatelliten in Zentromerbereichen wurde bisher in keiner anderen Pflanzenart beschrieben. Im Gegensatz dazu waren nur zwei der auf ESTs basierenden Markern mit einem Zentromer gekoppelt. Die anderen acht waren zufällig im Euchromatin verteilt.

Die Tatsache, daß die in dieser Arbeit isolierten Tomatenmikrosatelliten vornehmlich nahe dem Zentromer liegen, macht sie für die genetische Kartierung weniger wertvoll, bietet aber eine Möglichkeit zur molekularen Charakterisierung zentromerischer Regionen einzelner Tomatenchromosomen. Durch die Isolierung von YAC-Klonen, die homolog zu den

zentromerassoziierten Mikrosatelliten sind, konnte eine Reihe neuer repetitiver DNA-Sequenzen aus dem Tomatengenom charakterisiert werden.

6. References

Akkaya, M.S., Shoemaker, R.C., Specht, J.E., Bhagwat, A.A., Cregan, P.B. (1995). Integration of simple sequence repeat DNA markers into a soybean linkage map. Crop Sci. **35**:1439-1445.

Alfenito, M.R., Birchler, J.A. (1993). Molecular characterization of a maize B chromosome centric sequence. Genetics **135**:589-597.

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

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

Aragón-Alcaide, L., Miller, T., Schwarzacher, T., Reader, S. Graham, M. (1996). A cereal centromeric sequence. Chromosoma **105**:261-268.

Arens, P., Odinot, P., van Heusden, A.W., Lindhout, P., Vosman, B. (1995). GATA- and GACA-repeats are not evenly distributed throughout the tomato genome. Genome **38**:84-90.

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

Barton, D.W. (1950). Pachytene morphology of the tomato chromosome complement. Am.J.Bot. **37**:639-643.

Baum, M., Ngan, V.K., Clarke, L. (1994). The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. Mol. Cell Biol. **5**:747-761.

Baum, T.J., Gresshoff, P.M., Lewis, S.A., Dean, R.A. (1992). DNA amplification fingerprinting (DAF) of isolates of four common Meloidogyne species, and their host races. Phytopathology **82**:1095.

Becker, J., Heun, M. (1995). Barley microsatellites: allelic variation and mapping. Plant Mol.Biol. **27**:835-845.

Beckmann, J.S., Soller, M. (1990). Toward a unified approach to genetic mapping of eukaryotes based on sequence tagged microsatellite sites. *Bio/Technology* **8**:930-932.

Bedbrook, J., Jones, J., O'Dell, M., Thompson, R.D., Flavel, R.B. (1980). A molecular description of telomeric heterochromatin in *Secale* species. Cell **19**:545-560.

Bell, C.J., Ecker, J.R. (1994). Assignment of 30 microsatellite loci to the linkage map of *Arabidopsis*. Genomics **19**:137-144.

Bernatzky, R., Tanksley, S.D. (1986). Towards a saturated linkage map in tomato based on isozyme and random cDNA sequences. Genetics **112**:887-898.

Bolivar, F., Rodriguez, R.L., Greene, P.J., Betlach, M.C., Heynecker, H.L., Boyer, H.W., Crosa, J.H., Falkov, S. (1977). Construction and characterization of new cloning vehicles. II. A multipurpose cloning system. Gene **2**:95-113.

Bonierbale M.W., Plaisted R.L., Tanksley S.D. (1988). RFLP maps based on a common set of clones reveal modes of chromosomal evolution in potato and tomato. Genetics **120**:1095-1103.

Botstein, D., White, R.L., Skolnick, M., Davis, R. (1980). Construction of a genetic linkage map in man using restriction fragment polymorphisms. Am.J.Hum.Genet. **32**:314-331.

Brandes, A., Thompson, H., Dean, C., Heslop-Harrison, J.S. (1997). Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. Chromosome Res. **5**:238-246.

Bredemeijer, G.M.M., Arens, P., Wouters, D., Visser, D., Vosman, B. (1998). The use of semi-automated fluorescent microsatellite analysis for tomato cultivar identification. Theor.Appl.Genet. **97**:584-590.

Broun, P., Ganal, M.W., Tanksley, S.D. (1992). Telomeric arrays display high levels of heritable polymorphism among closely related plant varieties. *Proc.Natl.Acad.Sci.USA* **89**:1354-1357.

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

Brubaker, C.L., Paterson, A.H., Wendel, J.F. (1999). Comparative genetic mapping of allotetraploid cotton and its diploid progenitors. Genome **42**:184-203.

Burke, D.T., Carle, G.F., Olson, M.V. (1987). Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science **236**:806-812.

Burr, B., Burr, F.A., Thompson, K.H., Albertsen, M.C., Stuber, C.W. (1988). Gene mapping with recombinant inbreds in maize. Genetics **188**:519-526.

Carle, G.F., Olson, M.V. (1985). An electrophoretic karyotype for yeast. *Proc.Natl.Acad.Sci.USA* **82**:3756-3760.

Chu, G., Vollrath, D., Davis, R.W. (1986). Separation of large DNA molecules by contour-clamped homogeneous electric fields. Science **234**:1582-1585.

Cooper, D.N., Smith, B.A., Cooke, H.J., Niemann, S., Schmidtke, J. (1985). An estimate of unique DNA sequence heterozygosity in the human genome. Hum.Genet. **69**:201-205.

