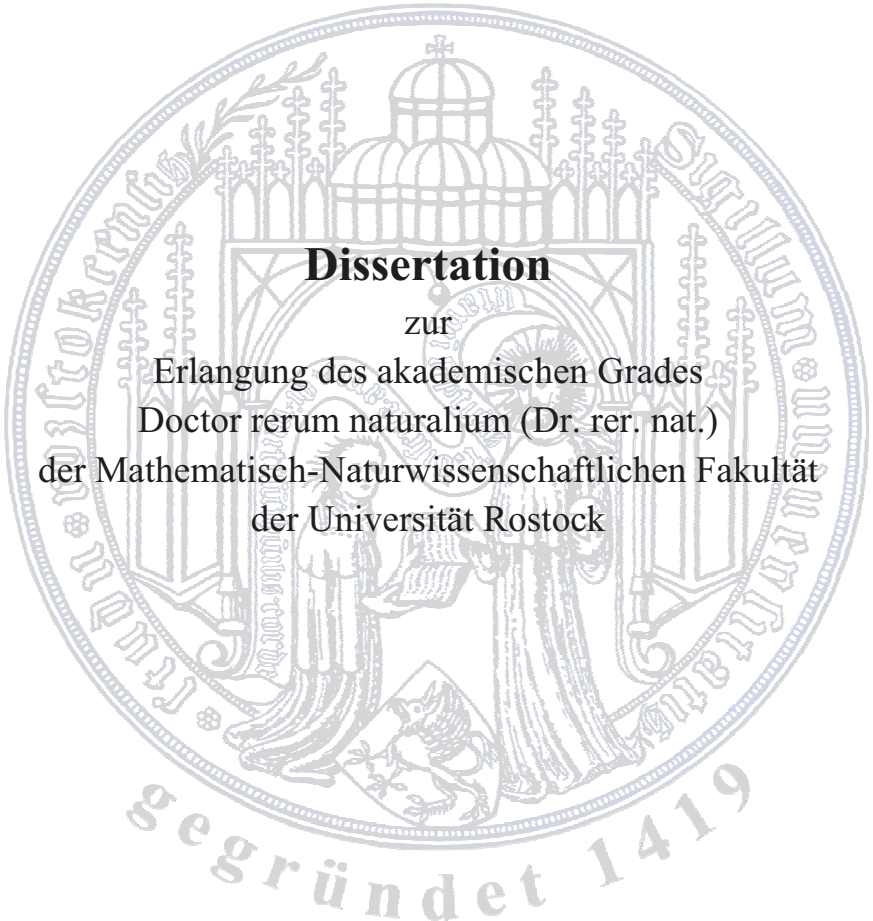
Chromosome walking at the fertility restorer locus *Rf1* in sunflower (*Helianthus annuus* L.)



Dissertation

zur

Erlangung des akademischen Grades
Doctor rerum naturalium (Dr. rer. nat.)
der Mathematisch-Naturwissenschaftlichen Fakultät
der Universität Rostock

vorgelegt von

Sonia Hamrit

geboren am 23. Januar 1965 in Constantine (Algerien)

Rostock, den 3. Juni 2009

urn:nbn:de:gbv:28-diss2009-0167-5

Gutachter:

Prof. Dr. Renate Horn
Institut für Biowissenschaften/Pflanzengenetik, Universität Rostock
email: renate.horn@uni-rostock.de

Prof. Dr. Dr. h.c. Wolfgang Friedt
Lehrstuhl für Pflanzenzüchtung, Universität Gießen
email: wolfgang.friedt@agrar.uni-giessen.de

Die öffentliche Verteidigung fand am 27. July 2009 an der Universität Rostock statt.

À mes parents

Qu'ils trouvent dans ce travail l'accomplissement de leur patience, leur tolérance, et surtout leur confiance.

Table of contents

1. Introduction

1.1 Sunflower characterization

Sunflower (*Helianthus annuus* L.) is one of the few crop species that originated in North America (Harter et al. 2004). The genus belongs to the family of *Asteraceae* and consists of 49 species and 19 subspecies with 12 annual and 37 perennial species (Seiler 1992). Sunflower has probably been first introduced to Europe through Spain in the sixteenth century as garden ornamental flower, and gradually moved eastward and northward in the European continent until it reached Russia in the 18th century where it was readily adapted for increasing oil content (Seiler, 1992). Sunflower (*Helianthus annuus* L.) as a member of the Compositae is characterized by a single terminal inflorescence consisting of 700 to 3,000 individual disc flowers or florets (Reviewed in Smart et al. 1994) (Fig. 1).

Fig. 1 Terminal inflorescence of sunflower.

Sunflower florets are hermaphroditic, as they are composed of both male and female organs. The anthers are present in the upper part of the floret and the ovaries are present in the lower part of the floret (Fig. 2). Sunflower florets exhibit protandry, that is, the male part of the flower (anthers) matures before the female. Within the inflorescence,

the florets develop sequentially in whorls from the periphery to the center at a rate of about one to four whorls a day (Hernández and Green 1993).

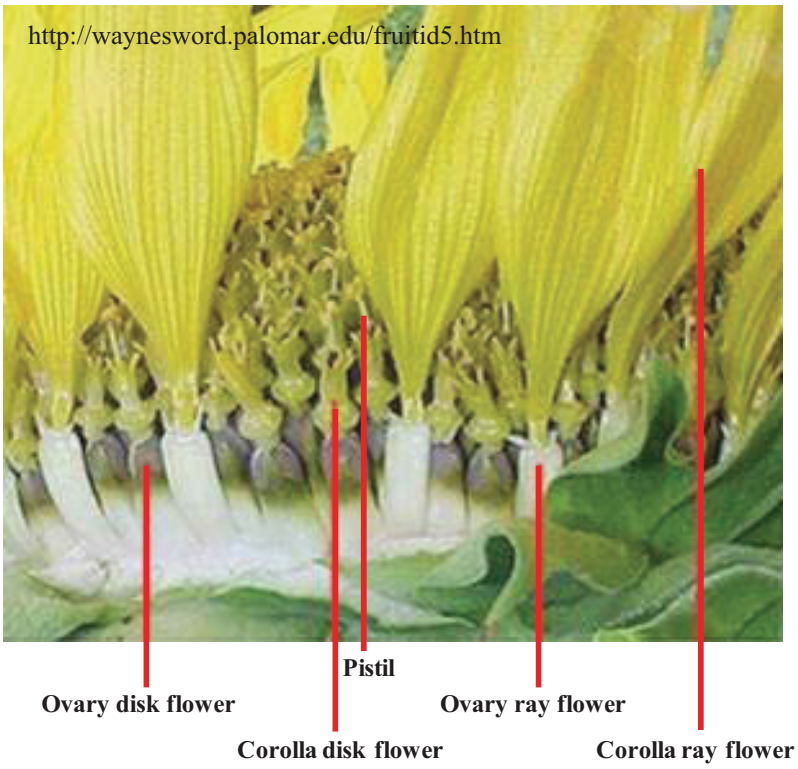


Fig. 2 A capitulum or head of sunflower. A portion of the inflorescence of sunflower showing an outer ring of large strap-shaped (petal-like) ray flowers surrounding a dense mass of small tubular disk flowers. The ovary of each flower is situated below the attachment of the corolla and stamens. The entire head is subtended by green overlapping bracts called phyllaries.

Today, sunflower is one of the most important crops in the world grown for edible oil, after soybean (*Glycine max* L.), rapeseed (*Brassica rapa* L.) and peanut (*Arachis hypogaea* L.) (Putt 1997). Sunflower seed oil contains about 11% of saturated fatty acids (palmitic and stearic acids), a moderate level (20% - 50%) of monounsaturated fatty acids (oleic acid), and a high concentration of polyunsaturated fatty acids (linoleic acid), with traces of linolenic acid (Dorell and Vick 1997; León et al. 2003).

The development of sunflower cultivars with a high level of saturated fatty acids could increase the utility of the oil for specific edible purposes (Kinney 1994; Pérez-Vich et al. 2000). High oleic acid varieties have a high nutritional value combined with a higher

oxidative stability (Pérez-Vich et al. 2002, 2006). High stearic acid content varieties would produce solid or semi solid fats (Pérez-Vich et al. 2002, 2006).

Nowadays, amounts of sunflower seed are consumed in other markets than the oil markets (Putt 1997). Some of them are used as whole, roasted seeds, much like peanuts as well as food for birds and small animals. In addition to the economically aspect, the genus *Helianthus* is an important evolutionary taxon that contains a number of wild species, which have been used as models for the study of the genetics of adaptation and speciation (Rieseberg et al. 1995; Lexer et al. 2003). Wild species are adapted to a wide range of habitats and possess a considerable amount of genetic diversity that may be a rich adequate source of alleles for continued improvement of the cultivated sunflower (Seiler and Rieseberg 1997; Burke et al. 2002).

Plant breeding generally aims at improving agronomically relevant traits, by combining characters present in different parental lines of cultivated species or their wild relatives (Winter and Kahl 1995). Conventional breeding methods reach this aim by screening the phenotype of pooled or individual plants of an F_2-population for the presence of the desirable trait. However, the improvement of sunflower by conventional breeding methods is not only restricted by the time-consuming and costly process of repeated back-crosses, selfing and testing but also by the limited availability of species in the gene pools (Durante et al. 2002). For these reasons, much attention has been focused on the concept of DNA-based markers to directly access any part of the plant genome since this process allows the connection of phenotypic characters with the genomic loci responsible for it. Actually, the development of new tools of molecular biology for genome and expression analyses is one of the main objectives in sunflower.

1.2 Genomic resources in sunflower

Sunflower is a diploid plant with an estimated genome size of about 3,000 Mb and 2n 2x 34 chromosomes (Arumuganathan and Earle 1991). A large number of molecular markers and several linkage maps of varying density and completeness have been developed and produced for cultivated sunflower *Helianthus annuus* L. (Knapp et al. 2001). One of the most important parameter of a genome linkage map is the genetic length measured in centiMorgan (cM), which corresponds to a rate of one recombination in 100 meioses. Extend of such recombination between markers was

taken as a measure of the distance separating the different markers within particular linkage group (Winter and Kahl 1995; Durett et al. 2002).

The genetic lengths of the published sunflower maps varied greatly, from 962 cM constructed from 117 SSRs (Micic et al. 2004) to 2,916 cM comprising 367 AFLPs markers (Al-Chaarani et al. 2004). Unfortunately there is no relationship between these different maps and the actual genetic length of the sunflower map is still unknown, because the vast majority of the markers are not publicly available (Carrera et al. 2002; Hu 2006).

1.2.1 Genetic maps in sunflower

Genetic maps in sunflower constructed by different marker systems have been described by several authors. The most important markers are: restriction fragment length polymorphisms (RFLP), amplified fragment length polymorphisms (AFLP), random amplified polymorphisms DNA (RAPD), and simple sequence repeats (SSR). Three independent restriction fragment length polymorphism (RFLP) maps have been constructed by Berry et al. (1995, 1996, 1997), Gentzbittel et al. (1995, 1999), and Jan et al. (1998). Berry et al. (1995) reported the first RFLP map of cultivated sunflower using an F_2-population derived from a cross between the inbred lines HA89 and ZENB8. In addition, Berry et al. (1996) merged nine F_2 maps and integrated 635 RFLP markers, which were arranged in 17 linkage groups corresponding to the haploid number of chromosomes in this species. This map covers 1,380 cM and had a mean density of 5.9 cM. In 1997, Berry et al. released to the public research institutions, 81 RFLP loci, detected by 81 cDNA probes.

Gentzbittel et al. (1995) described a consensus RFLP linkage map for cultivated sunflower based on 237 non-distorted segregating loci detected by 180 probe-enzyme combinations. This map had a mean marker density of about 7 cM and covered 1,150 cM with 23 linkage groups. Having 6 linkage groups more than the haploid number of chromosomes in sunflower, this map probably represents about 70% of the sunflower genome. The authors concluded that additional RFLP markers were necessary to obtain 17 linkage groups and to cover the regions lacking markers. With this aim, Gentzbittel et al. (1999) integrated seven individual maps and constructed a near-saturated linkage map, based on RFLP markers and including major phenotypic traits. This integrated map is arranged in 17 major linkage groups containing 238 loci and covers 1,534 cM.

Jan et al. (1998) mapped 271 RFLP loci in one F_2-population of a cross between inbred lines RHA271 and HA234 using 232 cDNA probes. This map had 20 linkage groups, covering 1,164 cM of the sunflower genome, and had an average marker-to-marker distance of 4.6 cM.

Rieseberg et al. (1995) employed 197 random amplified polymorphic DNA (RAPD) markers to study the effects of chromosomal structural differences on introgression in backcrossed progenies of the domesticated sunflower, *Helianthus annuus* L. and its karyotypically divergent wild relative, *Helianthus petiolaris*. The results indicated that the barriers to introgression in *Helianthus* appeared to include both chromosomal structural and genic factors.

Burke et al. (2004) described a joint SSR (simple sequence repeat)/RAPD genetic linkage map of the *H. petiolaris* genome and used it, along with an integrated SSR map derived from four independent *H. annuus* L. mapping populations, to examine the evolution of genome structure within *Helianthus*. The results of this work indicated the presence of 27 collinear segments resulting from a minimum of eight translocations and three inversions. Taken together, these rearrangements required a minimum of 20 chromosomal breakages/fusions. The *H. petiolaris* SSR/RAPD map spans 17 linkage groups and 1,592 cM.

Other genetic linkage maps were constructed by means of amplified fragment length polymorphism (AFLP) markers (Peerbolte and Peleman 1996; Kim and Rieseberg 1999; Gedil et al. 2003; Langar et al. 2003).

More recently, Tang et al. (2002) constructed the first genetic linkage map for sunflower on the basis of simple sequence repeat (SSR) markers. The map was constructed with recombinant inbreed lines (RILs) from a cross between confectionery and oil-seed fertility restorer lines (RHA280 × RHA801). The genetic linkage map was about 1,368 cM long, and had a mean distance of 3.1 cM per locus. The SSR markers described here supply a critical mass of DNA markers for constructing genetic maps of sunflower and create the basis for unifying and cross-referencing the multitude of genetic maps developed for wild and cultivated sunflower.

Subsequently, Yu et al. (2003) increased the density of the sunflower map by constructing a new RIL map (PHA × PHB) based on SSR markers. They integrated and cross referenced the Tang et al. (2002) SSR map with the RFLP maps of Berry et al. (1997) and Jan et al. (1998) using the Gedil et al. (2001) RFLP map as bridge. The latter

was constructed from HA370 × HA372 F_2-progenies and RFLP markers from the maps of Berry et al. (1997) and Jan et al. (1998).

The target region amplification polymorphism (TRAP) marker technique was also employed to define sunflower linkage group ends and to expand the published sunflower simple sequence repeat (SSR) linkage map (Hu et al. 2004, 2007).

Several other molecular tools for breeding and genotyping are available (Hongtrakul et al. 1997, 1998a, 1998b; Brahm et al. 2000; Quagliaro et al. 2001; Chen et al. 2006; Kolkman et al. 2007).

1.2.2 Expressed sequence tags in sunflower

For assessment of gene expression, multiple methods such as e.g. large-scale single pass sequencing of cDNA clones to generate expressed sequence tags (ESTs) can be utilized (Adams et al. 1993). This method can lead to genetic mapping of a gene that directly affects a trait or a specific sequence could be targeted for genetic mapping due to its predicted function based on sequence comparison (Cato et al. 2001). It provides a quantitative method to measure specific transcripts within a cDNA library by increasing the depth of sequencing within a library or by broading the diversity of tissues from which the libraries are constructed. By this the number of gene discoveries can be increased (Ronning et al. 2003). Consequently, ESTs generated from cDNA libraries should represent, ideally, all expressed genes in a target organ/tissue, at a specific development stage and/or in a specific environment (Fernández et al. 2003). The EST approach has been particularly useful in taxa whose whole genome sequences remain unavailable or are limited in their genetic resources like the Compositae (Lai et al. 2005). The EST libraries generated up to now have proven to be excellent resources for gene discovery, molecular marker development, analysis of gene expression at the level of the whole genome, and identification of candidate genes for phenotypes of interest (Clarke et al. 2003; Gupta and Rustgi 2004). The most recent advances towards the characterization of sunflower are the generation of a non-negligible number of public or private ESTs (Tamborindeguy et al. 2004; Ben et al 2005). As part of the Compositae Genome Project, a comprehensive annotated EST database has been developed for sunflower and lettuce, one other major Compositae crop (http://cgpdb.ucdavis.edu). For sunflower, ESTs were derived from *H. annuus* L, *H. argophyllus*, *H. ciliaris*, *H. exilis*, *H. paradoxus*, *H. petiolaris* and *H. tuberosus*.

They represent approximately 136,935 unigenes and are deposited in the NCBI database (http://www.ncbi.nlm.nih.gov/genomes/PLANTS/PlantListhn. html # EST).

Lai et al. (2005) described the use of this EST resource to generate and map 243 new genetic markers for sunflower using denaturing high-performance liquid chromatography (DHPLC) for SNP detection to create a transcript map for sunflower. These 243 loci mapped to the 17 linkage groups previously characterized by Tang et al. (2002). Comparison with previously mapped QTLs revealed some cases where ESTs with putatively related functions mapped near QTLs identified in other crosses for salt tolerance and for domestication traits (Lai et al. 2005). EST/QTL associations reported in this study represent an important first step toward identifying genes underlying ecologically and agriculturally important traits in sunflower. Pashley et al. (2006) used the publicly available cultivated sunflower EST database to develop SSR markers for use in the genetic analysis of a rare species, *H. verticillatus*, as well as the more widespread *H. angustifolius*. Noting that these two species represent the most divergent sections within the genus *Helianthus* (Schilling 1997), they showed that EST-derived SSR markers were three times more transferable across species than anonymous SSR markers.

1.2.3 Sunflower BAC libraries and their applications

Large-insert genomic libraries are one of the key components in plant genome research and have become one of the most useful resources in the map-based cloning of important genes. Clones containing large DNA fragments are required to obtain efficiently accurate and high resolution physical maps of eukaryote genomes (Cai et al. 1995). Bacteriophage λ and Cosmid vectors have a cloning capacity of 24 to 45 kb, respectively (Collins and Hohn 1978). Even, though the cloning efficiency of these systems is high (Cai et al. 1995), the relatively small DNA insert size and the instability of cloned DNA in cosmid vectors (Yokobata et al. 1990) limit their utility for chromosome walking. Yeast artificial chromosome (YAC) libraries were initially generated in order to construct large-insert genomic libraries (Burke et al. 1987; Libert et al. 1993). However, YAC libraries face several difficulties during their construction. First, they show low cloning efficiency (Smith et al. 1990). Second, it is more difficult to isolate cloned DNA from yeast cells than from bacterial systems, because yeast cells have a rigid cell wall and insert fragment co-migrate with the endogenous yeast

chromosome (Cai et al. 1995). Third, YAC clones are often chimeric, consisting of DNA fragments from different regions of the target genome (Cai et al. 1995). To obtain a vector system improved for cloning efficiency and stable maintenance of large genomic inserts, bacteriophage P1 (Sternberg 1990), bacterial artificial chromosome (BAC) (Shiyuza et al. 1992), and P1-derived artificial chromosome (PAC) (Ioannou et al. 1994) were developed. Bacterial artificial chromosome libraries have certain advantages over others systems. These include: high cloning efficiency, good range of the insert size (100-150 kb), a low rate of chimerism, high insert stability, easy mini-preparation of the DNA and compatibility with templates for sequence determination (Shizuya et al. 1992). BAC libraries have been developed for many crop species, such as sorghum (Woo et al. 1994), rice (Wang et al. 1995), soybean (Danesh et al. 1998; Tomkins et al. 1999a), sugarcane (Tomkins et al. 1999b), wheat (Lijavetzky et al. 1999), tomato (Budiman et al. 2000), and melon (Luo et al. 2001).