Copenhaver, G.P., Preuss, D. (1999). Centromeres in the genomic era: unraveling paradoxes. Current Opinion in Plant Biology **2**:104-108.

Cregan, P.B., Jarvik, T., Bush, A.L., Shoemaker, R.C., Lark, K.G., Kahler, A.L., Kaya, N., VanToai, T.T., Lohnes, D.G., Chung, J., Specht, J.E. (1999). An integrated genetic linkage map of the soybean genome. Crop Sci. **39**:1464-1490.

Dib, C., Faure, S., Fizames, C., Samson, D., Drouot, N., Vignal, A., Millasseau, P., Marc, S., Hazan, J., Seboun, E., Lathrop, M., Gyapay, G., Morisette, J., Weissenbach, J. (1996). A comprehensive genetic map of the human genome based on 5,264 microsatellites. Nature **380**:152-154.

Dietrich, W.F., Miller, J., Steen, R., Merchant, M.A., Damron-Boles, D., Husain, Z., Dredge, R., Daly, M.J., Ingalls, K.A., O'Connor, T.J., Evans, C.A., DcAngelis, M.M., Levinson, D.M., Kruglyak, L., Goodman, N., Copeland, N.G., Jenkins, N.A., Hawkins, T.L., Stein, L., Page, D.C., Lander, E.S. (1996). A comprehensive genetic map of the mouse genome. Nature **380**:149-152.

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

Edwards, J.H. (1994). Comparative genome mapping in mammals. Curr.Opinion in Genetics and Development **4**:861-867.

Feinberg, A.P. and Vogelstein, B. (1983). A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal.Biochem. **137**:266-267.

Flavel, R. (1980). The molecular characterization and organization of plant chromosomal DNA sequences. Ann.Rev.Plant Physiol. **31**:569-596.

Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T., Jones, G.H. (2000). Integrated cytogenetic map of chromosome arm 4S of *A. thaliana*: Structural organization of heterochromatic knob and centromere region. Cell **100**:367-376.

Frary, A., Presting, G.G., Tanksley, S.D. (1996). Molecular mapping of the centromeres of tomato chromosomes 7 and 9. Mol.Gen.Genet. **250**:295-304.

Galbraith, D.W., Harkins, K.R., Maddox, J.M., Ayres, N.M., Sharma, D.P., Firoozabady, E. (1983). Rapid flow cytometric analysis of the cell cycle in intact plant tissue. Science **220**:1049-1051.

Ganal, M.W., Broun, P., Tanksley, S.D. (1992). Genetic mapping of tandemly repeated telomeric DNA sequences in tomato (*Lycopersicon esculentum*). Genomics **14**:444-448.

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

Grandillo, S., Tanksley, S.D. (1996). Genetic analysis of RFLPs, GATA microsatellites and RAPDs in a cross between *L.esculentum* and *L.pimpinellifolium.* Theor.Appl.Genet. **92**:957-965.

Haaf, T., Warburton, P.E., Willard, H.F. (1992). Integration of human α-satellite DNA into simian chromosomes. Centromere protein binding and disruption of normal chromosome segregation. Cell **70**:681-696.

Haanstra, J.P.W., Wye, C., Verbakel, H., Meijer-Dekens, F., van den Berg, P., Odinot, P., van Heusden, A.W., Tanksley, S., Lindhout, P., Peleman, J. (1999). An integrated high-density RFLP-AFLP map of tomato based on two *Lycopersicon esculentum* x *L. pennellii* F_2 populations. Theor.Appl.Genet. **99**:254-271.

Harrington, J.J., Bokkelen, G.V., Mays, R.W., Gustashaw, K., Willard, H.F. (1997). Formation of de novo centromeres and construction of first-generation human artificial chromosomes. Nature Genet. **15**:345-355.

Jiang, J.S., Nasuda, F., Dong, C.W., Scherrer, S.-S., Wing, R.A., Gill, B.S., Ward, D.C. (1996). A conserved repetitive DNA element located in the centromeres of cereal chromosomes. *Proc.Natl.Acad.Sci.USA* **93**:14210-14213.

Kaszás, E., Birchler, J.A. (1996). Misdivision analysis of centromere structure in maize. EMBO J. **15**:5246-5255.

Kosambi, D.D. (1944). The estimation of map distance from recombination values. *Ann.Eugen.* **12**:172-175.

Kowalski, S.P., Lan., T.-H., Feldmann, K.A., Paterson, A.H. (1994). Comparative mapping of *Arabidopsis thaliana* and *Brassica oleracea* chromosomes revealed islands of conserved organization. Genetics **138**:499-510.

Kresovich, S., Szewc-McFadden, A.K., Bliek, S.M., McFerson, J.R. (1995). Abundance and characterization of simple-sequence repeats (SSRs) isolated from a size-fractionated genomic library of *Brassica napus* L. (rapeseed). Theor.Appl.Genet. **91**:206-211.

Kruglyak, I. (1997). The use of a genetic map of biallelic markers in linkage studies. Nature Genet. **17**:21-24.

Kwok, P.-Y., Deng, Q., Zakeri, H., Taylor, S.L., Nickerson, D.A. (1996). Increasing the information content of STS-based genome maps: Identifying polymorphisms in mapped STSs. Genomics **31**:123-126.

Lagercrantz,J., Ellegren, H., Andersson, l.. (1993). The abundance of various polymorphic microsatellite motifs differs between plants and vertebrates. Nucl.Acids Res. 2**1**:1111-1115.