The first BAC library reported in sunflower (Gentzbittel et al. 2002) was constructed from the inbred line HA821 (which is of particular interest for a gene controlling yield and seed filling), using *Hind*III as restriction endonuclease and pBeloBAC11 as vector. The BAC library had four- to five-fold genome coverage with an average insert size of 80 kb. The utility of this BAC library was evaluated by screening for the presence of genes for putative transmembrane receptors sharing epidermal growth factor (EGF) and integrin-like domains.

The second BAC library was constructed from the restorer line RHA325 using the same enzyme and same vector system. This BAC library had an 1.9-fold genome coverage and an average insert size of 60 kb (Özdemir et al. 2004). This BAC library was successfully used for isolating genomic copies of the *sf21* gene, originally identified as a pollen- and anther-specific gene, which later proved to be a gene family with members expressed in all tissues investigated (Lazarescu et al. 2006). This BAC library has also been used in the map-based cloning approach for the restorer gene *Rf1* (Horn et al. 2003; Kusterer et al. 2002, 2004a; Hamrit et al. 2008).

Feng et al. (2006) reported the construction of BAC and BIBAC (binary bacterial artificial chromosome) libraries from the widely used line HA89 using two restriction enzymes (*Bam*HI, *Hind*III) and two vectors (pECBAC1, pCLD04541). These libraries have much larger insert-sizes (140 and 137 kb, respectively) and deeper coverage (8.9-fold) than the existing sunflower BAC libraries. Using these two libraries, Feng et al.

(2006) identified a set of linkage group-specific BAC or BIBAC clones by overgo hybridization. These will be used to characterize a set of sunflower trisomic lines.

Bouzidi et al. (2006) described the design and the construction of a sunflower BAC library suitable not only for Southern hybridization but also for PCR screening. They screened the whole BAC library in less than 80 PCR reactions including positive and negative controls using a set of 25 SSR markers covering about 36 cM in the sunflower SSR map (Tang et al. 2002). The 112 BAC-clones identified were organized into 23 contigs, which roughly corresponded to 3.1 Mb. The contigs were anchored on the map by the same SSR markers.

In the recent years, Clemson University Genomics Institute (CUGI) released a sunflower BAC library for the public (http://www.genome.clemson.edu). This BAC library has been constructed from the cultivar HA383, using pIndigoBac536 vector and *Hind*III as restriction endonuclease. The BAC library has an average insert size of 125 kb and 8.3 genome equivalents. This BAC library has also been used in the map-based cloning approach for the restorer gene *Rf1* (Hamrit et al. 2008).

These genomic resources are valuable tools that can be used to obtain direct access to some genes of interest by map-based cloning or candidate gene approaches for physical mapping or for the development of markers (Horn and Hamrit 2009).

1.3 Map-based cloning strategy

The term map-based cloning, also called positional cloning or recombination mapping, refers to a technique for the isolation of genes controlling traits, for which very little or no information is available on the underlying molecular mechanism (Scheible et al. 2004). The concept behind map-based cloning is to find DNA markers closely linked to a gene of interest and then to progress or to walk along the chromosome (known as chromosome walking) to the gene of interest via overlapping DNA segments from genomic libraries (Tanksley et al. 1995) (Fig.3). This procedure is typically done by first identifying markers in the vicinity of the gene, and then saturating the region around those original molecular markers with other markers. The next step is to screen a large insert genomic library, for example a bacterial artificial chromosome (BAC) library with markers as probes in colony hybridization experiments to isolate BAC-clones that hybridize to the molecular markers. These BAC-clones can be arranged into contigs by fingerprinting. BAC-end sequences would then be used to develop probes for

continuing hybridizations until a closed contig around the gene of interest is formed, which is characterized by overlapping BAC-clones coming from both sides of the gene of interest. The last step of map-based cloning is to identify the candidate gene from large-insert BAC-clone by sequencing and verification of the gene by e.g. performing genetic complementation.

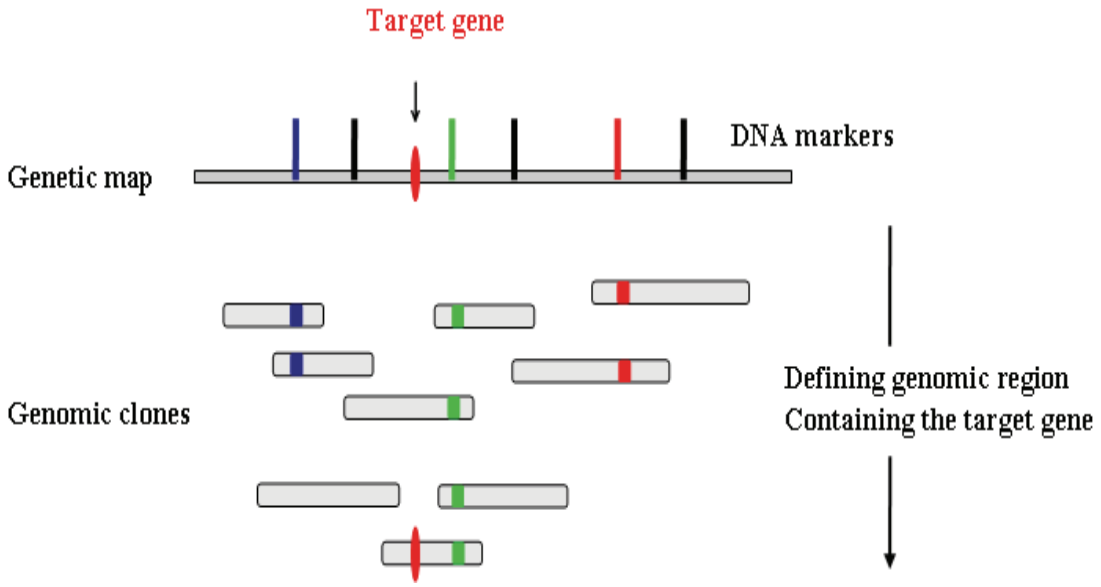


Fig. 3 Schematic representation of the map-based cloning steps. The genetic map represents the order of the DNA markers (coloured lines) and the target gene. The unit of the genetic map is centiMorgan (cM). The BAC-clones represent the level of the physical map.

1.3.1 Molecular marker techniques

Genetic maps have been constructed in many crop plants using several marker systems on a single segregating population. These marker types differ in information content, number of scorable polymorphism per reaction, degree of automation but do not affect the phenotype of the trait of interest because they are only located near or linked to the gene controlling the trait (Collard et al. 2005). Usually, the choice of method depends on the genetic resolution needed as well as on the technology available and the financial constraints. The maps are normally the product of a systematic effort to saturate a genome with polymorphic DNA markers (Winter and Kahl 1995). Among the various molecular markers, RFLPs (Botstein et al. 1980) have been developed in many plant

species including sunflower. RFLPs are co-dominant markers and can identify a unique locus. The polymorphism detected by RFLP markers is very reliable as it involves the recognition by specific restriction enzymes and hybridization with a specific probe (Mohan et al. 1997). However, RFLP analyses are labor intensive, time consuming and their success is inversely related to the size and complexity of the genome (Zhang and Wing 1997). In contrast, the newer approaches based on polymerase chain reaction (PCR) are relatively simple. Amplified fragment length polymorphism (AFLP) is based on PCR amplification of restriction fragments generated by specific restriction enzymes and oligonucleotide adapters (Vos et al. 1995). This method generates a large number of restriction fragments facilitating the detection of polymorphisms. The number of DNA fragments, which are amplified, can be controlled by choosing different numbers of selective nucleotides added to the adapter primer sequences. This approach is useful in saturation mapping and for discrimination between varieties. It is characterized by a high reproducibility and a high frequency of identifiable polymorphism (Lin et al. 1996). However, AFLPs are still expensive to generate because the bands are detected by silver staining, fluorescent dye or radioactivity (Mohan et al. 1997).

Microsatellites or Simple Sequence Repeats (SSRs) are short sequence elements arranged in a simple internal repeat structure (Tautz 1989). SSRs have been detected in the genome of every organism analyzed, so far (Li et al. 2002). The number of microsatellites has been shown to be highly variable within and between species. SSRs constitute the molecular marker type with the highest PIC (polymorphic information content), which is a feature of the number and frequency of alleles detected (Paniego et al. 2002). Nowadays, SSRs have become the markers of choice for population genetic analyses including DNA fingerprinting, genetic mapping, and molecular breeding in crop plants (Powell et al. 1996). Microsatellite markers also come in handy in map-based cloning approaches for comparative analyses between different crosses. In addition, the sequences can be directly used to design overgo probes for hybridization against large insert genomic libraries.

1.3.2 Mapping populations

The choice of the population used for mapping can have important consequences on the efficiency and accessibility of the mapping information (Burr and Burr 1991). The mapping is usually done using F_2-populations, backcrosses (BC), or recombinant inbred lines (RILs).

An F_2-population is developed by selfing F_1-individuals. The F_1-individuals are derived by crossing two parents that differ in the trait of interest. Backcross populations are developed by crossing F_1-individuals with one of the two parents used in the initial cross. F_2 and BC_1-populations are the simplest types of mapping populations developed for self pollinating species. Their main advantages are that they are easy and cheap to construct and require only a short time to be produced (Langar et al. 2003; Collard et al. 2005). However, the major disadvantages of using F_2- or BC_1-populations is the fact that these populations are not eternal and the sources of tissues to isolate DNA or protein will be exhausted at some point in time (Ferreira et al. 2006). Then one would have to begin mapping again in another population. The use of RILs can be a powerful solution to this problem. Recombinant inbred lines are developed by single-seed selections from individual plants of an F_2-population (called also F_2-derived lines). Single-seed descent is repeated for several generations, which results in a series of homozygous lines, each containing a unique combination of chromosomal segments from the original parents. They have several advantages. First, RILs represent an eternal population with unlimited mapping possibilities. This means, once the homozygosity has been obtained, recombinant inbred lines may be propagated indefinitely without further segregation (Burr and Burr 1991). Second, RILs can be repeatedly scored for morphological traits like disease resistance or quantitative traits such as yield, which can be compiled on a developing molecular map. Third, an important factor distinguishes RILs from F_2- or BC_1-populations: RILs undergo multiple rounds of meiosis before homozygosity is reached. As a result, linked markers and genes have a greater probability of recombination (Burr and Burr 1991). However, the time needed for producing RILs represents a major disadvantage, because RILs require between six to eight generations or even more to be regarded homozygous.

1.3.3 Building contigs and chromosome walking

The next step is to screen a large insert genomic library with the markers. The goal of the genomic library screening is to identify and to characterize clones, to which the markers hybridized. The step that follows is termed chromosome walking. This procedure involves creating new probes, usually from the sequences at the end of the identified BAC-clones and screening again the population in order to find another set of markers, which cosegregate with the gene of interest. A contig can be assigned

unambiguously to a specific genetic location if the contig is associated to one genetic location and the probe does not hit BACs in another contig (Cone et al. 2002). BAC-contigs form the platform on which full genome sequences are generally assembled (Green, 2001). This is a simple concept, but in reality constructing contigs is a challenging procedure.

1.3.4 Candidate gene approach

A candidate gene strategy is a putative way of relating a phenotype (Mendelian-inherited) trait to its biochemical basis by demonstrating a co-segregation between this candidate gene and the trait of interest (Gentzbittel et al. 1999). Beside expressed sequence tags (ESTs), a candidate gene sequence can be obtained by homology cloning (Gentzbittel et al. 1999). Using candidate gene approaches requires that the gene of interest has been isolated in another species and sequences are available to be used as probes or to design primers for PCR amplification in the crop. Major problem here is the degree of homology between genes in different species, which either allows to start with specific primers from conserved regions or to use degenerate primers if the sequence homology proves only moderate between species. Primers are then used to amplify the candidate gene either from genomic DNA or cDNA. The large amount of EST data now available for sunflower will prove to be helpful in applying candidate gene approaches (Horn and Hamrit 2009).

1.4 Mapping of major gene loci in sunflower

The identification and characterization of genes underlying important traits is a major objective of plant functional genomics (Lai et al. 2005). In sunflower, important genes and loci for disease resistance (Mouzeyar et al. 1995; Vear et al. 1997; Gentzbittel et al. 1998; Lawson et al. 1998; Lu et al. 1999), for defense reactions (Mazeyrat et al. 1998, 1999), and quantitative resistance have been documented (Mestries et al. 1998). The influence of drought on gene expression patterns has been described by Ouvrard et al. (1996), Cellier et al. (1998), Giordani et al. (1999). Other developmental genes expressed during floral and anther development, embryogenesis and organogenesis have been identified (Domon et al. 1990; Dudareva et al. 1996; Kräuter-Canham et al. 1997; Berrios et al. 2000a, 2000b, 2000c). León et al. (1996) identified RFLP markers linked to factors affecting hypodermis pigmentation (*hyp-1*) and assessed relationship between

this trait and oil percentage in sunflower seed. Based on a bulked segregant approach, Lu et al. (2000) reported the development of SCAR (sequence-characterized amplified region) markers from RAPD-markers linked to the sunflower *Or5* gene conferring resistance to a highly virulent Spanish population of broomrape (*Orobanche cumana*) that belongs to race E. These markers can be used either in a marker assisted selection (MAS) procedure for selection of new resistance lines or as a starting point to study the molecular genetics of broomrape resistance in sunflower. Leónardo et al. (2005) mapped the sunflower chlorotic mottle virus resistance gene (*Rcmo-1*) to linkage group (LG) 14 of the SSR map of sunflower (Tang et al. 2002). Genes for putative transmembrane receptors sharing epidermal growth factor and integrin-like domains have been identified by Gentzbittel et al. (2002). The oleoyl-phosphatidyl choline desaturase (*FAD2*) is necessary for the synthesis of linoleic acid from oleic acid. Three *FAD2* genes (*FAD2-1,-2,-3*) have been identified in sunflower (Hongtrakul et al. 1998a; Martinez-Rivas et al. 2001). It has been shown that *FAD2-1* is correlated with *Ol*, which is a chemically induced, incomplete dominant mutation, greatly increasing oleic acid content in developing seeds in sunflower. *FAD2-1* cosegregating with *Ol*, has been mapped to LG 14 (Lacombe and Bervillé 2001; Pérez-Vich et al. 2002). Hongtrakul et al. (1998a) developed dominant and codominant RFLP markers and a codominant SSR marker for *FAD2-1*. The codominant RFLP distinguished between *Ol* locus genotypes and can be used to accelerate the development of high oleic lines in sunflower. Schuppert et al. (2006) screened for the presence of tandem and inverted *FAD2-1* repeats, isolated and sequenced the intergenic region separating tandem repeats of *FAD2-1*. They developed dominant and codominant insertion-deletion (INDEL) markers and codominant SSR markers for *FAD2-1* for forward genetic analyses and marker-assisted selection in hybrid sunflower programs.

1.5 Male sterility in sunflower

Male sterility is a phenotypic trait in plants in which the male gametophytic function is prevented without affecting the female gametophyte. On the basis of inheritance patterns, two types of male sterility occur in sunflower: nuclear male sterility (NMS) and cytoplasmic male sterility (CMS) (Miller and Fick 1997).

1.5.1 Nuclear male sterility

Nuclear male sterility is generally the result of a single recessive gene. After crossing a nuclear male sterile line with any fertile genotype, the F_1-progenies are all male fertile (Jan 1997). Allelic relationships among NMS genes have been reported by Vranceanu (1970). He tested 10 NMS sources isolated from Romanian germplasm for allelism and reported the presence of five independent NMS genes, named *ms1* to *ms5*. Jan and Rutger (1988) evaluated the effectiveness of seven mitomycin-C and streptomycin induced NMS mutants from the cultivated line HA89, which were confirmed and placed in four different allelic groups, each representing a unique NMS gene, designated *ms6-ms9* (Jan 1992). Nuclear male sterility genes from two released male sterility lines, P21 and B11A3, have been designated *ms10* and *ms11* (Jan 1992), and were subsequently mapped on LG 11 and LG 8 by Pérez-Vich et al. (2005). Chen et al. (2006), identified DNA markers linked to the *ms9* gene in an F_2-population derived from the cross between NMS360 and RHA271 and mapped the *ms9* gene to LG 10 of the sunflower SSR linkage map (Yu et al. 2003).

The most interesting and useful application of NMS has been the linkage of nuclear male sterility with hypocotyl and leaf petiol color (Miller and Fick 1997). This type of nuclear male sterility has been used for creating testers for crossing with early generation inbred lines to produce testcross hybrids.

1.5.2 Cytoplasmic male sterility

If the trait leading to male sterility is not inherited according to Mendelian rules, but instead, represents a maternally transmitted trait, this type of male sterility is referred to as cytoplasmic male sterility (Laughnan and Gabay-Laughnan 1983).

The first report of such extranuclear inheritance of male sterility in plants was reported by Bateson and Gairdner (1921). They studied two strains of flax that produce male-sterile F_2-progenies, when crossed in one direction but not in the other direction. There have been over 70 sources of cytoplasmic male sterility (CMS) reported in the genus *Helianthus* since the original discovery by Leclercq in 1969 (Serieys 1996, 2005). CMS can occur spontaneously (Serieys and Vincourt 1987). It can also be induced by intraspecific crosses (Heiser 1982), interspecific crosses that introduce a nuclear background into a foreign cytoplasm (Leclerq 1969; Whelan 1980; Smart et al. 1994), or induced by mutagenesis (Jan 1988, Jan and Vick 2007).

There have been several CMS sources discovered from wild *H. annuus* and *H. argophyllus*, as well as more cytoplasmic sources from *H. petiolaris*. Several perennial species have also contributed to the development of new CMS sources (Tab. 1).