Lander, E.S., Green, P., Abrahamson, J., Barlow, A., Daly, M.J., Lincoln, S.E., Newburg, L. (1987). MAPMAKER: An interactive computer package for constructing primary genetic maps of experimental and natural populations. Genomics **1**:174-181.

Lapitan, N.L.V. (1992). Organization and evolution of higher plant nuclear genomes. Genome **35**:171-181.

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

Lapitan, N.L.V., Ganal, M.W., Tanksley, S.D. (1991). Organization of the 5S ribosomal RNA genes in the genome of tomato. Genome **34**:509-514.

Laterrot, H. (1987). Near isogenic tomato line in Moneymaker type with different genes for desease resistances. Tomato Genet.Coop.Rep. **37**:91.

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

Lin, X. et al. (1999). Sequence and analysis of chromosome 2 and 4 of the plant *Arabidopsis thaliana*. Nature **402**:761-777.

Ling, H.-Q., Koch, G., Bäumlein, H., Ganal, M.W. (1999). Map-based cloning of *chloronerva*, a gene involved in iron uptake of higher plants encoding nicotianamine synthase. *Proc.Natl.Acad.Sci.USA* **96**:7098-7103.

Litt, M., Luty, J.A. (1989). A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am.J.Hum.Genet. **44**:397-401.

Liu, Z.-W., Biyashev, R.M., Saghai Maroof, M.A. (1996). Development of simple sequence repeat DNA markers and their integration into a barley linkage map. Theor.Appl.Genet. **93**:869-876.

Livingstone, K.D., Lackney, V.K., Blauth, J.R., van Wijk, R., Jahn, M.K. (1999). Genome mapping in *Capsicum* and the evolution of genome structure in the *Solanaceae*. Genetics **152**:1183-1202.

Maluszynska, J., Heslop-Harrison, J.S. (1991). Localization of tandemly repeated DNA sequence in *Arabidopsis thaliana.* Plant J. **1**:159-166.

Martin, G.B., Ganal, M.W., and Tanksley, S.D. (1992). Construction of a yeast artificial chromosome library of tomato and identification of cloned segments linked to two disease-resistance loci. Mol.Gen.Genet. **233**:25-32.

Martinez-Zapater, J.M., Estelle, M.A., Somerville, C.C. (1986). A highly repeated DNA sequence in *Arabidopsis thaliana*. Mol.Gen.Genet. **204**:417-423.

Maughan, P.J., Saghai Maroof, M.A., Buss, G.R. (1995). Microsatellite and amplified sequence length polymorphisms in cultivated and wild soybean. Genome **38**:715-723.

McCouch, S.R., Chen, X., Panaud, O., Temnyukh, S., Xu, Y., Cho, Y.G., Huang, N., Ishii, T., Blair, M. (1997). Microsatellite marker development, mapping and application in rice genetics and breeding. Plant Mol.Biol. **35**:89-99.

Michelmore, R.W., Paran, I., Kesseli, R.V. (1991). Identification of markers linked to desease-resistance genes by balked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc.Natl.Acad.Sci.USA* **88**:9828-9832.

Milbourne, D., Meyer, R., Bradshaw, J.E., Baird, E., Bonar, N., Provan, J., Powell, W., Waugh, R. (1997). Comparison of PCR-based markr systems for the analysis of genetic relationships in cultivated potato. Mol.Breeding **3**:127-136.

Milbourne, D., Meyer, R.C., Collins, A.J., Ramsay, I..D., Gebhardt, C., Waugh, R. (1998). Isolation, characterization and mapping of simple sequence repeat loci in potato. Mol.Gen.Genet. **259**:233-245.

Miller, J.C., Tanksley, S.D. (1990). RFLP analysis of phylogenetic relationships and genetic variation in the genus *Lycopersicon*. Theor.Appl.Genet. **80**:437-448.

Morgante, M., Olivieri, A.M. (1993). PCR-amplified microsatellites as markers in plant genetics. Plant J. **3**:175-182.

Nei, M., Li, W.H. (1979). Mathematical model for studying genetic variation in terms of restriction endonucleases. *Proc.Natl.Acad.Sci.USA* **76**:5269-5273.

Olson, M.V., Hood, L., Cantor, C., Botstein, D. (1989). A common language for physical mapping of the human genome. Science **245**:1434-1435.

Paran, I., Michelmore, R.W. (1993). Development of reliable PCR-based markers linked to downy mildew resistance genes in lettuce. Theor.Appl.Genet. **85**:985-993.

Peacock, W.J., Dennis, E.S., Rhoades, M.M., Pryor, A.J. (1981). Highly repeated DNA sequence limited to knob heterochromatin in maize. *Proc.Natl.Acad.Sci.USA* **78**:4490-4494.

Pedersen, C., Linde-Laursen, I. (1994). Chromosomal locations of four minor rDNA loci and a marker microsatellite sequence in barley. Chromosome Res. **2**:65-71.

Pelissier, T., Tutols, S., Deragon, J.M., Tourmente, S., Genestier, S., Picard, G. (1995). *Athila*, a new retroelement from *Arabidopsis thaliana*. Plant Mol.Biol. **29**:441-452.

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

Pillen, K., Ganal, M.W., Tanksley S.D. (1996). Construction of a high resolution genetic map and YAC-contigs in the tomato *Tm-2a* region. Theor.Appl.Genet. **93**:228-233.

Podani, J. (1990). SYN-TAX III-pc supplement3: Macintosh version. Abstr.Bot. **14**:23-29.

Powell, W., Morgante, M., Andre, C., Hanafey, M., Vogel, J., Tingey, S., Rafalski, A. (1996). The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. Mol.Breeding **2**:225-238.

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

Presting, G.G., Malysheva, L., Fuchs, J., Schubert, I. (1998). A *Ty3/gypsy* retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. **16**:721-728.

Richards, E.J., Goodman, H.M., Ausubel, F.M. (1991). The centromere region of *Arabidopsis thaliana* chromosome contains telomere-similar sequences. Nucl.Acids Res. **19**:3351-3357.

Rick, C.M. (1975). The tomato. *In: King RC (ed) Handbook of genetics, vol.2. Plenum Press,* New York, pp.247-280.

Rick, C.M., Khush, G.S. (1969). Cytogenetic exploration of the tomato genome. *In: Bogarat R (ed) Genetics Lectures, Oregon University press, Eugene.* **1**:45-69.

Röder, M.S., Plaschke, J., König, S.U., Börner, A., Sorrells, M.E., Tanksley, S.D., and Ganal, M.W. (1995). Abundance, variability and chromosomal location of microsatellites in wheat. Mol.Gen.Genet. **246**:327-333.

Röder, M.S., Korzun, V., Gill, B.S., Ganal, M.W. (1998a). The physical mapping of microsatellite markers in wheat. Genome **41**:278-283.

Röder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.-H., Leroy, P., Ganal, M.W. (1998b). A microsatellite map of wheat. Genetics **149(4)**:2007-2023.

Round, E.K., Flowers, S.K., Richards, E.J. (1997). *Arabidopsis thaliana* centromere regions: genetic map positions and repetitive DNA structure. Genome Res. **7**:1045-1053.

Saal, B., Wricke, G. (1999). Development of simple sequence repeat markers in rye (*Secale cereale* L.). Genome **42**:964-972.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). *Molecular Cloning: A Laboratory Manual.* Cold Spring Harbour, NY: Cold Spring Harbour Laboratory Press.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc.Natl.Acad.Sci.USA* **74**:5463-5467.

Schmidt, R., West, J., Love, K., Lenehan, Z., Lister, C., Thompson, H., Bouchez, D., Dean, C. (1995). Physical map and organization of *Arabidopsis thaliana* chromosome 4. Science **270**:480-483.

Schmidt, T., Heslop-Harrison, J.S. (1996). The physical and genomic organization of microsatellites in sugar beet. *Proc.Natl.Acad.Sci.USA* **93**:8761-8765.

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

Smith, J.G., Caddle, M.S., Bulboaca, G.H., Wohlgemuth, J.G., Baum, M., Clarke, L. (1995). Replication of centromere II of *Schizosaccharomyces pombe*. Mol. Cell Biol. **15**:5165-5172.

Smulders, M.J.M., Bredemeijer, G., Rus-Kortekaas, W., Arens, P., Vosman, B. (1997). Use of short microsatellites from database sequences to generate polymorphisms among *Lycopersicon esculentum* cultivars and accessions of other *Lycopersicon* species. Theor.Appl.Genet. **97**:264-272.

Stubbe, H. (1971). Weitere evolutionsgenetische Untersuchungen in der Gattung *Licopersicon*. Biol. Zentralbl. **90**:545-559.

Su, P.-Y., Brown, T.A. (1997). *T3/gypsy*-like retrotransposon sequences in tomato. Plasmid **38**:148-157.

Sun, X., Wahlstrom, J., Karpen, G. (1997). Molecular structure of a functional *Drosophila* centromere. Cell **91**:1007-1019.

Tanksley, S.D., Ganal, M.W., Prince, J.P., de Vicente, M.C., Bonierbale, M.W., Broun, P., Fulton, T.M., Giovannoni, J.J., Grandillo, S., Martin, G.B., Messequer, R., Miller, J.C., Miller, L., Paterson, A.H., Pineda, O., Röder, M.S., Wing, R.A., Wu, W., Young, N.D. (1992). High density molecular linkage map of the tomato and potato genomes. Genetics **132**:1141-1160.

Tanksley, S.D., Miller, J.C., Paterson, A.H., Bernatzky, R. (1988). Molecular mapping of plant chromosomes. *In:Gustafson J., Appels, R., eds. Chromosome structure and function. Plenum Press,* New York, pp.157-172.

Taramino, G., Tarchini, R., Ferrario, S., Lee, M., Pe', M.E. (1997). Characterization and mapping of simple sequence repeats (SSRs) in *Sorghum bicolor*. Theor.Appl.Genet. **95**:66-72.

Taramino, G., Tingey, S. (1996). Simple sequence repeats for germplasm analysis and mapping in maize. Genome **39**:277-287.

Tautz, D. (1989). Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucl.Acids Res. **17**:6463-6471.

Tautz, D., Renz, M. (1984). Simple sequences are ubiquitous repetitive components of eukaryotic genomes. Nucl.Acids Res. **12**:4127-4138.

Tautz, D., Trick, M., Dover, G.A. (1986). Criptic simplicity in DNA is a major source of genetic variation. Nature **322**:652-656.