Tab.1 The FAO code, common name, species of origin, and reference of some of the known cytoplasmic male sterility (CMS) sources in sunflower (Crouzillat et al. 1991; Serieys 1996; Miller and Fick 1997)

FAO code	Common name	Species of origin	References
1. Spontaneous CMS			
ANN1	*H. annuus* 397	wild *H. annuus*	Serieys and Vincourt 1987
ANN2	*H. annuus* 517	wild *H. annuus*	Serieys and Vincourt 1987
ANN3	*H. annuus* 519	wild *H. annuus*	Serieys and Vincourt 1987
ANN4	*H. annuus* 521	wild *H. annuus*	Serieys and Vincourt 1987
2. Intraspecific crosses			
ANL1	Kouban Anashenko	*H. ann. lenticularis*	Anashenko 1974
ANL2	Indiana 1	*H. ann. lenticularis*	Heiser 1982
ANT1	Fundulea 1	*H. ann. texanus*	Vranceanu et al. 1986
3. Interspecific crosses			
PET1	French, Leclercq	*H. petiolaris*	Leclercq 1969
PET2	CMS 89(PET2), CMG1	*H. petiolaris*	Whelan 1980
PEF1	Fallax	*H. petiolaris fallax*	Serieys 1996
GIG1	CMS 89(GIG), CMG2	*H. giganteus*	Whelan 1981
MAXI	CMS 89(MAX1), CMG3	*H. maximiliani*	Miller and Wolf 1991
BOL1	*H. bolanderi*	*H. bolanderi*	Serieys 1984
EXI1	Exilis 1	*H. exilis*	Serieys and Vincourt 1987
ANO1	Anomalus	*H. anomalus*	Serieys and Vincourt 1987
PEP1	PET/PET	*H. petiolaris*	Serieys and Vincourt 1987
ARG3-M1	*Argophyllus* 3	*H. argophyllus*	Christov 1990
4. Induced by mutagenesis			
MUT1	Hemus	Irradiation of Hemus	Christov 1993
MUT2	Peredovick	Sonification of Peredovick	Christov 1993
ANN14	HA89	Irradiation of HA89	Jan and Vick 2006*
MUT7	HA89	Irradiation of HA89	Jan and Vick 2006*

* Newly identified cytoplasmic male sterility.

These cytoplasmic male sterilities could be an alternative to reduce the genetic vulnerability of sunflower to pathogens and increased the variability of genes inherited in a non-Mendelian way (Horn and Friedt 1997). Whatever its origin, it is generally said that CMS is due to a nuclear-cytoplasmic incompatibility and that different mechanisms are likely to exist in different systems (Horn 2006). However, many of the newly identified CMS sources do not have a complementary fertility restorer gene, or the cytoplasm is not stable. This has limited their use for commercial purposes (Seiler and Rieseberg 1997). To be useful for hybrid seed production, a CMS line needs complete male sterility and female fertility (Jan 2000).

In sunflower, commercial hybrid breeding is based on a single source of male sterility, PET1. This male sterility, which originates from an interspecific cross between *H. petiolaris* Nutt and *H. annuus* L. (Leclercq 1969), is a well-studied example of CMS resulting from anther degeneration (Chase, 2006). The PET1-cytoplasm causes premature programmed cell death (PCD) of the tapetal cells, which then extends to other anther tissues (Balk and Leaver 2001).

Alterations in the mitochondrial genome of PET1 compared to the fertile cytoplasm are limited to a 17-kb region and consist of two mutations: a 12-kb inversion and a 5-kb insertion/deletion which lead to an altered transcription pattern of the *atpA* gene (Siculella and Palmer 1988). The whole 5-kb insertion, found in PET1, is also present in all PET1-like CMS sources (Horn and Friedt 1999).

In PET1, the expression of a novel open reading frame, *orfH*522 in the 3′-flanking region of *atpA* gene is associated with the CMS phenotype (Köhler et al. 1991) which encodes a 16-kDa polypeptide (Horn et al. 1991; Laver et al. 1991). This 16-kDa protein was also detected by in organello translation in nine additional male sterile cytoplasms (Horn et al. 1996).

Nuclear genes, called restorer of fertility (*Rf*), have the function to suppress the effect of CMS-associated mitochondrial abnormalities on male fertility. In PET1, this involves the anther-specific reduction in the level of the co-transcript of *atpA-orfH*522 in the male florets of restored hybrid plants (Monéger et al. 1994; Smart et al. 1994; Schnable and Wise 1998; Gagliardi and Leaver 1999; Hanson and Bentolila 2004). Several restorer lines are available for PET1. Most of them carry the genes *Rf1* and *Rf2*, both responsible for restoring pollen fertility. *Rf2* was described to be present in nearly all inbred lines, including maintainer lines of PET1 (Miller and Fick 1997) and only the *Rf1* gene is introduced by the restorer lines to produce fertile sunflower hybrids.

Cytoplasmic male sterility in combination with fertility restoration is a widespread reproductive system that provides a useful tool to exploit heterosis in hybrid breeding and it is a desirable feature for use in hybrid seed production as it eliminates the need for hand emasculation.

1.6 Male fertility restoration and inheritance

Extensive research confirmed the complex genetic control of male-fertility restoration in sunflower (Miller and Fick 1997). Restoration of CMS is controlled by an array of genes, with variable reactions between genes and cytoplasms (Tab. 2).

Tab. 2 Inheritance of male fertility restoration of various cytoplasmic male sterility sources in sunflower (Miller and Fick 1997)

Cytoplasm	Genes	Source or gene action	References
PET1	*Rf1*	T66006-2-1	Kinman 1970
	Single dom	Acc(MO) 1338	Enns et al. 1970
	Rf2	MZ-1398	Vranceanu, Stoenescu 1971
	Rf1, *Rf2*	Two complementary genes	Fick and Zimmer 1974
	Two genes	Complementary, dominant	Vranceanu & Stoenescu 1978
PET1	Three genes	Complementary, dominant	Vranceanu & Stoenescu 1978
	Two genes	Cumulative, nonallelic	Vranceanu & Stoenescu 1978
ANL1	One gene	Single dominant	Kukosh 1984
PET2	One gene	Single dominant	Horn and Friedt 1997
	Two Complement.	CMG1	Whelan 1980
	Two dominant	RPET2, CMG1	Kural and Miller 1992
GIG1	One dominant	RGIG1, CMG2	Kural and Miller 1992
MAX1	Two complement.	RMAX1, CMG3	Kural and Miller 1992
	One dominant	RHA 274 (*Rf1*)	Kural and Miller 1992
	Two genes	Complementary dominant	Horn and Friedt 1997
ANL1	Two genes	Complementary dominant	Horn and Friedt 1997
ANL2	One gene	Single dominant	Horn and Friedt 1997
ANT1	Two complement.	Rf-ANT	Iuoras et al. 1992
ANN2	One dominant	P21, RMAX1, PI 413178	Jan 1991
ANN3	One dominant	P21, RPET2, PI 413180	Jan 1991
ANN4	Two genes	Dominant complementary	Horn and Friedt 1997
RIG1	Two complement.	RPET2, Luch	Jan et al. 1994
PEF1	Two complement.	Zaria, HA60	Miller 1996
	One gene	Single dominant	Horn and Friedt 1997

Male fertility restoration capability was reported from the sunflower line T66006-2-1, and it was determined that a single dominant gene, *Rf1*, conditioned male fertility restoration (Reviewed in Miller and Fick 1997). Several restorer lines used for hybrid seed production have been derived from this T66006-2-1 source (Korell et al. 1992). Working with Canadian lines, obtained from wild *H. annuus* and cultivated sunflower crosses, Enns et al. (1970) reported control of fertility restoration by a single dominant gene. Vranceanu and Stoenescu (1971) initiated studies to determine if inbred lines of various geographic and genetic backgrounds had factors for male fertility restoration. They found that only one source, MZ-1398, selected through eight generations of selfing from the local sunflower population Mezehedeshy showed restoration capacity. Fick and Zimmer (1974) confirmed the presence of two complementary dominant genes in crosses between the non-oil seed lines CMS HA267 and RHA280. Research on 16 sources of pollen fertility restoration revealed that this trait was controlled in eight cases by one single dominant gene, in three cases by two complementary genes, in four cases by three complementary genes and in one case by the cumulative action of two non-allelic dominant genes (Vranceanu and Stoenescu 1978). Studies reported a single dominant gene, *Rf1*, which restored male fertility to the cytoplasm derived from *H. annuus* ssp. *lenticularis*, or CMS ANL1 (Reviewed in Miller and Fick 1997). Extensive research confirmed the complex genetic control of male fertility in sunflower. Fourteen restorer lines were investigated for genetic control of PET1 male sterility. Five lines were found to control restoration by a single dominant gene. Five other lines controlled restoration by two independent dominant genes and four lines controlled restoration by two complementary dominant genes (Miller and Fick 1997). Segregation analyses of the F_2-populations indicate that a single dominant restorer gene was sufficient to restore pollen production of hybrids based on ANL2, PEF1 and PET2 (Horn and Friedt 1997). Horn and Friedt (1997) showed that restoration of ANN4 required two dominant complementary genes and two dominant genes were involved in restoration of fertility in the crosses based on ANL1 and MAX1. Consequently Horn and Friedt (1997) identified on the basis of six new CMS cytoplasms stable fertile hybrids. Inheritance studies are complicated by complementary genes existing in several female lines, as well as genes present in the restorer lines (Miller and Fick 1997). Isolation of the restorer gene may help to elucidate this complexity.

1.7 Molecular studies on genes restoring male fertility in sunflower

A map-based cloning approach was started by several groups over the world to isolate gene (s) restoring male fertility in sunflower. Gentzbittel et al. (1995, 1999) reported the presence of two distinct fertility restoration loci of the PET1 cytoplasmic male sterility. *Rf1* was mapped on LG 6 and *Msc1* mapped on LG 12 in the RFLP map and the consensus map. For this study, they analyzed seven F_2-populations, for a total of 1,115 individuals. The authors presented no evidence as to whether these two loci are duplications of a single gene, but their genetic independence would make it possible to combine the two loci in an inbred line without difficulty. Jan et al. (1998) mapped the *Rf1* fertility restoration gene to one end of the LG 2 at a distance of 25 cM from an adjacent marker in their map. For their investigations, they used 93 F_2-plants of a cross between inbred lines RHA271 and HA234. A screening of 1,200 arbitrary decamer primers (Horn et al. 2002) and 1,024 AFLP primer combinations had been investigated by bulked segregant analysis to identify markers closely linked to the restorer gene *Rf1* (Kusterer et al. 2002). Horn et al. (2003) developed PCR-based markers for the restorer gene *Rf1* of sunflower hybrids based on the PET1 cytoplasm. This study was based on 183 individuals of an F_2-population (RHA325 × HA342). Additionally, 75 SSR markers (Tang et al. 2002; Burke et al. 2002) were analyzed and were mapped in the F_2-population. The SSR-marker ORS-1030 allowed the assignment of the restorer gene *Rf1* to the LG 13 of the general sunflower SSR-based consensus map of Tang et al. (2002) (Kusterer et al. 2005, 2004b). The linkage group around the restorer gene *Rf1* contained 43 markers, i.e. 35 AFLP markers, 7 RAPD markers and 1 SSR marker covering 250.3 cM (Fig. 4). The restorer gene *Rf_PEF1*, restoring pollen fertility in hybrids based on one of the new CMS sources, PEF1, could be linked with AFLP markers (Schnabel et al. 2008). In addition, SSR marker analyses demonstrated that *Rf_PEF1* is not located on LG 13 as *Rf1*.

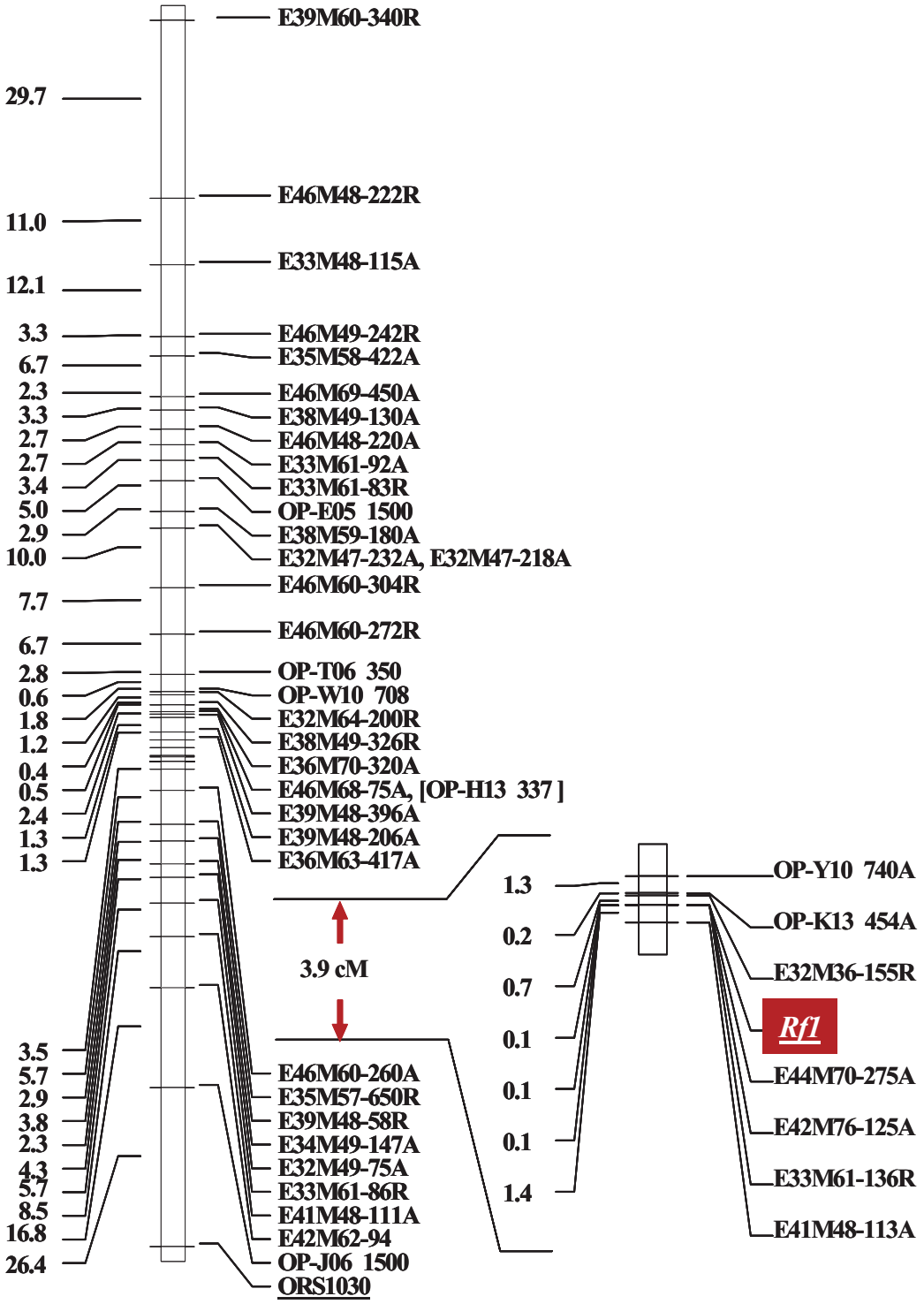


Fig. 4 Marker-saturated map of LG 13 carrying the restorer gene *Rf1*. AFLP-markers, RAPD-markers and SSR markers were mapped in the F_2-population of the cross RHA325 × HA342. A region of 3.9 cM around the *Rf1* gene is shown enlarged (Kusterer et al. 2005)

Recently, a restorer of fertility gene *Rf3*, which is different from both *Rf1* and *Rf2*, was assigned to the confection line RHA280 (Jan and Vick 2007). Feng and Jan (2008) identified a new completely dominant restorer gene, *Rf4*, which restored pollen fertility in the presence of a new type of cytoplasmic male sterility named CMS GIG2 from an interspecific cross between *H. giganteus* and *H. annuus* L. cv. HA89. The *Rf4* gene was mapped to the LG 3 with SSR markers and RFLP-derived STS-markers, and was localized about 0.9 cM away from the SSR marker ORS1114. For this study, Feng and Jan (2008) used 200 SSR markers randomly selected from the map of Tang et al. (2002) and Yu et al. (2002, 2003) based on a segregation population of 933 individuals. They concluded that this restorer of fertility is different from *Rf1* of CMS PET1 because all the restorer lines for CMS PET1 did not restore pollen fertility to CMS G1G2 (Feng and Jan 2008). All these markers linked with the restorer of fertility genes provide an essential basis for genetic analysis of these genes and marker-assisted selection in sunflower breeding. In the future, the isolation of these fertility restorer genes from sunflower will hopefully help to understand the different molecular mechanisms behind cytoplasmic male sterility and its fertility restoration. Comparison of different restorer genes within sunflower will be of special interest (Horn and Hamrit 2009).

1.8 CMS-restorer systems identified in other plants

CMS-restorer systems have been identified in many higher plants. To date, a number of restorer fertility genes have been identified and cloned. In maize the *Rf2* gene, the first fertility restorer gene to be cloned, was found to encode an aldehyde dehydrogenase (Liu et al. 2001; Cui et al. 1996). In contrast, the other restorer genes cloned from petunia (Bentolila et al. 2002), radish (Imai et al. 2003; Koizuka et al 2003; Brown et al. 2003; Desloire et al. 2003), rice (Kazama and Toriyama 2003; Akagi et al. 2004; Komori et al. 2004; Wang et al. 2006), as well as from sorghum (Klein et al. 2005) encode pentatricopeptide repeat (PPR) proteins. In *Arabidopsis thaliana*, the PPR containing protein family contains over 400 members, with the number of repeats per protein within this family varying from 2 to 26. PPR proteins have been implicated in processing of organellar transcripts (Small and Peeters 2000) which are targeted to mitochondria and play a role in regulating the expression of the abnormal CMS-associated genes at the transcriptional and/or translational level (Aubourg et al. 2000).

So far, only a few PPR genes of various organisms have been studied in detail, and little is known about their molecular functions (Lurin et al. 2004).

Male sterility in petunia is encoded by an abnormal mitochondrial gene termed *pcf* (petunia CMS-associated fused gene), which disrupts pollen development (Hanson et al. 1999). The petunia *Rf* gene was isolated by a map-based cloning strategy, using a cross between *Rf/rf* and *rf/rf* lines (Bentolila and Hanson 2001). In lines containing the *Rf* allele, the *pcf* transcript is altered and the amount of the PCF protein is greatly reduced (Bentolila et al. 2002). In petunia, the *Rf* locus is composed of duplicated genes containing 14 repeats of a PPR motif (35-aa) (Bentolila et al. 2002). The authors identified in a non-restoring genotype a homologous gene that exhibits a deletion in the promoter region. Advanced analysis of the restorer protein revealed that it is part of a soluble mitochondrial inner-membrane (Gillman et al. 2007).