Thompson, H., Schmidt, R., Brandes, A., Heslop-Harrison, J.S., Dean, C. (1996a). A novel repetitive sequence associated with the centromeric regions of *Arabidopsis thaliana* chromosomes. Mol.Gen.Genet. **253**:247-252.

Thompson, H.L., Schmidt, R., Dean, C. (1996b). Identification and distribution of seven classes of middle repetitive DNA in the *Arabidopsis thaliana* genome. Nucl.Acids Res. **24**:3017-3022.

Vallejos, C.E., Tanksley, S.D., Bernatzky, R. (1986). Localization in the tomato genome of DNA restriction fragments containing sequences homologous to the ribosomal RNA 45S, the major chlorophyll a-b binding polypeptide and the ribulose bisphosphate carboxylase genes. Genetics **112**:93-106.

Van der Biezen, E.A., Overduin, B., Nijkamp, H.J.J., Hille, J. (1994). Integrated genetic map of tomato chromosome 3. Tomato Genet.Coop.Rep. **44**:8-10.

Van Deynze A.E., Sorrels, M.E., Park, W.D., Ayres, N.M., Fu, H., Cartinhour, S.W., Paul, E., McCouch, S.R. (1998). Anchor probes for comparative mapping of grass genera. Theor.Appl.Genet. **97**:356-369.

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

Vos, P., Hogers, R., Bleeker, M., Reijans, M., van de Lee, T., Hornes, M., Frijters, A., Pot, J., Peleman, J., Kuiper, M., Zabeau, M. (1995). AFLP: a new technique for DNA fingerprinting. Nucl.Acids Res. **23**:4407-4414.

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

Vosman, B., Arens, P., Rus-Kortekaas, W., Smulders, M.J.M. (1992). Identification of highly polymorphic DNA regions in tomato. Theor.Appl.Genet. **85**:239-244.

Wang, D.G., Fan, J.-B., Siao, C.-J., Berno, A., Young, P., Sapolsky, R., Ghandour, G., Perkins, N., Winchester, E., Spencer, J., Kruglyak, L., Stein, L., Hsie, L., Topaloglou, T., Hubbell, E., Robinson, E., Mittmann, M., Morris, M.S., Shen, N., Kilbum, D., Rioux, J., Nusbaum, C., Rozen, S., Hudson, T.J., Lipshutz, R., Chee, M., Lander, E.S. (1998). Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science **280**:1077-1082.

Wang, Z., Weber, J.L., Zhong, G., Tanksley, S.D. (1994). Survey of plant short tandem DNA repeats. Theor.Appl.Genet. **88**:1-6.

Weber, J.L. (1990). Informativeness of human $(dC\text{-}dA)_n\text{-}(dG\text{-}dT)_n$ polymorphisms. Genomics **7**:524-530.

Weber, J.L., May, P.E. (1989). Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am.J.Hum.Genet. **44**:388-396.

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

Willard, H.F. (1990). Centromeres of mammalian chromosomes. Trends Genet. **6**:410-416.

Williams, J.G.K., Kubelik, A.R., Livak, K.J., Rafalski, J.A., Tingey, S.V. (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl.Acids Res. **18**:6531-6535.

Wu, K.-S., Tanksley, S.D. (1993). Abundance, polymorphism and genetic mapping of microsatellite in rice. Mol.Gen.Genet. **241**:225-235.

Zamir, D., Tanksley, S.D. (1988). Tomato genome is comprised largely of fast-evolving, low copy-number sequences. Mol.Gen.Genet. **213**:254-261.

Isolated tomato centromeric DNA sequences (TCSs)

26N6
GGAGTGTCACGTACTGACACAAGAGGAATAATGAATATGAGGGATCGGAGTGTCACGTA
CCGACATAAGAGGAATAATGAATATGAAGG

33E11
GGGTGTCACGTACTGACACAAGAGGAATAATGAATATGAGGGATCGGAGTGTCACGTAC
CGACACAAGAGGAATAATGAATATGAGG

54M24
AGAGTGTCACGTTCCTACACGTAGTATTAAGGTATCGGGTGTTACGTTCCGGCACGTAGA
ATTAGGGGATCGGAGTGTCAAGTTCCGACACTTACTATTAAGGGATCGGGTGTCACGTTC
CGACACGTAGCATTAAGG