In radish, Ogura cytoplasm male sterility is caused by an aberrant mitochondrial gene, *orf138* (Iwabuchi et al. 1999), which prevents the production of functional pollen. A single radish nuclear gene, *Rfo*, alters the expression of *orf138* at the post-transcriptional level (Desloire et al. 2003). Using a map-based cloning approach relying on synteny between radish and *Arabidopsis* the *Rfo* gene has been cloned (Brown et al. 2003; Desloire et al. 2003). Another group isolated a Kosena CMS-associated gene, *orf125*, with extensive similarity to the *orf138* (Koizuka et al. 2003). The *Rfk1* restorer gene has been identified (Imai et al. 2003) and isolated (Koizuka et al. 2003), which regulates the expression of *orf125* at the translational or post-translational level in Kosena CMS radish (Koizuka et al. 2000). *Rfo/Rfk1* encodes a predicted protein of 687 amino acids comprising 16 copies of the PPR motif. Presence of *Rfo/Rfk1* reduces the amount of the CMS-associated mitochondrial protein (ORF138/ORF125) without changing the level of mRNA. Koizuka et al. (2003) reported the isolation of male sterile allele of this gene from a CMS Kosena radish. This allele has four substituted amino acids compared to the fertility restoring allele. Recently, Wang et al. (2008) identified and mapped a new restorer gene termed *Rfo2* at a distance of 1.6 cM from the *Rfo* locus.

In rice, one particular CMS system has been obtained by combining the cytoplasm of Chinsurah Boro II (*Indica* rice) with the nuclear genome of Taichung 65 (*japonica* rice), which is called the *ms-bo* type or BT type (reviewed in Kazama and Toriyama 2003). A fertility restorer gene *Rf-1* has been reported to promote the processing of the aberrant *B-atp6* RNA when introduced to the CMS line (Kazama and Toriyama 2003). The *Rf-1*

gene encodes a 791-aa protein containing a mitochondrial target signal and 16 PPR motifs (Akagi et al. 2004). Sequence analysis revealed that the recessive allele *rf-1* lacks one nucleotide in the putative coding region, presumably resulting in encoding a truncated protein because of a frame shift (Komori et al. 2004). Two fertility restorer genes, *Rf1a* and *Rf1b*, were identified at the classical locus *Rf-1* as members of multigene cluster that encode PPR proteins. RF1A and RF1B are both targeted to mitochondria and can restore male fertility by blocking ORF79 production via endonucleolytic cleavage (RF1A) or degradation (RF1B) of dicistronic B-*atp6*/*orf79* mRNA (Wang et al. 2006).

Klein et al. (2005) utilized a positional cloning approach in conjunction with microcolinear analysis of sorghum and rice to genetically and physically delimit the *Rf1* locus of sorghum. They have tentatively identified the sorghum *Rf1*, one of the major loci controlling fertility restoration in presence of the A1 cytoplasm in sorghum and showed homology to the PPR protein family on the amino acid level.

1.9 Objectives of this study

The CMS-associated region in plant species studied so far appear to carry novel recombinant genes or open reading frames that are usually co-transcribed with normal mitochondrial genes (Hanson and Bentolila 2004). Cytoplasmic male sterility is a classical example of a genome conflict: maternally inherited mitochondrial genes, which induce male sterility, interact with nuclear genes, which restore male fertility (Budar et al. 2003). All identified fertility restorer genes, except for *Rf2* in maize (Liu et al. 2001), are members of the PPR family; petunia (Bentolila et al. 2002), radish (Imai et al. 2003; Koizuka et al 2003; Brown et al. 2003; Desloire et al. 2003), rice (Kazama and Toriyama 2003; Akagi et al. 2004; Komori et al. 2004; Wang et al. 2006), as well as sorghum (Klein et al. 2005). These fertility restorer genes have to block or compensate the cytoplasmic dysfunctions that lead to disturbances during pollen development. The mechanisms by which restoration occurs are probably as diverse as the mechanisms by which mitochondrial mutation cause CMS (Schnable and Wise 1998).

The diversity in restoration systems extends to the number of restorer genes. In some systems, one or two major restorer loci confer complete restoration, in others, full fertility restoration requires the combined action of several genes, many of which provide only small incremental effects (Schnable and Wise 1998).

In almost all reports, male fertility restoration in sunflower is controlled by two or more complementary dominant genes (Crouzillat et al. 1991). However, the introduction of one dominant, nuclear-encoded restorer gene *Rf1* is in most cases sufficient to restore fertility in sunflower (Kusterer et al. 2002). Additional information about the restorer gene and its alleles should give insight into the question of its origin, and the mechanism behind its action. The isolation and characterization of the restorer gene *Rf1* will have practical significance for the plant breeder and will facilitate the understanding of the interaction between nucleus and cytoplasm. The availability of markers tightly linked to *Rf1* (Horn et al. 2003; Kusterer et al. 2005), and the establishment of the putatively closed contig (Kusterer et al. 2004a) represent an appropriate basis to isolate the *Rf1* gene using a map-based cloning approach.

The main objectives of this study were:

1. to verify the preliminary contig around the restorer gene *Rf1*,
2. to use additional markers tightly linked to the gene *Rf1* as overgo probes against two sunflower BAC libraries (RHA325 and HA383) in order to identify positive BAC-clones for chromosome walking,
3. to characterize the positive BAC-clones by fingerprinting and pulse field gel electrophoresis,
4. to isolate BAC-ends from the positive BAC-clones in order to generate new probes to rescreen the two BAC libraries and to walk towards the *Rf1* gene, and
5. to assembly a closed contig spanning the *Rf1* region.

2. Materials and Methods

2.1 Materials, equipments and suppliers

Solutions, buffers, media, and materials (e.g. tips, Eppendorf tubes, glassware) used during this study were autoclaved for 25 min and 2.1 bars at 120 °C. The tooth picks used for picking up the BAC-clones were sterilised for 8 h at 160 °C. The companies' supplier, materials and equipments are listed in the appendix (Tab. 10).

2.2 Plant material

The initial work has been carried out with a segregating sunflower F_2-population derived from the cross between the lines RHA325 and HA342 (Horn et al. 2003). RHA325 is an American public restorer line carrying the PET1 cytoplasm and HA342 is a public American high oleic maintainer line (Kusterer et al. 2005).

For fine-mapping of the *Rf1* gene, the segregating population was enlarged to 1,571 F_2-individuals. F_2-plants, F_3-families and BC_1F_2 were evaluated for male fertility and male sterility at the field station Gross-Gerau, near Frankfurt am Main.

Two bulks of ten homozygous male fertile and ten male sterile F_2-plants were used for bulked segregant analyses (BSA).

Tab. 3 Oligonucleotides constructed from markers closely linked to the restorer gene *Rf1* and used to screen the two BAC libraries (RHA325 and HA383).

Primers		Primer probe name	Sequence 5' → 3'
AFLP	E42M76-125A	E42M76 125-Ova E42M76 125-Ovb	CTCCATACACTCCGCAGCAGAATG TTCCTCTCTGCTACTCCATTCTGC
	E32M36-155R	E32M36 161-Ova E32M36 161-Ova	ACCAAACTGGTTATTCCATTTTGG CCATTTCGAGTAGTAACCAAAATG
	E33M61-136R	E33M61-136-Ova E33M61-136-Ovb	GAGAGCGAAAGATCAAGACTAAAG ATCTTCTGTCTTCAGTCTTTAGTC
	E41M48-113A	E41M48-18 Ova E41M48-18 Ovb	GTTCCACTTTCTGTCAACGAGGAA CAGCTACGAAAGCTAGTTCCTCGT
	E44M70-275A	Ova E44M70 Ovb E44M70	GGATTCATCCCTAGAAGTGCAAGAT CTTCCTTGAACACCTTCAATCTTGCA
RAPD	OP-K13-454A	OP-K13-Ova OP-K13-Ovb	CCCTTCCATTAAAACCCAGCATCA GTGTACATGTCGAAAGTGATGCTG
	OP-Y10-740A	Y10 Ova Y10 Ovb	CGTGGGAGAGAGGTGGCCAACCCC AAAATGTCTCCTAATTCCCCAACC
	OP-H13-337A	H13 Ova H13 Ovb	TCTGTGTTCAATTGAGAATTCGCA CCAATTTCATGGGTTGACGCTTAA

2.3 Bacterial Artificial Chromosome libraries

For the construction of the contig around the restorer gene R*f1*, two Bacterial Artificial Chromosome (BAC) libraries were screened.

The first BAC library was constructed from the restorer line RHA325 using pBeloBAC11 as vector and *Hind*III as restriction endonuclease (Özdemir et al. 2002, 2004). This BAC library comprises 104,736 BAC-clones which are stored in 335 microtiter plates. Each microtiter plate contains 384 individual BAC-clones in glycerol stock solution. The average insert size is about 60 kb and covers 1.9 sunflower genome equivalents. The BAC library was gridded onto four high density nylon membrane filters (Özdemir et al. 2004). The BAC-library RHA325 was enlarged with 20,000 BAC-clones, picked and double-spotted onto 2 high density membrane filters at Clemson University Genomics Institute (CUGI, http://www.genome.clemson.edu). The enlarged BAC-library is thought now to have larger genome coverage.

The second BAC library was purchased from CUGI. This BAC library named HA_HBa was constructed from the maintainer line HA383 using pIndigoBac536 vector and also digested with *Hind*III as restriction endonuclease. The BAC library has an average insert size of 125 kb and represents about 8.3 genome equivalents. It consists of 202,752 BAC-clones distributed in 528 microtiter plates and double- spotted on 11 high density filters.

Hybridization using overgo probes were performed against these two BAC libraries in order to obtain positive BAC-clones.

2.4 Overgo design

Overlapping oli**go**nucleotides, also called “overgos”, are primer pairs that usually constitute of two 24 bp oligonucleotides that contain an 8-bp overlapping region at the 3′ end. This structure allows the overgo primer pair to prime on each other and synthesize their complementary strands by incorporating radioactive labelled nucleotides by the Klenow fragment leading to a radioactively labelled double-stranded DNA fragment of 40 bp (McPherson, http://www.tree.caltech.edu) (Fig. 5).

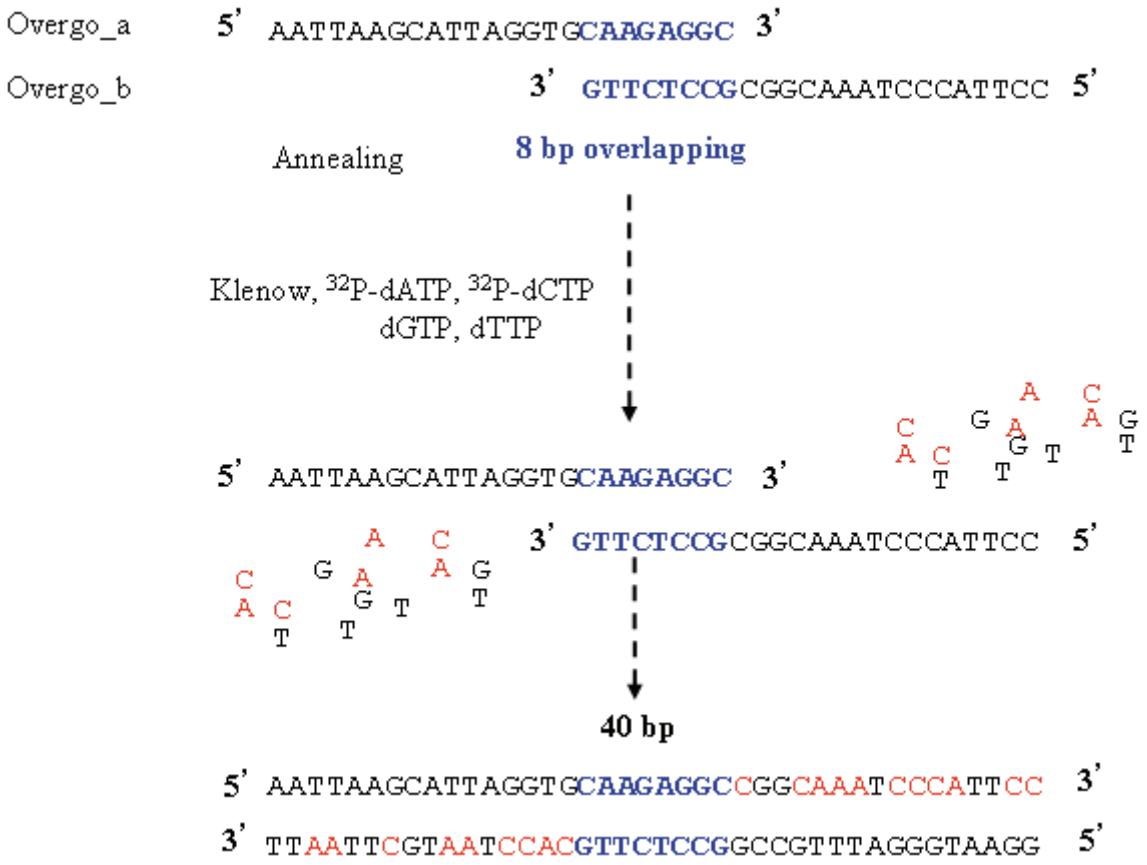


Fig. 5 Schematic representation of an overlapping oligonucleotide probe (overgo) structure and its labelling. The blue colour represents the 8 bp overlapping region. A and C nucleotides represent the [^{32}P] labelled portion of the probe. G and T are non-radioactive nucleotides.

Overgo labelling buffer: OLB [-A,-C,-N6]

- Solution O: 1.25 mM Tris-HCl, pH 8.0 ; 125 mM $MgCl_2$
- Solution A: 1 ml solution O; 18 µl 2-mercaptoethanol; 5 µl 0.1 M dTTP; 5 µl 0.1 M dGTP
- Solution B: 2 M HEPES-NaOH, pH 6.6
- Solution C: 3 mM Tris-HCl, pH 7.4; 0.2 mM EDTA

OLB: A: B: C, 1: 2.5: 1.5 [aliquot and stored at -20 °C]

2.4.1 Design of overgos from AFLP and RAPD markers

The cloned RAPD markers and the AFLP markers (Tab. 3) most closely linked to the restorer gene *Rf1* (Kusterer et al. 2005) have been used as overgo probes to screen the

two large insert bacterial artificial chromosome libraries in order to walk towards the restorer gene *Rf1*.

2.4.2 Overgo annealing and labelling

Overgos were annealed by combining 10 µl (20 pmol/µl) of each oligo to obtain a final volume of 20 µl that was heated in a thermocycler for 5 min at 80 °C, followed by 10 min at 37 °C, and cooled on ice. Overgo labelling with [α-^{32}P] dATP and [α-^{32}P] dCTP was performed as described by McPherson (http://www.tree.caltech.edu). Briefly, 1 µl annealed overgo was mixed with 0.5 µl bovine albumin serum (2 mg/ml), 2 µl OLB (oligo labelling buffer without dATP, dCTP, and random hexamers), 1 µl Klenow fragment (2 U/µl), 0.5 µl [α-^{32}P] dATP (3,000 Ci/mmol), 0.5 µl [α-^{32}P] dCTP (3,000 Ci/mmol), and 4.5 µl ddH_2O. The mixture was incubated for 1 h at 20 °C in a PCR thermocycler. The labelled overgos were denatured for 10 min at 94 °C before using them for hybridization.

2.5 Filter hybridization and washing

2.5.1 Prehybridization

BAC filter membranes were rolled (up to 5 filters) and inserted into a 300 ml glass hybridization tube containing 20 ml of hybridization solution. In case of Southern blot membranes, only 10 ml of hybridization solution were used. Membranes were prehybridized for at least 1 h in the incubation oven at 58 °C.

2.5.2 Hybridization

The overgo probe was denaturated for 10 min at 90 °C and immediately added to the 20 ml (or 10 ml) of hybridization solution that had been used for pre-hybridization. The hybridization was incubated for 18-24 h overnight at 58 °C.

Hybridization solution (7% SDS)

- 1% BSA; 1 mM EDTA; 7% SDS; 0.5 M Sodium phosphate
- 1 M sodium phosphate, pH 7.2: 134 g Na_2HPO4; $7H_2O$, add 4 ml 85% H_3PO_4 and fill up to 1 l with ddH_2O

2.5.3 Filter washing

In order to remove non-specific bindings and to reduce the background without affecting the specific probe hybridization, two 10 min washes in 2×SSC, 0.1% SDS (wash 1), and 1×SSC, 0.1% SDS (wash 2) at 58 °C were performed. Wash 1 and wash 2 were pre-warmed overnight at 58 °C. Wash 1 was realised in the hybridization bottles in the rotisserie oven but wash 2 was done in trays for the filters. For the Southern blots, the washing was done in the hybridization bottles.

2.5.4 Stripping BAC filters

To remove radioactivity, filters were washed in 100 mM NaOH, 10 mM EDTA, and 0.1% SDS for 10 min at room temperature with gentle shaking. If necessary, this step was repeated until no signals on filters were detected. The stripping solution was removed by rinse with ddH_2O and followed by another wash in 5×SSC for 10 min.

Washing solution

- 20×SSC: 3 M Sodium chloride; 0.3 M tri-Sodium-Citrat$_2$-Hydrat

2.5.5 Autoradiography

The filters were sealed in plastic bags and exposed to X-Ray films for 4 to 5 days at -80 °C. After scanning, all filter sets were stored in sealed plastic bags at room temperature for further use. Each positive BAC-clone generates a characteristic two spot pattern. After identification, all the positive BAC-clones were picked from the BAC library and isolated.

2.6 Spot pattern of the BAC libraries

2.6.1 BAC library HA383

The BAC library HA383 constructed by CUGI consists of 11 filters (A to K). Each filter or membrane is divided into six fields. Each field contains 384 squares (16 × 24). Within each square there are 16 positions (4 × 4) where 8 BAC-clones are spotted in duplicate (Fig. 6). Each filter contains 36,864 spots corresponding to 18,432 individual BAC-clones. During the double-spotting of the filters B, D, F, H, J, the positions 3 and 4 have been mixed up. This means that the position 3 correctly refers to the position 4, and vice versa.