52M17
TCTTAATACTACGTGTCGGAACGTGACACTCCGATCCCCTAATTCTACGTTTTGGAACGTG
ACAGCCGATCCCATAATACTACGTGTTGGAACGTGACAGCC

42G2
GGGTGTCACGAACCGACACGTAGATTTAGGGGATCGGGTGTCACGAACCGACACATAGA
TTTAGGG

41L8
GGGTGTCACGAACCGACACGTAGATTTAGGGGATCGGGTGTCACGAACCGACACATAGA
TTTAGGG

56B12
CGTGTGTGTCACGTTCCGACACCAGGATAGAATATATGGATCGGGTGCCACGTTCCGGTA
CCAGGATAGAATATGGATCGGGTGTCACGTTCCGACACCAGGATAGAATATATGGATCG
GGTGCCACGTTCCGGTACCAGGATAGTATATGAGGATCGGAGTGTCACGTTCCGACACCA
GGATAGAATATGGATCGGGTGCCACGTTCCGGTACCAGGATAGAATGAGGATCGGAGTG
TCACGTTTCGACACCAGGCTAGTATATTGAGGAGCGGAGTGTCACGTACCGACACGAGG
GGAATAAAGATAATGAATCTTGAAATGTTAATATATCCAATCTAATGAACTAAATTCCCA
AATGAGTATGATGAGGAGGCGTAAGTCCTCATTGATGTGCTTGGTGTTGTAACCAAGGGT
TATGGTAACTGTAAATGCTGCATGCTAAGGATATTAGTTGATTTTATGATATTGCTGAAT
ACATACTATTTTCTATTTTGAGTTGGCCGATGATATCTACTCAGTATCCGTGTTTCGTACT
GACCCYTACTTTTATTGTTTTCTTCTTGTTTATTTGTGGAGTGCAACAAACGCGCCATCGT
CTTCAACTCAACAGTAATTCAAGCCAGTCTTACTACATCAGAAATTCAGGGTGAGCTAAT
GCTTCTAGCTTGGACTG

42P20
CAGTCCAAGCTAGCAAGCATTAGCTCACCCTGATATCCGAGAGTAATGAAGACTGGTTAG
ATTTACTGTTGAGTCGAAGATCGCACGGCACGTTTGCTGTACTCCACAACATAAATAAGA
AGAACACATAAAAGTAGGGGTCAGTACAAAACACGGGTACTGCAGAAGATATCATCGGC
CAACTCAAAATAGA

33A14
AAATTACTAAATTAAGAGTATTCTAAAAAGCTAAAACAAGTAAAAGCTAGTCCATGCCA
GADCTTCAAGGCATCAAGACATGAAGAGGTAGACCCAGTCCAAGCTACAAGAATTAGCT
CACCCTGAAATCCGTTGTAATTGAAGACTGGCTAGAATTACGGTTGAGTTGAAGACGGCG
GTACGTTTGCTGCACTCCACAAATAACAAAGAAGAAAATATAAAAGTAGGGGGTCAGTA
CAAACACAAGTACTGAGTAGGTATCATCGGCCAACTCCAAATAGAAAACAATATATATC
AAGTAATACCATAAATCAACTACAATACTCAACATGTAGCAACAACAAGCACTATAATC
ATTACCAAGTACCGCCAAGTTCAAACATGAGGACTCAAGCCTCAATACCATACTCATTTT
GGAATTAGGTACATTAGATTGAGTATATTAAC

22N14
AAATTACTAAATTAAGAGTATTCTAAAAAGCTAAAACAAGTAAAAGCTAGTCCATGCCA
GADCTTCAAGGCATCAAGACATGAAGAGGTAGACCCAGTCCAAGCTACAAGADTTAGCT
CACCCTGAAATCCGTTGTAATTGAAGACTGGCTAGAATTACGGTTGAGTTGAAGACGGCG
GTACGTTTGCTGCACTCCACAAATAACAAAGAAGAAAATATAAAAGTAGGGGGTCAGTA
CAAACACAAGTACTGAGTAGGTATCATCGGCCAACTCCAAATAGAAAACAATATATATC
AAGTAATACCATAAATCAACTACAATACTCAACATGTAGCAACAACAAGCACTATAATC
ATTACCAAGTACCGCCAAGTTCAAACATGAGGACTCAAGCCTCAATACCATACTCATTTT
GGAATTAGGTACATTAGATTGAGTATATTAAC

15J19
GGGTGCACGTTCCGGTACCAGGATAGTATATGAGGATCGGAGTGTCACGTTCCGACACCA
GGATAGTATATGGATC

12N12
TCATATATTATCCTGGTACCGGAACGTGGCACCCGATCCATATTCTATCCTGGTGTCGAA
ACGTGACACTCCAATCCTCATATGCTATCCTGGTACCGGAACGTGGCACCT

11M6
AGTCCAAGCTAGAAGCATTAGCTCACCTTGAATTTCCGATGTAGTAAGACTGGCTTGAAT
TACTGTTGAGTCGAAGATGACGGCACGTTTGCTGCACTCCACAAATAAACAAGAAGAAA
ACAATAAAAGTAGGGGTCAGTACAAAACATGGGTACTGAGTAGATATCATCGGCCAACT
CAAAATAAAAAACAGTATATATTATGCAATATCATAAAATCAACTAATATCCTTAGCATG
CAGCATTTACAGGTTACCATAACCCTTTGGTTACAACACCAACCACAAATTBGNGGACTC
ACGCCTCCTCATCATACTCATTTGGGAATTCAGTTCATTAGATTGAGTATATTAACATTTT
TCAAGATTCGTTATTTTTATTCCCCTCGTGTCGGTACGTGACACTCCGCTCCTCAATATAC
TATCCTAGTGTCGAAACGTGACACTCC

12F15
ACCTGCTAAAGTGTCCAACACCGCTTTATTATTAACATCTCCTTGGGGTGCTCCGGTTTTG
TTGGTAAAGAAGAAGGATGGGAGTTTTCGGATGTGAATAGACTACAGACAACTGAATAA
AGTAACTATTAAGAACAAGTACCCTCTTCCCTACATTGATGACTTGTTCGACCAGTTACA
AGGTGCTTGTGTCTTCTCTAAGATTGATTTGCGATCCGGTTATCATCAATTGAAAATACGG
GCAACGGATGTTCCAAAGACTGCTTTTCGAACGAGGTATGGGCATTACGAATTTGTAGTG
ATGTCTTTTGGTCTAACGAATGCCCCTGCTGCGTTCATGAGCTTGATGAACGGGATTTTTA
AACCATATCTGGACCTCTTCGTGATCGTATTTATTGACGATATACTAGTATACTCAAAGA
GCAAGAAGGAACATGAAGAGCATTTGAGAATGGTATTGGAAATGTTGAGGGAGAAAAAG
CTTTATGCCAAGTTCTCTAAGTGTGCGTTTTGGCTAGATGTAGTGTCCTTCTTTGGACACG
TGGTTTCTAAGGATGCAGTGTCCTTCTTGGGGGCACGTGGTTTCTAAGGATGGAGTGATG
GTG

25G22
TGCAGGGAACACATCCAGAAACTCACGAACTACTGAAACCGACTAAATCGTAGGTACTT
GGGTAGTGTCATCCTTGAGATGTGCCAAGAAAGCTAAACACCCTTTACTAATCATTTTCC
TAGCACAAAGAAAGGAGACGATGCGGACCGTATTGGGAGTGTAGTCACCCTCCCACACT
AACGGGTCTGTCCCAGGCTTGGCTAACGTCACCGTTTTAGCATTACAATC

27E20
TAGTACTCAGATTCCAAGAGTTAAAGTATTTTGAAACGAAGACCCTCGATGGACCGTTCT
GCCTATGACGGTCCGTTACACTTGCCGTCGAGGGGAATGAAGAAAGAAGCAGAAGAWAT
TTAACCAAGTATGGGACGACAGAGTCCACGACGGTCCGTTATGACCACGACGGTCAGTC
GCGCGGTCCATCGACCCAGCCGCATTTTGGCAGATTTCCTTAATTTCAATCCTTGTTTAAT
TAGGGTTTTTACTTTTTATAAACAGTTCCGAAAAACCTCATTTTGGGRGGGGWWARCTCT
GCTAGATTAGACATCATATTATTAGAGTTTCTGTATTAGTGTTTGATTTTTGGAGATTCTT
ACAAGTGCTTTTTGGTGATTAATCAAGCAAATTTTCGGACTTTATTCTTTCTCATTGAAGT
AAGTACATGAATTCTTATTTAATATA

33B13
GAGTGTCACGAACCGACACGAGAATTAGGGGAYGGGTGCACGACCTACACGTAGATATA
GGG

51C15
CGTATAGATCTCACACTTAACCCCATATAAATAGTGTCTCTATTGCTTTAATGCAAACACC
ACCGCAGCCAATTCCAAATCGTGGGTCGGATAGTTACGTTCATGCACCTTTAGTTGCATT
GAAGCATAAGCA

53O22
GCATGCCTGCAGGTCGACTCTAGAGGAGCCAGCCACTGGGGCTGTAGCTCGAGGAAGAG
AAGCGGCAAGAGGCCGTGGTAGAGGTCGTGGGAGGACGTCCTCTAGGGGAAGAGGACG
AGCACCTAGCCATCTTATACTAGGGCAGTGACTCTCCACCGACTGAGGAAGTAATAAGAG
AAGGGGAGGATGGGGAAACTGAATAAGTGCAGAATGAGGGATTGCCACCCCAACCTACC
CCAGAGAT

41M21
TGTCTCATCATATTTATACATGCTATCACTATTACACTTGAATTATGTGGAATCAGACCAT
GACATAAACTCACCACGCACTAACATCTATCGTCTTTAATATCTCACCGAGGTGCTAATA
TTTCTGGTATTGCAACTAGTGTCCTCACTTCAAAATAATAACCTCATTTTTCATACTGTGA
TTTTAGTTTGAGTATAAGGATTACTTTTCAACTGTAATGACCTGTTTAGTCGTTTTGAGCA
GCAGATTTTATTTCTGGAAAAACAGGCTGAGATGACGGAANCCACGAGGGACCGTCATG
GGCACGACGGACCGTCGAGGGGGTCTCGTTCCAAAACACTTAGATTTCTGAATTTGGGTA
CTGAAATCGACTCTCTGAACTTCATGATGGAATGGCAGGACGGACCGTCACAGGCGTGAC
GGGCCGTCACA

46O13
GACTCGGTTCCTTGTCTTGGCGCAATTCTCCTGATTTGAGCTTGGGCACTCGGTCGGAACC
CCCTGTCTGCTCATTTACCGCATAGAAATCCTCCTCATAATTACATTCATCATTTGGTGGT
GGTGGTTTTGACAAGTAGTTAACTGCATTTATCTTTTCTGCACCCCCAATGACATGTTTTA
ATACCAACCCAAGCTCAGTTCTCATCTGAGCCATCTCTTCACGAATCTCATCTGTGTCTGG
GTTGTGAGTGGATTGTACTGCGAAGGTATTTYTCYCVATWWCTGAATTCCTAGTAACCTT
TATTTTTTGGGAGATTTTCTCTAACTTTTTGGCAATCTCGGCATAAGGACACTCCCCATAA
GAACCACTCTCTATAGTATCCAATACTGCTTTATTATTATCATCCTGTCCCCGATAGAAGT
ATTCTTTCAGTGACTCATCATCTATGCGGTGATTGGGGATATTTATAAAA