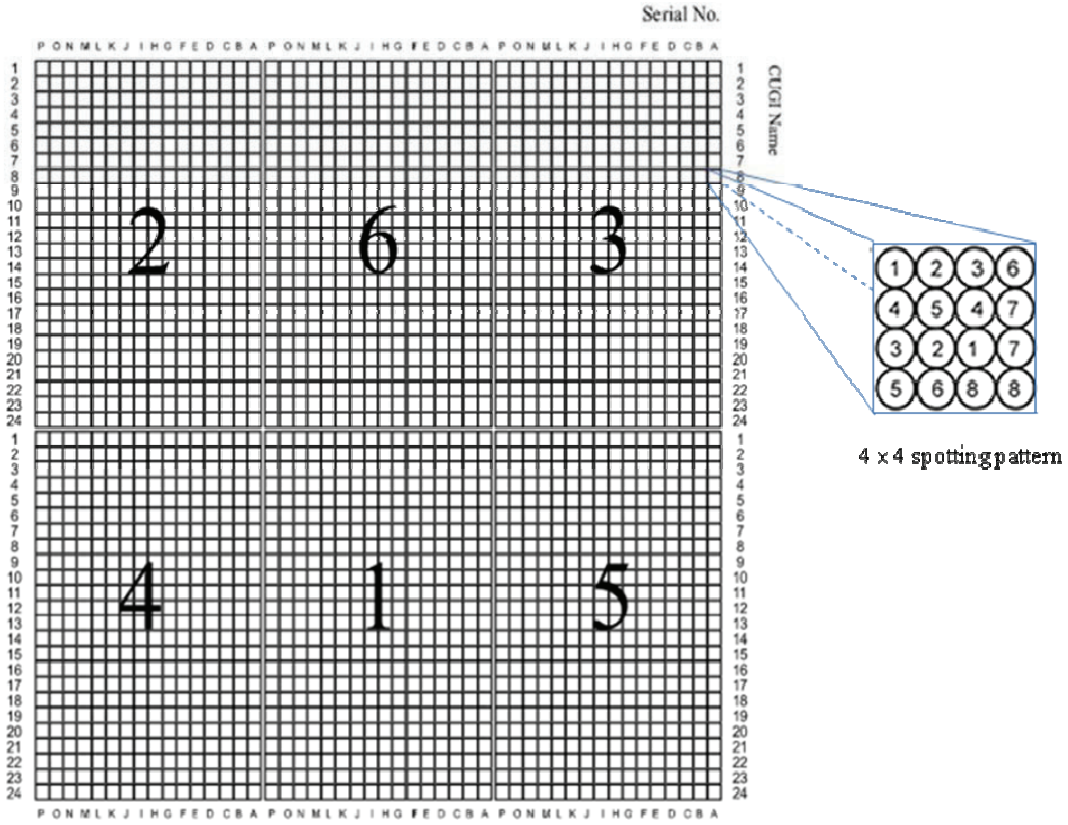


Fig. 6 Filter illustration of the BAC library HA383. Standard CUGI filter layout and spotting pattern.

2.6.2 BAC library RHA325

The BAC library RHA325 (Özdemir et al. 2004) consists of four filters (A to D). Each filter or membrane is divided into 6 fields. Each field contains 384 squares (16 × 24). Within each square there are 25 positions (5 × 5, the middle of the square is left free) where 12 BAC-clones are spotted in duplicate (Fig. 7). Each filter contains 55,296 spots corresponding to 27,648 individual BAC-clones. The original set of filters was enriched with two new filters, HornA and HornB, from the BAC library RHA325. They were spotted at CUGI, following CUGI's 4 × 4 duplication pattern and they represent 20,000 individual BAC-clones.

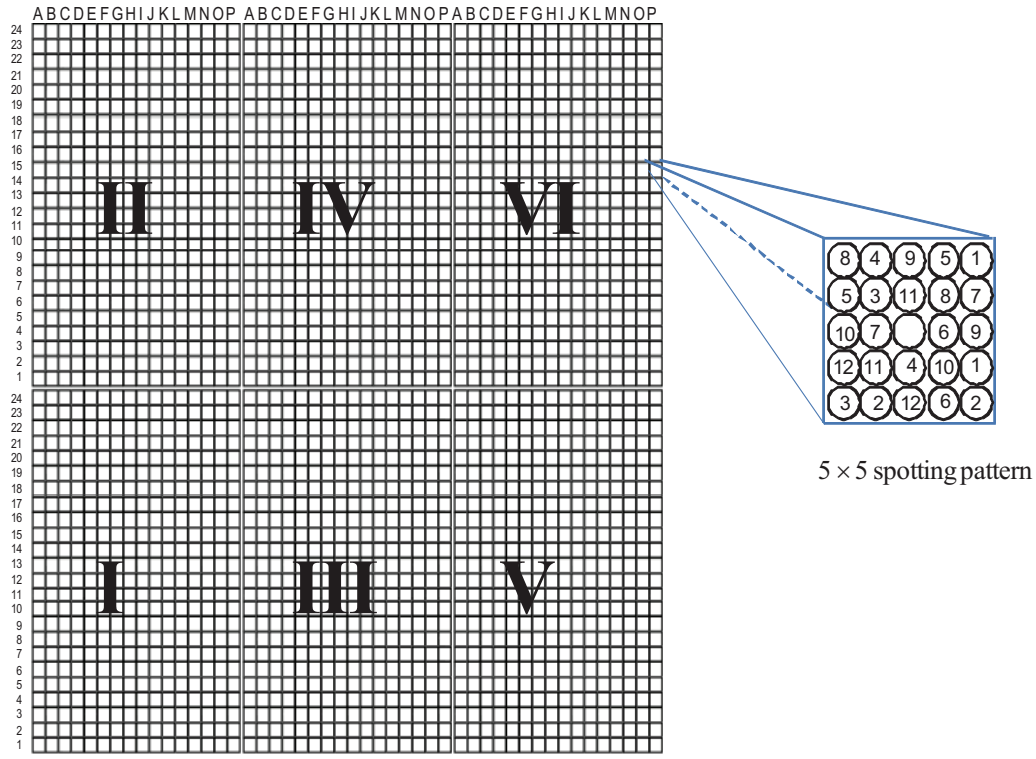


Fig. 7 Filter illustration of the BAC library RHA325. Filter layout and spotting pattern.

2.7 Isolation of the DNA from BAC-clones by alkaline lysis

2.7.1 Principle of the isolation method

The BAC-clone DNA was isolated by the alkaline lysis method (Birnboim and Doly 1979). It is a rapid and small-scale isolation of plasmid DNA from bacteria. Bacteria containing the DNA were first grown and then lysed with a strong alkali buffer consisting of the detergent SDS and the strong base NaOH. The detergent breaks the membrane's phospholipids bilayer and the alkali denatures proteins involved in maintaining the structure of the cell wall. Through a series of steps involving agitation, precipitation, centrifugation, and the removal of the supernatant, cellular debris was removed and the plasmid was isolated and purified.

2.7.2 Minipreparation

A single colony was grown in 5 ml LB (Lurea Bertani) medium, for approximately 20 h at 37 °C with shaking (200 rpm); chloramphenicol (final concentration 12,5 µg/ml) was added to the medium to select for the vectors pBeloBAC11 or pIndigo536. About 150

µl of each overnight culture were added to 150 µl of sterilised 50% glycerol and stored as glycerol stock at -80 °C for further use. After centrifugation for 2 min at 5,000 rpm, the bacterial cells were resuspended in 300 µl of isotonic solution I (GTE) containing 10 mg/ml RNase. Then 300 µl of freshly prepared solution II (0.2 N NaOH and 1% SDS) were added, incubated for 5 min, gently mixing the samples. The pH of the lysate was then neutralized in 300 µl of an acetate buffered solution III containing a chaotropic salt (3M potassium acetate), and mixed gently at room temperature for 5 min. Following centrifugation at 13,000 rpm, 10 min, a 0.6-fold volume of isopropanol was added to the supernatant and centrifuged for 10 min at 13,000 rpm to precipitate the DNA. The DNA was washed once with 1 ml 70% ethanol, and then spun for 2 min. The supernatant was discarded, and the pellet was dried in a speed vac for 1 h and resuspended in 25 µl TE. Efficiency of the plasmid Minipreparation was analysed on 0.8 % agarose gels. All positive BAC-clones were individually fingerprinted and sized (see 2.8).

Mini-preparation solutions

- LB medium: Trypton 1%; yeast extracts 0.5%, NaCl 1%
- Suspension solution (GTE): Glucose 50 mM; Tris/HCl 25 mM; EDTA 10 mM
- Lysis solution: NaOH 0.2 N; SDS 1%
- Suspension solution (TE): 10 mM Tris/HCl pH 8.0; 1 mM EDTA pH 8.0

2.8 Fingerprinting and BAC insert sizing

Ten µl of DNA was digested with 20 U *Hind*III (10 U/µl) for 3 h at 37 °C. Then, 1.5 µl of loading buffer were mixed with the digested DNA and all was subjected to electrophoresis on 0.8% agarose gel in 0.5×TAE buffer together with the 1 kb ladder marker loaded at the both sides of the gel. The gel was then stained in ethidium bromide (10 mg/ml) for 20 min, and washed with water. The bands were imaged and sized using the Bio-Rad imaging devices supported by the Quantity One Software. The insert size of each BAC-clone was determined by adding up all insert DNA fragments. BAC-clones were also sized by pulse field gel electrophoresis (PFGE) on 1% (w/v) agarose gel in 0.5×TBE together with the marker MidRange II PFG (*BioLabs*) loaded at the both sides of the gel using a CHEF (contour-clamped hexagonal electric field) DR III apparatus (BioRad, Richmond, California). Electrophoretic conditions were: 6 V/cm, a constant initial pulse time of 1 s, a final pulse time of 25 s, 120°, for 22 h, at 14 °C in 0.5×TBE buffer. The gel was stained in ethidium bromide for 30 min, washed with

water and photographed. A band of about 7.5 kb appearing in all lanes represents the cloning vector. Fingerprinted BAC-clones were blotted on nylon membranes for further Southern hybridizations.

Electrophoresis buffer

- 50×TAE: 2 M Tris-Base; 0.05 M EDTA; 57.2 ml Acetic acid, adjust pH to 8.0
- Loading buffer for 10 ml TAE: 0.05 g Bromophenol blue; Xylene cyanol; 5 ml Glycerol; 0.186 g EDTA
- 10×TBE: 108 g Tris-Base; 55 g Boric acid; 40 ml (0.5 M; pH 8.0) EDTA

2.9 Southern hybridization

2.9.1 Principle of Southern hybridization

Southern hybridization is a method used to check for the presence of a DNA sequence in a DNA sample. Southern hybridization combines agarose gel electrophoresis for size separation of DNA fragments with methods to transfer the size-separated DNA to a filter membrane for probe hybridization.

The transfer method is a technique for transferring DNA molecules from an agarose gel to a thin membrane made out of nitrocellulose or nylon. Gel embedded DNA is depurinated in the presence of HCl followed by denaturation in an alkaline solution. Denaturation prior to blotting is essential for dissociating the two polynucleotide chains, which are then hybridized against a probe. Denatured molecules move upward by capillary action of the buffer and on coming in contact with the membrane bind to it. A brief UV treatment facilitates firm binding of DNA to the membrane (Fig. 8).

2.9.2 Electrophoresis

Isolated BAC-DNAs were digested with *Hind*III restriction endonuclease. The DNA fragments were electrophoresed on 0.8% agarose gel containing 1×TAE buffer at 80V until the Bromophenol blue indicated that the sample has run for a sufficient distance. The gel was visualized and photographed using Quantity One Software (BioRad).

2.9.3 Southern blot

For depurination, the gel was treated with HCl (0.25 M) for 8 min at room temperature with gentle agitation. This step is necessary if the fragments are greater than 10 kb. DNA was incubated in a sufficient volume of denaturation buffer for 45 min at room temperature with gentle agitation. The gel was then neutralized in neutralizing buffer for 25 min. A rinse in distilled water followed each step. The capillary blotting apparatus was assembled according to Figure 8.

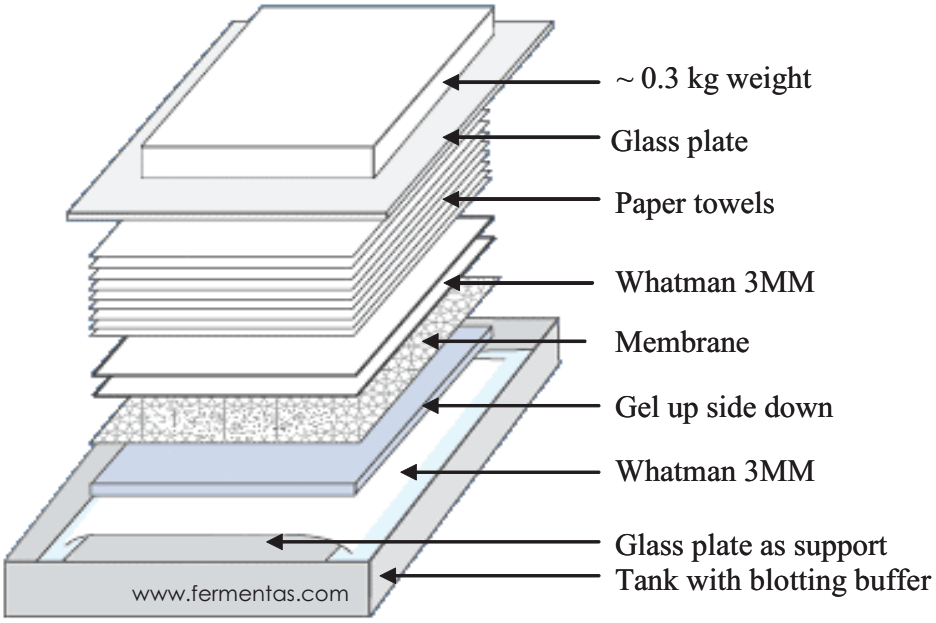


Fig. 8 Principle of Southern Blot

A bridge of glass support was made on the plastic tray, which was filled with about 1.5 to 2.5 cm of 10×SSC as transfer buffer. Two layers of Whatman 3 MM paper, cut according to the gel size, were prewetted with 10×SSC and placed on top so that the ends of the paper were submerged in the buffer below. Air bubbles were smoothed out between the paper and the glass plate using a glass pipette.

The gel was placed on the moist Whatman paper. Nylon Hybond membrane was cut to the proper size of the gel, soaked in 2×SSC and placed on the gel. Any air bubbles were smoothed out between the membrane and the gel. Strips of plastic were laid down along each of the four edges of the membrane to cover any exposed parts of the gel. The membrane was then covered with two pieces of Whatman 3 MM paper of the same size

as the membrane, soaked in 2×SSC, and air bubbles were smoothed out between the paper and the membrane.

About 5 to 8 cm thick layer of paper towels were placed on the top of the Whatman papers, followed at the top by a glass support, which carried a weight (about 300 g). The DNA was transferred overnight.

On the next day, the apparatus was dissembled, the membrane was taken carefully, washed in 6×SSC buffer to remove any residual agarose, was dried at room temperature and the DNA was fixed to the membrane by UV cross-linking. The membrane can be used for non-radioactive and radioactive hybridization.

Southern blot buffers

- Depurination buffer: 250 mM HCl
- Denaturation buffer : 1.5 M NaCl; 0.5 M NaOH
- Neutralization buffer: 0.5 M Tris-HCl pH 7.5; 1.5 M NaCl
- 2×SSC: 0.3 M tri sodium citrate; 3 M NaCl
- 6×SSC: 0.3 M tri sodium citrate; 3 M NaCl

2.10 Non-radioactive Southern hybridization

The ECL™ (**E**nhancer **C**hemi**l**uminescence) direct nucleic acid labelling detection system is based on enhanced chemiluminescence. The system involves direct labelling of the probe DNA with the enzyme horseradish peroxidase. The non- radioactive hybridization was performed according to the protocol described in the manual of the manufacturer (Fig. 9). For labelling of the probe, 10 µl of the probe were denaturated for 10 min at 95 °C. The probe was cooled down then, peroxidase, which has been complexed with a positively charged polymer, was added in the same amount (10 µl) as the probe for 90 s at 37 °C. Loose attachments are formed with the nucleic acid by charge attraction. Ten µl of glutaraldehyde, which causes the formation of chemical cross-links so that the probe is covalently linked to the enzyme, were added and incubated for 10 min at 37 °C. Prior to the labelling process, the membrane was pre-hybridized in a hybridization solution delivered with the kit for at least 1 h at 42 °C. Then, the labelled probe was added. The hybridization was performed overnight. On the next day, the membrane was washed to remove unbound probe, twice 20 min with wash buffer 1 and once for 5 min with wash buffer 2. Both washings were performed at 42 °C. The membrane was placed immediately between two Saran wrap and 1 ml of detection reagents (500 µl reagent 1 + 500 µl reagent 2) delivered with the kit, which

should be mixed immediately before use, were added for 60 s at room temperature. The detection reagents were drained from Saran wrap, and the membrane was exposed to a film (Amersham) for 30 min or more, depending on the intensity of the signal.

ECL wash solutions

- Wash buffer 1: Urea 6 M; SDS 4 %; 0.5×SSC
- Wash buffer 2: 2×SSC : 0.3 M tri sodium citrate; 3 M NaCl

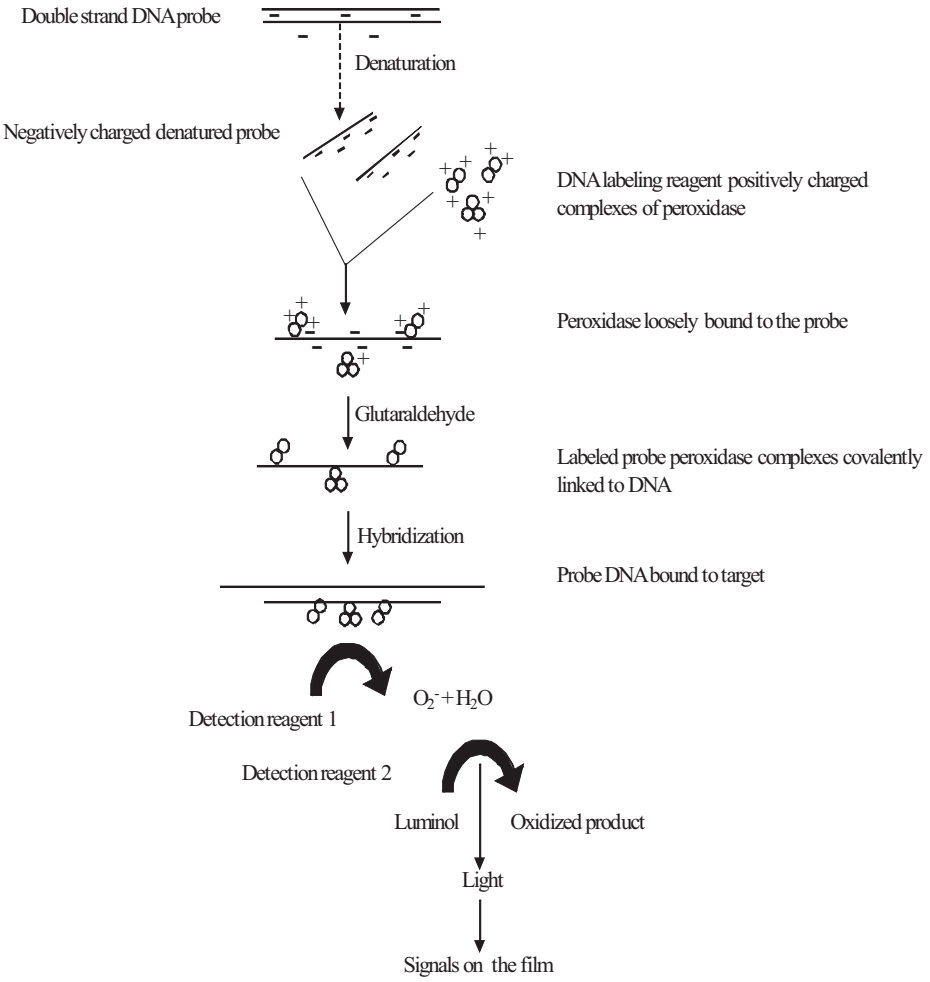


Fig. 9 Principle of the enhancer chemiluminescence direct nucleic acid labelling and detection system (GE healthcare).

2.11 Isolation of BAC-ends

The use of BAC-end sequences (sequences adjacent to the cloning site) has been proposed as a means for selecting minimally overlapping BAC-clones for sequencing large genomic regions (Venter et al. 1996). A modified plasmid end rescue method (Kelley et al. 1999) was used to obtain the BAC-ends. The method used is based on the ligation of restriction fragments of BAC-clones digested with *Bam*HI into a universal cloning vector followed by double antibiotics selection.

2.11.1 BAC-end isolation method

Ten µl of DNA from the positive BAC-clones were digested with the restriction enzyme *Bam*HI for 3 hours at 37 °C. Four µl of the digested DNA was ligated into pUC18 in 20 µl final solution of 1×ligase buffer, sterile ddH_2O, and 5 U of T4 DNA ligase (5 U/ µl) for 16 hours at 8 °C.

Prior to ligation, the vector pUC18 had been digested with the restriction enzyme *Bam*HI and dephosphorylated with calf intestinal alkaline phosphatase. In the mean time, *E. coli* (DH5α) bacteria were treated with calcium chloride method to make them competent (see 2.13) (Sambrook and Russell 2001).

For the transformation of *E. coli*, 4 µl purified ligation was added to 100 µl of competent *E. coli*, incubated on ice for 20 min, and heat shocked for exactly 60 sec at 42 °C. The tubes were immediately returned on ice for a minimum of 5 minutes. Transformed cells were incubated in 250 µl SOC (super optimal broth with catabolite repression) media for 1 hour at 37 °C. Then 100 µl were plated on LB agar medium with X-Gal (50 µg/ml), IPTG (25 µg/ml), and two antibiotics: chloramphenicol (12.5 µg/ml) for the vector pBeloBAC11 and pIndigo536 and ampicillin (50 µg/ml) for the pUC18 vector. The cultures were incubated overnight at 37 °C to allow the colonies to grow. Only the colonies containing pBeloBAC11 (or pIndigo536) and pUC18 could grow on both antibiotics. Potential recombinants (white BAC-clones) were used to inoculate 5 ml of LB medium containing two antibiotics, ampicillin (100 µg/ml) and chloramphenicol (12.5 µg/ml) and grown overnight at 37 °C.

To estimate the insert size of the BAC-end, DNA from the obtained BAC-subclone was isolated by the alkaline lysis method as described before (see 2.7.2), resuspended in 100 µl ddH_2O for sequencing or in 25 µl TE for complete digestion with *Bam*HI and separated on a 0.8% agarose gel. The gel was stained with ethidium bromide, washed in

water and photographed. Only two bands must appear, the smaller fragment, which is present in all lanes, is the pUC18 vector band and the second one is the peBeloBAC11 (or pIndigo536) vector together with the BAC-end fragment.

To check the reliability of the method, additional hybridizations using pBeloBAC11 vector as probe against the cloned BAC-end were performed. When the result was positive, then the cloned BAC-ends were sequenced for the purpose of obtaining new sequences for chromosome walking.

For the cycle sequencing, SP6 universal primer promoter sequences that flank one cloning site of pBeloBAC11 or pIndigo536 vectors were used.

2.12 Digestion and dephosphorylation of pUC18

During ligation, DNA ligase will catalyze the formation of a phosphodiester bond between adjacent nucleotides only if one nucleotide contains a 5'-phosphate group and the other a 3'-hydroxyl group. Recircularization of plasmid DNA can therefore be minimized by removing the 5'-phosphate group from both ends of the linear DNA with calf intestinal phosphatase (CIAP). A foreign DNA fragment with 5'-terminal phosphate groups can be ligated efficiently to dephosphorylated plasmid DNA to give a circular molecule.

For the enzyme digestion, in 45 µl final solution, 39 µl of the vector were added to 4.5 µl of 10×buffer, and digested with 1.5 µl of *Bam*HI (10 U/ µl) for 3 h at 37 °C.

For the phosphatase treatment, 1 µl of 10×phosphatase reaction buffer and 4 µl of calf intestinal alkaline phosphatase (1 U/µl) were added to the 45 µl of the digestion, and incubated for 60 min at 37 °C, followed by heat inactivation for 15 min at 85 °C.

The dephosphorylation was followed by a phenol-chloroform-isoamyl extraction where 50 µl of TE and 100 µl of phenol-chloroform-isoamyl were added, mixed and centrifuged at 14,000 rpm for 5 min. The upper phase was taken, and 7.1 µl of 3 M sodium acetate and 200 µl of 96% ethanol were added, the tubes were stored for 30 min at -20 °C, centrifuged at full speed (14,000 rpm) for 30 min. The supernatant was removed, the sediment dried and finally, 20 µl of TE were added, and the dephosphorylated vector was checked for one single band on 0.8% gel agarose.

2.13 Bacterial transformation

The introduction of exogenous DNA into bacterial is a process called transformation. Before successfully using *E. coli* cells for transformation, they must be made competent to efficiently take up foreign DNA. For this purpose, the calcium chloride method was used (Sambrook and Russell 2001).

A glycerol stock of the *E. coli* strain (DH5α) stored at -80 °C, was thawed and 1 µl was added to 50 ml of SOB (Super Optimal Broth) media in a 500-ml flask, and incubated overnight at 37 °C with moderate agitation (200 rpm). From the overnight culture, 50 µl were added to a 500-ml flask containing 50 ml of SOB media and incubated again at 37 °C with moderate agitation (200 rpm) until an OD_{660} of approximately 0.35 to 0.4 was achieved (3 h to 4 h). The bacterial cells were transferred into ice-cold 50 ml sterile tubes, and chilled on ice for 10 min. The cells were sediment by centrifugation at 4,100 rpm for 10 min at 4 °C and drained thoroughly. Each pellet was resuspended by swirling in 30 ml ice-cold solution containing 80 mM $MgCl_2$ and 20 mM $CaCl_2$. The cells were recovered by centrifugation as before (4,100 rpm, 10 min, and 4 °C). The pellet was resuspended in 2 ml of ice-cold 0.1 M $CaCl_2$ solution (for each 50 ml of original culture), and 20% glycerol was added for a long term conservation. The competent cell solution was divided into 100 µl-aliquots, and was stored frozen at -80 °C until needed.

Transformation media

- SOB: 2% w/v Trypton (20 g); 0.5% w/v yeast extract (5 g), 10 mM NaCl (0.584 g); 2.5 mM KCl (0.186 g); ddH_2O to 1 l; pH 7.0
- SOC: 2% w/v Trypton (20 g); 0.5% w/v yeast extract (5 g), 10 mM NaCl (0.584 g); 2.5 mM KCl (0.186 g); 10 mM $MgCl_2$ (0.952 g); 20 mM glucose (3.603 g); ddH_2O to 1 l; pH 7.0

2.14 Sequencing of BAC-ends

The cloned BAC-ends were isolated using standard alkaline lysis miniprep techniques as described in 2.7.2. Sequencing reactions were set up according to manufacturer's instructions for the Big Dye terminator chemistry (Applied Biosystems). The sequences were performed first using universal primer SP6. Two independently cloned BAC-ends were always compared for the reliability of the sequences.

The alignment was done by BioEdit sequence alignment editor software (www.mbio.ncsu.edu/BioEdit/BioEdit.html). New primers for sequencing were designed using the programme web primer designer (http://seq.yeastgenome. org/cgi-bin/web-primer).

2.15 Contig assembly by fingerprints and hybridisations

Overgo probes were constructed from the end sequence to rescreen the BAC libraries. *Hind*III was used to digest a BAC DNA and the fragments were fractioned on 0.8% agarose gel. The fragment size of each set of fingerprint data was used to assemble the contig maps. For the chromosome walking, Southern hybridisations using BAC-end overgo as probe were performed to look for overlapping between the contigs.

3. Results

3.1 Chromosome walking at the restorer locus *Rf1*

3.1.1 Hybridizations using the markers OP-K13-454A and E33M61-136R

The hybridization of the overgos derived from markers OP-K13-454A and E33M61-136R against RHA325 BAC library (Özdemir et al. 2004) resulted in five positive BAC-clones. BAC fingerprinting using *Hind*III restriction enzymes as well as hybridization were performed (Fig. 10).

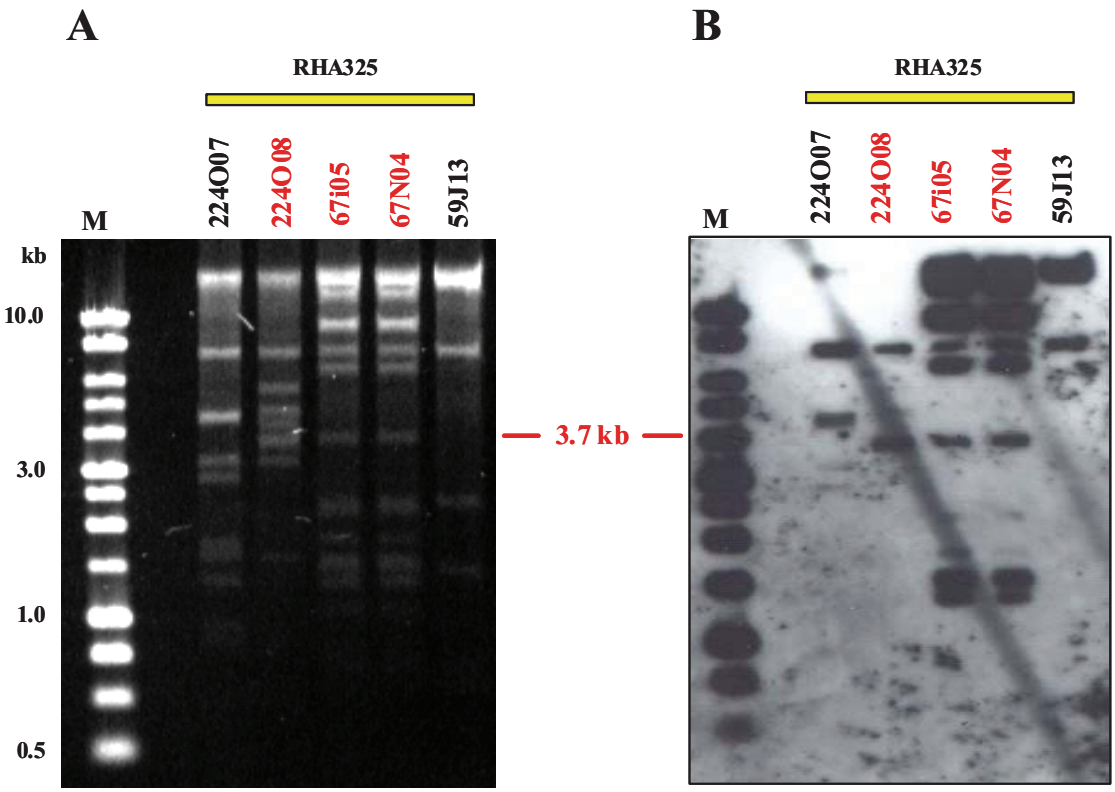


Fig. 10 Characterization of the BAC-clones identified by overgos derived from OP-K13-454A and E33M61-136R in the BAC library RHA325. (A) BAC-fingerprinting of the five identified positive BAC-clones. (B) Hybridization using *Hind*III digested BAC-clone 67N04 as probe.

From one side the BAC-clones 67i05 and 67N04 showed an identical banding pattern which overlapped with the smaller BAC-clone 59J13. From the other side, the BAC-clones 224O08 and 224O07 also shared a *Hind*III fragment. The banding pattern was confirmed by Southern hybridization using the *Hind*III digested BAC-clone 67N04 as probe. A 3.7-kb-fragment shared between the BAC-clones 224O08, 67N04 and 67i05

was identified. This result allowed the assumption of a preliminary putative contig around the restorer gene *Rf*1 (Fig. 11).

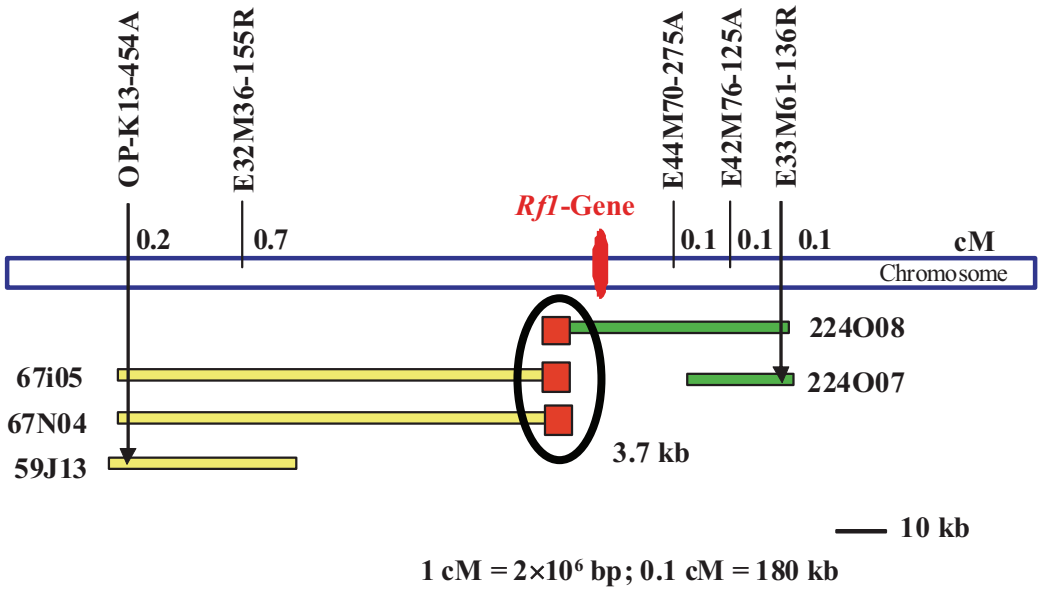


Fig. 11 Preliminary putative contig around the restorer gene *Rf*1.

Along chromosomes (Mb/cM) ratio was calculated on the basis of a DNA content of 3,000 Mb haploid sunflower genome (Arumuganathan et al. 1991). Considering the entire genetic length of the sunflower map of 1,645 cM in 18 linkage groups (Kusterer 2003), an average value of 1.8 Mb/cM (~ 180 kb/0.1 cM) can be assumed.

3.2 Verification of the putative closed contig

The presumed putative closed contig around the *Rf*1 gene of the PET1 cytoplasm needed to be verified. The 3.7 kb-fragments from the BAC-clones 67N04 and 224O08 were cloned. Four restriction enzymes *Hind*III, *Pst*I, *Eco*RI and *Kpn*I were used to verify the identity of the putative overlapping 3.7-kb-fragment (Fig. 12). Surprisingly, the restriction patterns obtained with *Pst*I and *Eco*RI were not identical. This indicated that we cloned two different fragments of the same size. Therefore we do not have closed contig around the *Rf1* gene.

In addition, the 3.7-kb-fragment from the 67N04 BAC-clone was sequenced. The analysis showed homology to retrotransposons which might explain the observed cross hybridization between the 3.7 kb fragments of the two BAC-clones 67N04 and 224O08. It has been reported that overlaps can be detected simply by hybridization but it is not a satisfactory criterion because dispersed repeats can generate false-positives (Hong 1997). Therefore, to obtain reliable fingerprints, digestions with several restriction enzymes should be part of the whole process.

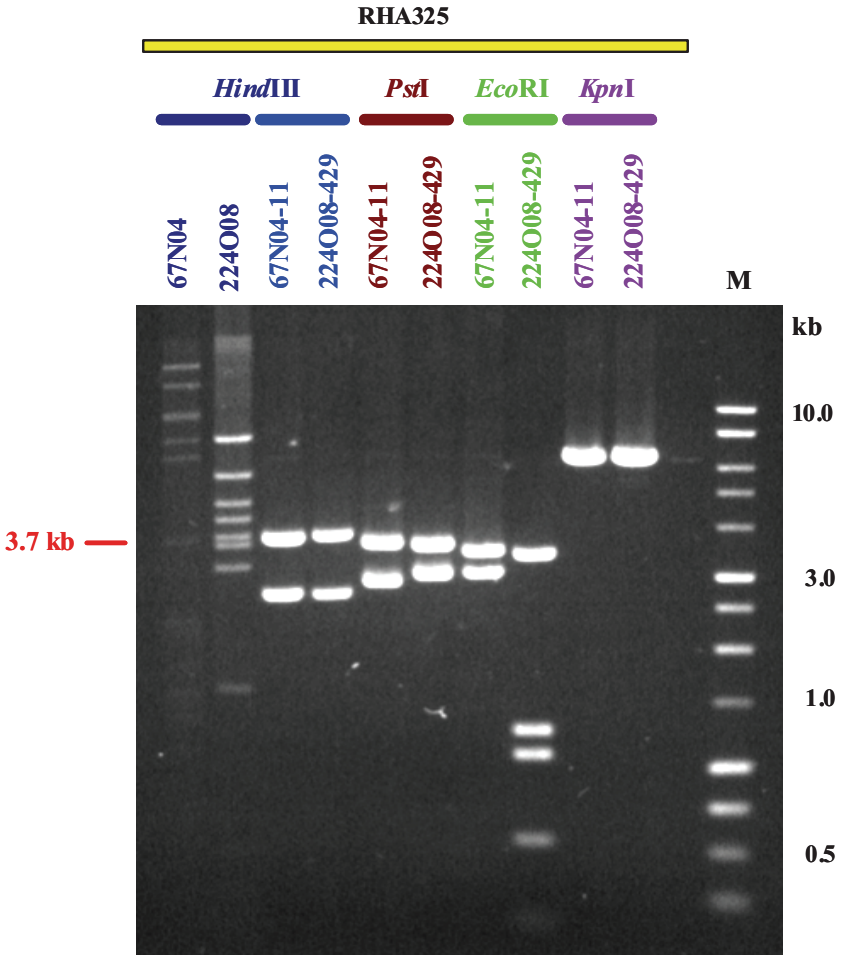


Fig. 12 Investigation of the putative closed contig. The cloned 3.7-kb-fragments from 67N04 and 224O08 BAC-clones were digested with four restriction enzymes.

3.3 Identification of new molecular markers linked to the *Rf1* gene

The linkage map of the sunflower constructed by Kusterer et al. (2005) was used as reference starting point to further saturate the *Rf1* locus with additional more closely linked markers.

For this purpose, a parallel work was performed (Horn 2008, personal communication), in which 256 AFLP primer combinations were used in bulked segregant analysis (Michelmore et al. 1991). Among the 256 *Eco*RI/*Mse*I primer combinations, 73 were polymorph between the bulks of homozygous fertile F_2-plants and homozygous male sterile plants and generated a total of 85 potential markers. Thirty-nine polymorphic fragments were present in the male sterile bulks while 46 polymorphic fragments were present in the fertile bulks. In order to validate the result of the BSA, a second round of PCR analyses with 55 primer combinations was carried out with the DNA of the individual plants constituting the bulks, the parental DNA. Those primer combinations that did not show any recombinants were then mapped in 91 individuals of F_2-population segregating for *Rf1* gene.