46N17

AATACAGAAGTCAATGTCCCTATCTGGTGGCATACCAGGAAGATCACCTAGACACTTTTT
AAGCATCGAAACATGAAACACTGGATGAATAGAAGCCAAGTTATTTGGTAACTTCAGTTC
ATAGGCAACATTCCCTACTCGCTGTAAGACCCACATTTGTTTATCCCAAATACCTAGTAA
ACTAGCAATGTGAAAGCACCTATGTCTAGGGTCGTAAACACGGCAATTTCAAAGATTTTT
AACCTAGGGTTAGATAAATTTTATATAAAAGGARTRCAATATCAATAAGACTAGGCTAAA
ACTAATTGATAAACAAAWATGCAAAATAAATGAGTAGAAATTAGAGACACATACCTGAG
AATTGGAAAAAAACAAGATGGAAAAGTACTTGGTGATCAAGAAGACACCCACAGCAGGA
ACTCCGACTAGTAGATGATAACTTTTAGGGCTTAGGAGAATTTTGGGGGAACAATGATTG
GTTGATAGAGGGAAATAGGGGGAGTAGTGAAGGAATGGGGGTGAAATGAGGGGTTGGG
GAGTTTAAATAGTAAGATTTTGTAAAAHAAACTCCAGCCGGGTCAGGTCGGGTACACCTC
ATTAATGACGCGACGATTCCTCGACGACGGTCCGTCTCAGTTGTGACGATCCATCGTTGG
TTCCATCGTGTGTGCCCTAGTTTGAWAATWATAGADAGACATACCTGGGCACGATGGAT
TCATGCGACGGTCCGTCACAGGTGCAACGGTCCGTCGCTGTATCCGTCAGTGACTGCTGC
ATAGTATTTTCTGCAGAATTTCCTGGTGATGTGCTTGCAAATTTAAAACCCATTAGTAGA
AAATGCTACC

Acknowledgements

This research has been performed at the Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben. I would like to thank all colleagues of the Institute and in particular from the „Gene and Genome Mapping“ group for the scientifically stimulating environment and friendly atmosphere that has promoted this work.

I am especially grateful to Dr. Martin Ganal, head of the „Gene and Genome Mapping“ group, for the opportunity to gain valuable experience in molecular genetics under his careful supervision, stimulating discussions during the work and critical reading the manuscript.

I am grateful to Prof. Dr. Ulla Bonas and PD Dr. Christiane Gebhardt for reviewing the manuscript.

I thank Dr. Marion Röder, expert in the field of microsatellite marker development and application, for her helpful advices regarding theoretical and practical scientific problems and for reading this manuscript.

I am grateful to Mrs. Susanne König for sequencing hundreds of clones. I would also like to address my thanks to technical assistants Mrs. Rosemarie Czihal, Mrs. Heidi Haugk, Mrs. Doris Kriseleit and Mrs. Katja Wendehake for introducing me to techniques and equipment during various phases of the project.

I would like to say many thanks to Dr. Dagmar Schmidt and to Elena Pestsova for the collaboration during the time and their friendly willingness to help in various situations.

My parents and friends I would also like to thank for sympathy and encouragement.

This work has been supported by the German-Israeli Foundation for Research and Development (GIF) under the grant number G-342-028.12/94 which is gratefully acknowledged.

DECLARATION

Hereby I declare that all the work presented in this manuscript is my own, carried out solely with the help of the literature and aid cited.

Gatersleben, March 2000

Tatyana Areshchenkova

CURRICULUM VITAE

Name:	Tatyana Areshchenkova
Address:	Selkeweg 4a D-06466 Gatersleben Germany
Date of birth:	1 January, 1967
Place of birth:	Kiev, Ukraine
Nationality:	Ukrainian, citizen of Ukraine
Marital state:	single
Languages:	Ukrainian, Russian, English, German

Education and employment

1974-1984	secondary school Nr.137 in Kiev, Ukraine
1984-1989	student at the Kiev Technological Institute of Food Industries, Department of Microbiology. Honours degree of higher education
1989-1991	engineer at the Institute of Microbiology and Virology of the Ukrainian Academy of Sciencies, Department of Industrial Microorganisms Physiology
1991-1995	post-graduate course in biotechnology at the Institute of Microbiology and Virology of the Ukrainian Academy of Sciencies
March 1995 - September 1995	Grant from „Deutsche Forschungsgemeinschaft“ (DFG) for working in the Mutational Genetics group at the Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany
1996 - present	Ph.D. student in the Gene and Genome Mapping group at the Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, Germany

Publications and posters

Areshchenkova T. and Ganal M.W. Simple sequence repeats in tomato are predominantly associated with centromeres. *6. Tagung der AG Moleculare Marker der GPZ,* 14-15 September, 1998, Köln, Germany.

Areshchenkova T. and Ganal M.W. Tomato microsatellite marker development: allelic diversity and genetic map position. *In 2nd International Symposium on Plant Biotechnology,* 4-8 October, 1998, Kiev, Ukraine, Abstract p.14.

Areshchenkova T. and Ganal M.W. (1999). Long tomato microsatellites are predominantly associated with centromeric regions. Genome **42:**536-544.

Areshchenkova T. and Ganal M.W. Frequency, allelic variability and centromeric location of long microsatellites in the tomato genome. *4th Gatersleben Research Conference*, 17-21 June, 1999, Schloss Meisdorf/Harz, Germany.

Areshchenkova T. and Ganal M.W. Microsatellite markers for tomato are primarily centromere-associated. *Molecular Biology of Tomato Conference*, 13-16 July, 1999, York, UK.