Only four markers mapped closely linked to the *Rf1* gene. The new markers E56M54-235A, E54M55-428A, E55M59-370R were located on one side of *Rf1* at distances of 3.1, 2.3 and 1.6 cM, respectively, while the marker E62M52-249A was mapped on the other side of *Rf1* at a distance of 0.1 cM (Fig. 13). The total genetic distance for the 13 markers surrounding the gene is about 9.7 cM with 0.74 cM/marker within the range of 3.5 cM. A greater marker density in the vicinity of the *Rf1* might be required for a map-based cloning approach (Deng et al. 2001). The markers OP-K13-454A and E32M36-155R were previously separated and now cluster and cosegregate with the restorer gene *Rf1* in the new map. Probably, the small mapping population used accounts for this discrepancy (Yin et al. 2006).

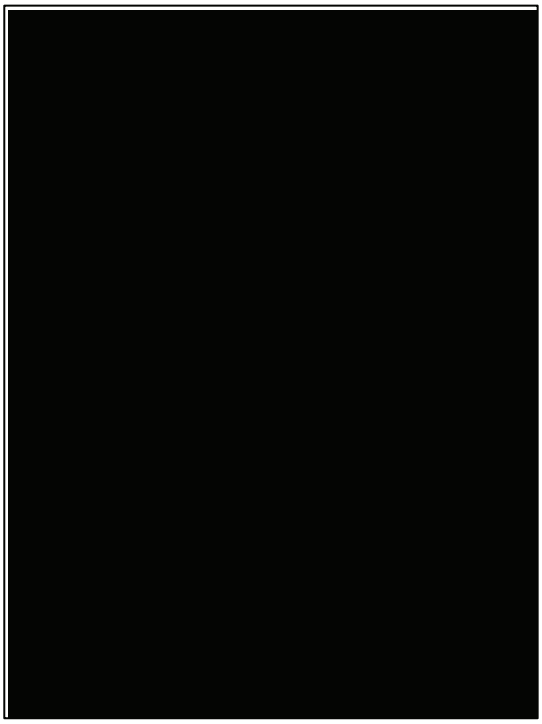


Fig. 13 Markers surrounding the *Rf1* gene (Horn 2008, personal communication). The linkage order between the *Rf1* and 13 DNA markers is shown on the right; genetic distances in centiMorgan (cM) are shown on the left. The four new AFLP markers are presented in bold. The total genetic distance of the region shown around the *Rf1* is about 9.7 cM.

3.4 Screening two sunflower BAC libraries

Screening of BAC libraries is possible by either a PCR-based method or a hybridization-based method (Han et al. 2000). The PCR-based screening requires pooling of BAC-clones involving combination of plates, rows, and columns (Bruno et al. 1995).

In our study, we used the strategy of colony hybridization against high density filters of two sunflower BAC libraries (RHA325 and HA383) using radioactively labelled probes. The two libraries combined contain 327,488 BAC-clones and are equivalent to approximately 10.2-fold haploid genomes of sunflower. Therefore, it was expected that if both libraries together were sufficiently representative, some of the BAC-clones

identified by the DNA markers would be close to the gene. The contig to be constructed will be a valuable resource, providing a dense network of hybridization probes for the isolation of overlapping BAC-clones from the restorer BAC library RHA325.

3.4.1 Colony hybridization against sunflower BAC libraries

Seventeen high-density BAC filters were screened, six from the restorer BAC library RHA325 and 11 from the maintainer BAC library HA383 for candidate BAC-clones potentially close or containing the *Rf1* gene using the most linked markers as overgo probes in hybridizations (Fig. 14**)** .

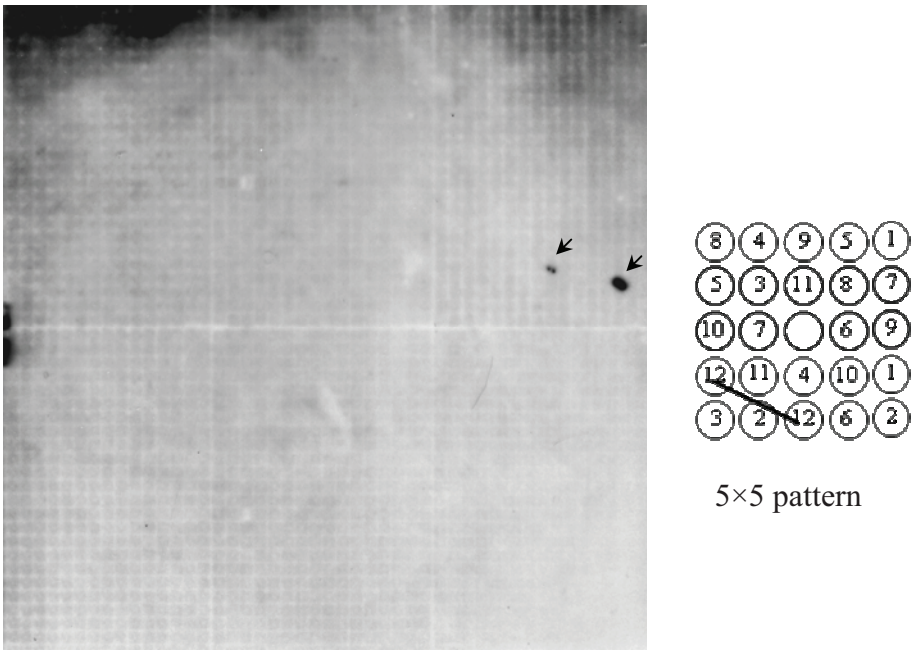


Fig. 14 Filter image of hybridization of BAC library RHA325 using the overgo derived from the marker OP-K13-454A radioactively labelled as probe. Colonies were double spotted in high density on a 22.5 × 22.5 cm filter in a 5 × 5 pattern. Each filter contains 27,648 BAC-clones. Positive BAC-clones are indicated by arrows. The BAC-clones represent position 12 in the 5×5 spotting pattern.

One of the major advantages of these probes is that the sequences can be screened by BLAST (Basic Local Alignment Search Tool) search (Altschul et al. 1990) to avoid repeated sequences present with higher probability in a longer conventional DNA fragment probe and to minimize cross hybridization problems associated with large-genome DNA library screening (Feng et al. 2006).

During this study, twenty overgo probes have been designed from three sources including AFLP markers, RAPD markers and BAC-end sequences (Tab. 4). The overgos 216F17-B-end and 401E15-B-end were used as control probes (see 3.8.1)

Tab. 4 Overview of the overgos generated in this study from AFLP markers[1], RAPD markers[2], and BAC-end sequences[3].

Overgos		Profile length bp	Forward overgo sequence	Reverse overgo sequence
1	E32M36 155R	161	ACCAAACTGGTTATTCCATTTTGG	CCATTTCGAGTAGTAACCAAAATG
	E42M76 125A	125	CTCCATACACTCCGCAGCAGAATG	TTCCTCTCTGCTACTCCATTCTGC
	E33M61 136R	155	GAGAGCGAAAGATCAAGACTAAAG	ATCTTCTGTCTTCAGTCTTTAGTC
	E41M48 113A	113	GTTCCACTTTCTGTCAACGAGGAA	CAGCTACGAAAGCTAGTTCCTCGT
	E44M70 275A	275	GGATTCATCCCTAGAAGTGCAAGAT	CTTCCTTGAACACCTTCAATCTTGCA
	E62M52 249A	249	GTGGGGTATGGAAAAAACAACTTGA	GTGTCGTCTTCCTTGCATTCAAGTTG
2	OP K13 454A	454	CCCTTCCATTAAAACCCAGCATCA	GTGTACATGTCGAAAGTGATGCTG
	OP H13 337A	337	TCTGTGTTCAATTGAGAATTCGCA	CCAATTTCATGGGTTGACGCTTAA
	OP Y10 740A	740	CGTGGGAGAGAGGTGGCCAACCCC	AAAATGTCTCCTAATTCCCCAACC
3	67N04 B end	650	ACGATGCCTTTTTCTTCTTTGAGG	GTAACTGGAGATGAGCCCTCAAAG
	225D09 B end	535	TACCTCCATAATATATCGCCGCAT	CCATTACAGCAGTGGTCATGCGGCG
	225D09 sf2 end	756	ATAGCTCTATACGCAGCGTATATAA	GGCACAGACAACCAGAGTTTATATAC
	216F17 B end	991	TGGAATTTAGGGACTCTGAAGACA	CACGTAACTCCGGTCTTGTCTTCA
	224O07 B end	882	TTGTAAGAGATAGATCAGGGCACG	CTGTTTCCTCGATTTCCGTGCCCT
	401E15 B end	1,522	TGCTCGCCTATAAGTGGACACTCA	CTGGGAACTCTTGGGGTGAGTGTC
	126N19 B end	3,486	GTTTCGACAGCCGTACGCTGTGG	TAGTGTGTCTTATTGCCACAGCG
	261D17 B end	2,128	TCATACTTCTTATCCTTTTGAGTAA	CTCAACCGCTGACCTCCCTTACTCAA
	261F19 B end	3,953	CATGTTCACCCCTACAATTTTCAG	CCGAACCGAGCTTAAACTGAAAAT
	450B6 B end	576	AAGGGTTGATTGCCGGAAAAATTGGT	CTTCGTTAGCATCGGATTACCAATTT
	480G4 B end	369	ATGCAAGCTAATCGATAAGGATTTG	CGAAGCCCACAAAGCAACCAAATCCT

3.4.2 Overgo characterization

There was a wide fluctuation in the number of BAC-clone hits for the used overgos (Tab. 5). Four overgos (two constructed from RAPD markers and two constructed from BAC-ends) failed to hit any BAC-clones. This could be due to several reasons: (i) overgos were derived from regions with deletions (ii) poor overgo design and/or failed radiolabelling, and (iii) overlapping segments of the sunflower genome are not present in the two BAC libraries used in this study (Cao et al. 1999; Hohmann et al. 2003). Five overgo probes were repetitive: three of them were derived from AFLP markers. This can be explained by the repetitive nature of AFLPs. Two of the repetitive overgo probes were derived from BAC-ends. Other overgo, like e.g. E62M52-249A, was positive with the maintainer BAC library but not with the restorer BAC library. In contrast E32M36-155R hit two positive BAC-clones in the RHA325 BAC library but did not hit any BAC-clone in the HA383 BAC library. This could be due probably to the different representation of fractions of a genome (Nakamura et al. 1997). The overgo 225D09-B-sf_2 of 653 bp of the sequence of the BAC-end 225D09 has a 81% homology at nucleotide level with 2.00E ^{91} (E value) to *sf_2*, which encodes an anther specific protein in sunflower (Domon et al. 1990). The overgo designed from this sequence was repetitive in the restorer BAC library but only hit two BAC-clones in the maintainer BAC library. The *sf_2* was described to hybridize exclusively to RNA from anthers in the late developmental stage, and highly expressed in male-fertile plants (Domon et al. 1990; Gentzbittel et al. 1999). The two overgos E33M61-136R and E42M76-125A were repetitive. Comparing their filters, three common BAC-clones were found among the huge number of signals. Similarly, one common BAC-clone was found by comparing the filters of E42M76-125A with those of E44M70-275A.

The low hit rate of the BAC library RHA325 confirms that a higher library size of a genome equivalent than 1.9 is necessary to ensure coverage of the region of *Rf1* gene.

Tab. 5 Overview of screening two sunflower BAC libraries (RHA325 and HA383) with 18 overgo probes designed from AFLP and RAPD markers linked to the *Rf1* gene and the BAC-end sequences.

Probe Markers	BAC-Filters		Probe BAC-ends	BAC-Filters	
	RHA325	HA383		RHA325	HA383
E32M36-155R	2	No hit	67N04-B-end	2	11
E33M61-136R	2	Repetitive	225D09-B-end	1	5
E42M76-125A	3	Repetitive	224O07-B-end	No hit	17
E44M70-275A	Not done	1	126N19-B-end	No hit	
E41M48-113A	1	4	261D17-B-end	Repetitive	
E62M52-249A	No hit	3	261F19-B-end	1	9
OP-K13-454A	3	4	480G04-B-end	No hit	
OP-H13-337A	No hit		450B06-B-end	Repetitive	
OP-Y10-720A	No hit		225D09-B-sf2	Repetitive	2
Hit/non-repetitive overgo				1.1	4.0

3.5 Number of BAC-clones obtained

A total of 97 positive BAC-clones were identified (data not shown) by colony hybridization and 71 of these were confirmed by fingerprint hybridization, which corresponds to a rate of 73% of positive BACs identified. The percentage of false-positives was about 27%. The positive BAC-clones identified by the marker and BAC-end overgo probes are illustrated in Table 6. The average number of BACs that were positive for each non-repetitive overgo probe ranged from one to 17 BACs. Only 23 BACs were identified with marker-derived overgo probes. The BAC library HA383, which represents 8.3×genome equivalents, is expected to have an average of eight positive BAC-clones per probe. The average number of hits per overgo (4.0) was only half of the expected number. Thus, the number of positive BAC-clones obtained per probe does not reflect the number of genome equivalents.

For RHA325, the number of hits per overgo was 1.1, which is also half of the expected number of positive BAC-clones (1.9-fold genome equivalents) (Özdemir et al. 2004). In total, we estimated that about 40% of our overgo probes failed. The low hit rate confirms the difficulties in ensuring genome coverage of the region of interest and to be able to construct closed contigs.

Sizes of the isolated BAC-clones were analysed in two ways: (i) DNA fingerprinting using *Hind*III restriction enzyme, and (ii) Pulse Field Gel Electrophoresis (PFGE) using *Not*I restriction endonuclease.

Tab. 6 Positive BAC-clones identified in the two BAC libraries (RHA325 and HA383).

Probes	Library	BACs	Probes	Library	BACs	Probes	Library	BACs
225D09-B-end	RHA325	115P09	67N04-B-end	RHA325	139A17	224O07-B-end	HA383	36B04
	HA383	175L13			179K02			43K09
		384H08		HA383	178G06			56D10
		231M03			177K11			105N18
		466H15			180A24			189N06
		194A21			486L01			196G02
261F19-B-end	RHA325	39P12			153F08			240A03
	HA383	131C20			467E16			248J09
		158M19			312E12			295B10
		497K04			177M08			303D18
		390M14			89P04			340L17
		482H17			60K10			370M06
		278B16			192D15			407M09
		269i11	225D09-*sf2*	HA383	35i08			432E15
		524G22			454H6			464H13
		513E21	E32M36-155R	RHA325	16M16			485H12
OP-K13-454A	RHA325	67N04			16P13			490A04
		59J13	E41M48-113A	RHA325	94F15	E62M52-249A	HA383	6N12
		67i05		HA383	100L22			447N06
	HA383	216F17			147A03			480G04
		307N02			233O05	E42M76-125A	RHA325	126N19
		225D09			482D10			261D17
		401E15	E44M70-275A	HA383	450B06			261F19
E33M61-136R	RHA325	224O07						
		224O08						

3.5.1 BAC-fingerprinting of positive BAC-clones

Positive BAC-clones were first subjected to fingerprinting (Fig. 15). Purified BAC-clone DNA was digested with *Hind*III restriction enzyme. Fragments were separated on 0.8% agarose gels, and the fingerprints were analysed with IMAGE Quantity One software based on the migration distances of the restriction fragments. The fragment

number per BAC-clone ranged from 1 to 18 for RHA325 (average 9) and from 10 to 25 for HA383 (average 18).

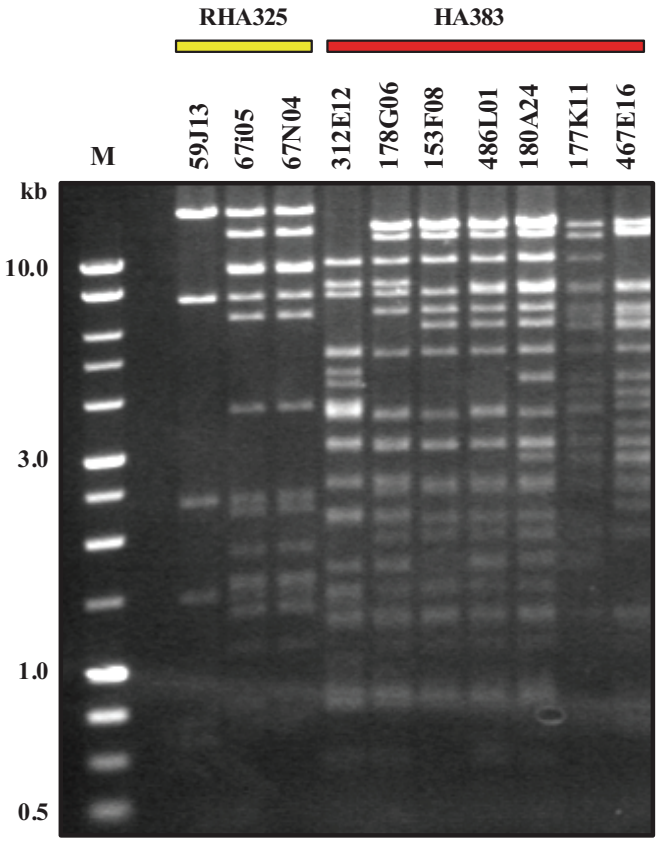


Fig. 15 DNA fingerprints of 10 positive BAC-clones. BAC-clones digested with *Hind*III and separated on a 0.8% agarose gel, 80 V for 65 min in 0.5×TAE buffer. The insert sizes of the BAC-clones shown on this gel ranged from 8 kb to 67 kb for RHA325 and 54 kb to 87 kb for HA383.

The average insert size of the positive BAC-clones was 32 kb for the BAC library RHA325 ranging from 5 kb to 68 kb, and 74 kb for the BAC library HA383 ranging from 39 kb to 101 as determined by BAC-fingerprinting. The majority of positive BAC-clones carried inserts of less than 20 kb for RHA325 and 60-80 kb for HA383 (Fig. 16). The average insert sizes obtained were smaller than the estimated average insert size of RHA325 (60 kb) and HA383 (125 kb) BAC libraries.

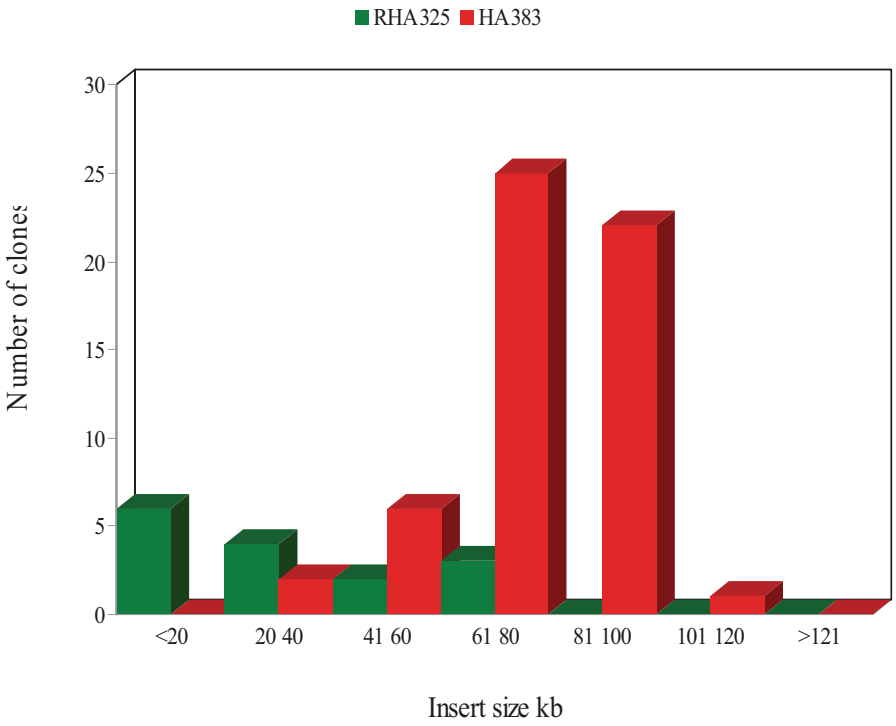


Fig. 16 Distribution of insert size of the 71 positive BAC-clones in the two sunflower BAC libraries (RHA325 and HA383). DNA samples of 15 BAC-clones from BAC library RHA325 (green columns) and 56 BAC-clones from BAC library HA383 (red columns) were analysed. The insert sizes were determined by fingerprinting using Quantity One software. The major peak corresponds to insert lengths of 61-80 kb for HA383.

3.5.2 Insert size of BAC-clones using pulse field gel electrophoresis

Fifty BAC-clones were analyzed by *Not*I digestion and pulse field gel electrophoresis. The insert sizes of the BAC-clones of HA383 ranged from 91 to 157 kb with an average insert size of 115 kb, which is still smaller than the expected one (125 kb) but was larger than the average insert size of fingerprinting method. This is probably due to the fact that several fragments have been overlooked in the fingerprint sizing. It was difficult to separate two or even more fragments which differed in a few base pairs. In contrast, the average insert size in RHA325 did not differ between the two methods because the BAC-clones were smaller and had fewer fragments, which did not hinder the fingerprint estimation. The number of *Not*I sites in the inserts of the BAC-clones ranged from null to three, most BAC-clones had no *Not*I site in their inserts, as

indicated by a single insert band on pulse field gel electrophoresis (Fig. 17). However, few BAC-clones (7%) contained one to two internal *Not*I restriction sites in their inserts, yielding two to three bands upon *NotI* digestion (data not shown).

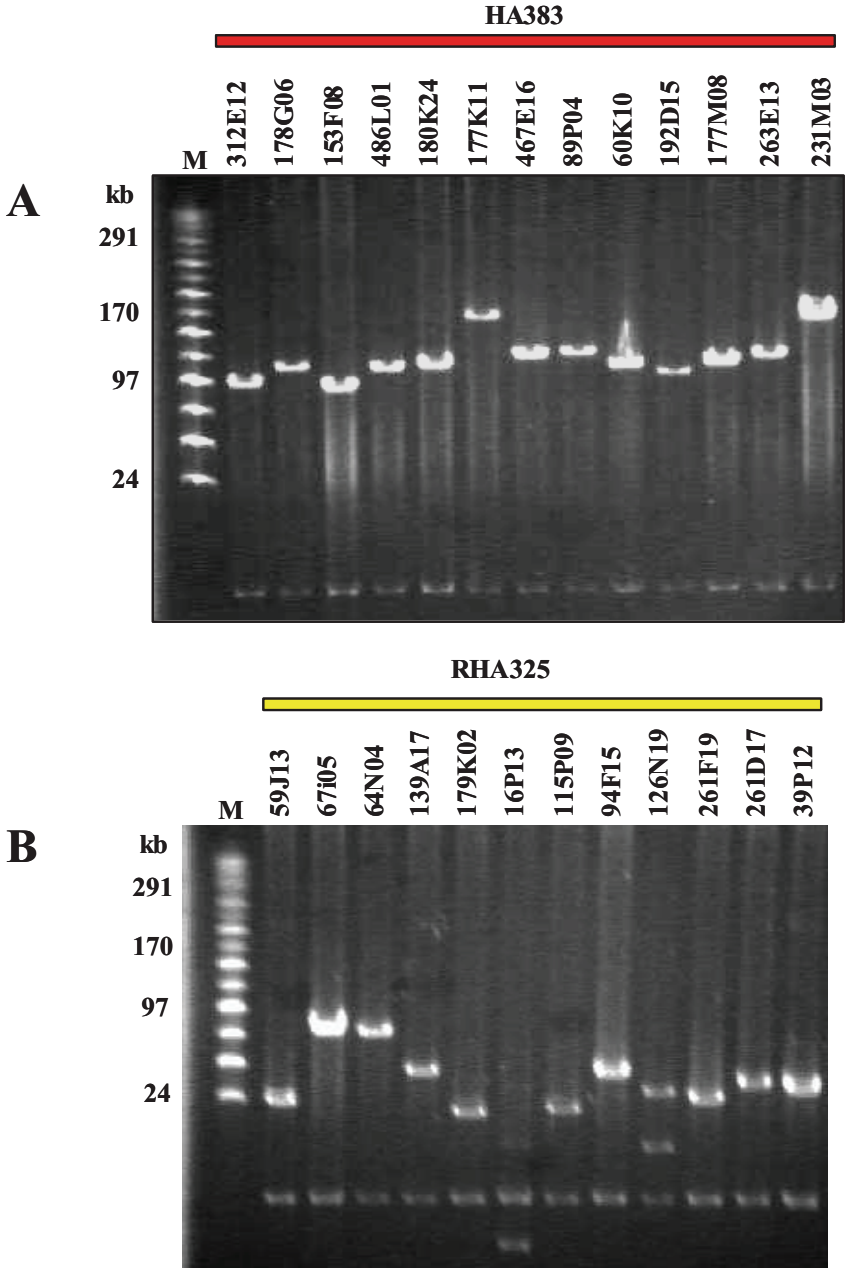


Fig. 17 DNA analyses of BAC-clones by pulse field gel electrophoresis. DNA samples were digested with *Not*I and separated on an 1% agarose gel with ramp pulse of 1-25 s at 6 V/cm at 14°C in 0.5×TBE buffer for 24 h. M: Marker MidRange II PFG (*BioLabs*), and resolved by IMAGE Quantity software. (A) In HA383, the common band is the vector pIndicoBac536. The insert sizes of the 13 BAC-clones shown on this gel ranged from 91 to 157 kb. (B) In RHA325, the common band is the vector pBeloBAC11. The insert sizes of the 12 BAC-clones ranged from 5 to 73 kb.

The insert size distribution of 50 BAC-clones (13 from RHA325 and 37 from HA383) is illustrated in Figure 18. This analysis showed that 30 BAC-clones (60 %) have an insert greater than 100 kb. Pulse field gel electrophoresis is more precise because *Not*I releases the whole insert from the BAC vector. As result, the real insert size of the fragment is obtained especially for the insert size higher than 100 kb.

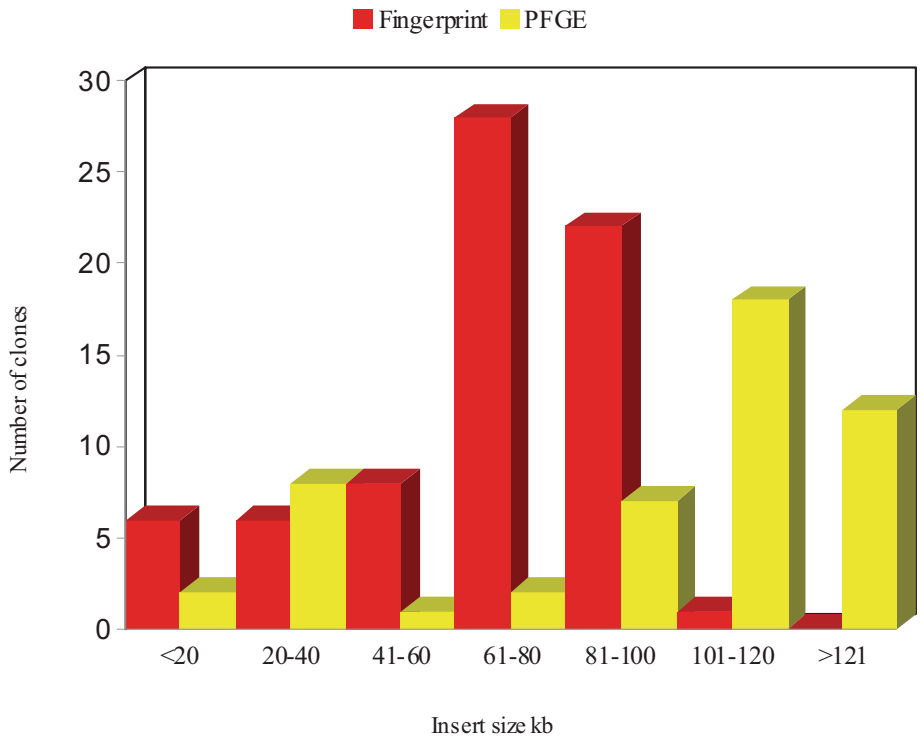


Fig. 18 Comparison of insert size distribution of positive BAC-clones in the two sunflower BAC libraries (RHA325 and HA383). The insert sizes were determined by fingerprint (red columns) and pulse field gel electrophoresis (yellow columns) and resolved by IMAGE Quantity One software. The major peak corresponds to insert lengths of 101-120 kb from the pulse field gel electrophoresis.

3.6 Isolation of BAC-ends

3.6.1 Method used to isolate BAC-ends

The objective of this study was to conduct a map-based cloning strategy. This involves screening of BAC libraries with cloned markers and assembly of contigs by chromosome walking. Chromosome walking requires the isolation of BAC-end-fragments to generate primer probes for re-screening the BAC library for further steps.

Figure 19 indicates the strategy followed to isolate BAC-ends. The restriction enzyme *Bam*HI that cuts several times within the insert and once within the BAC vector was used. The resulting fragments were ligated into pUC18 vector and used to transform *E. coli*.

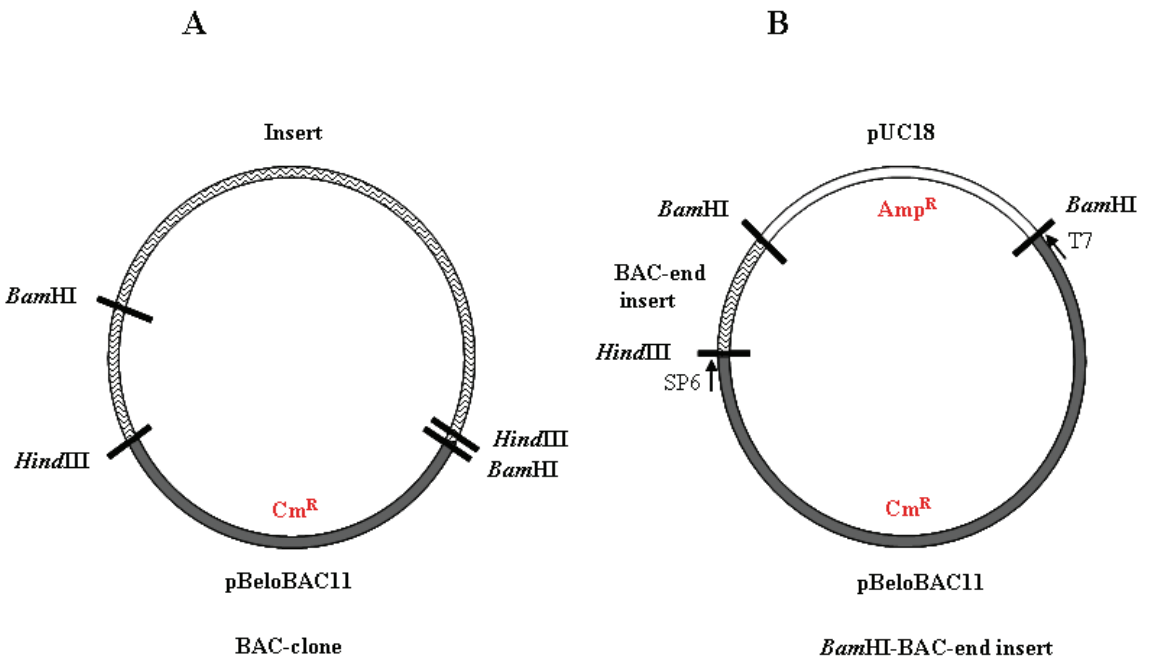


Fig. 19 Schematic diagram for cloning BAC-ends. (A) BAC-vector with cloned *Hind*III fragment. (B) Cloned *Bam*HI-BAC-end, which consists of pBeloBAC11 (RHA325) or pIndigoBac536 (HA383) vector and the most extreme end of the cloned *Hind*III fragment, which does not have any further internal restriction enzyme site for *Bam*HI (Hamrit et al. 2008).

By using two antibiotics: ampicillin resistance for the pUC18 vector and chloramphenicol resistance for pBeloBAC11 or pIndigo536 vectors, selected white BAC-clones should only contain the BAC vector, pBeloBAC11 or pIndigo536, and the most extreme end of the insert which does not have any further internal *Bam*HI restriction sites (Fig. 20).

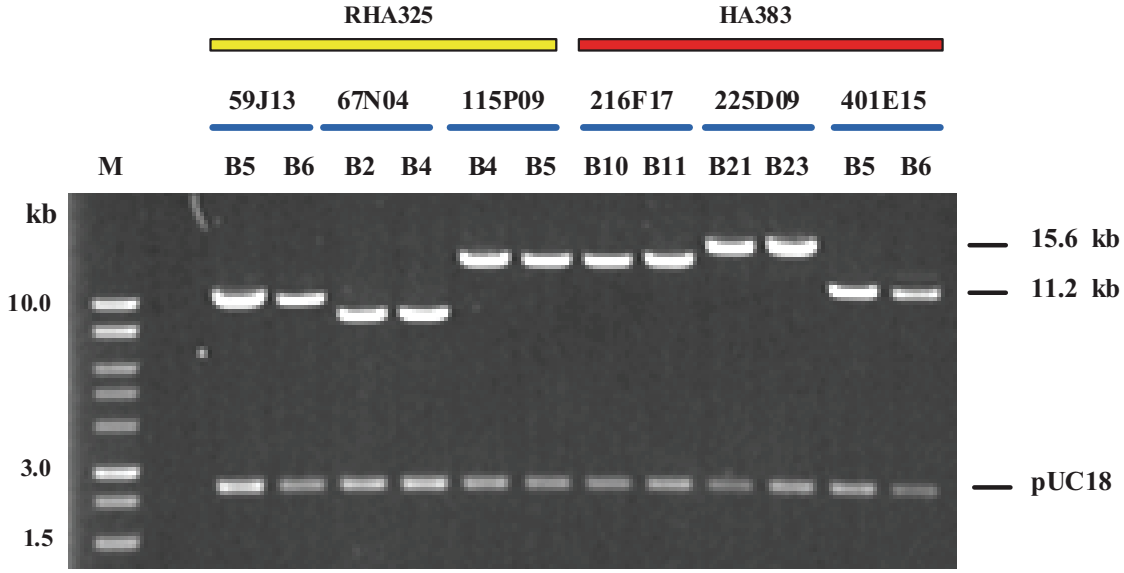


Fig. 20 Isolation of *Bam*HI-BAC-ends ligated into pUC18 vector and digested with *Bam*HI restriction enzyme from six different BAC-clones. The fragments were separated on a 0.8% agarose gel in 0.5×TAE buffer.

3.6.2 Characterization of the BAC-ends

Using this method, 27 BAC-ends have been isolated out of 29 positive BAC-clones in order to develop overgo probes (Tab. 7). The failure of cloning the two remaining BAC-ends could be due to the restriction fragments generated by *Bam*HI digestion that were too large for an efficient ligation or too small to be detected. Knowing the sequence of the vector, the BAC-end sequence could be easily separated from the vector sequence for further overgo probe design. The size of the BAC-ends ranged from 1.2 kb to 15.2 kb. To design new overgo probes from the BAC-ends to further screen the BAC libraries, SP6 promoter sequences that flank the multiple cloning site of pBeloBAC11 and pIndigo536 were used for sequencing. A total of 48 pairs primers (Tab. 9) have been progressively designed generating in total 57 kb sequences of high quality (including 35 kb forward and 22 kb reverse sequences for the check) with a success rate of 99%. High-quality sequences were defined as those having more than 75% high-quality bases other than vector and *E. coli* sequences (Budiman et al. 2000).

Tab. 7 Overview over the cloned BAC-ends. Each BAC-end was characterised by size (without the vector). Sequences (forward and reverse) from each BAC-end have been used to design new overgos to re-screen both BAC libraries.

BAC-clone	Size kb	BAC-end size kb	Sequence size bp		BAC-clone	Size kb	BAC end size kb	Sequence size bp	
			For	Rev				For	Rev
16M16	5	3,1	492	-	115P09	19	6,0	542	-
16P13	5	3,6	-	-	175L13	132	8,6	596	-
139A17	44	5,4	1,376	1,011	384H08	106	1,2	528	-
179K02	16	6,6	-	-	231M03	145	2,1	544	-
224O07	23	3,1	1,721	1,732	194A21	95	5,2	655	-
224O08	28	4,6	2,056	1,285	94F15*	41	3,7	1,905	1,555
126N19	34	8,5	3,486	3,354	100L22	95	5,8	592	515
466H15	69	-	-	-	147A03	119	2,7	1,090	630
261D17	61	0,4	350	-	233O05	98	3.5	-	-
		2,3	2,128	1,026	482D10	103	-	-	-
261F19	21	5,5	3,953	3,247	6N12	107	15,2	614	-
67N04*	73	1,5	1,509	1,275	447N06	135	15,2	1,042	669
59J13	21	3,0	-	-	480G04	135	15,2	369	-
216F17	148	6,7	1,482	712	450B06	126	15,2	576	-
401E15	93	3,7	1,522	1,161	225D09	121	8,1	5,138	3,347

3.6.3 Verification of cloned BAC-ends by hybridization

In addition, to further verify the cloned *Bam*HI-BAC-ends, the BAC-ends were released from the pUC18 vector by complete digestion with *Bam*HI enzyme (Fig. 21A), and blotted onto membranes. Southern hybridizations using the vector pBeloBAC11 as probe against the BAC-ends were performed. The results confirmed that obtained sub-clones contained one band representing the pBeloBAC11 (or pIndigo536) and the BAC-end fragment (Fig. 21B). The obtained fragments had a larger size than the size of the vectors (7.5 kb). However, using this method we were only able to clone the right band side of the original BAC insert. Other Southern hybridizations were performed only

with BAC-ends that were further used to design new overgos for chromosome walking purpose. Results for the other cloned *Bam*HI-BAC-ends are shown in the Figures 22, 23, 24, and 25.

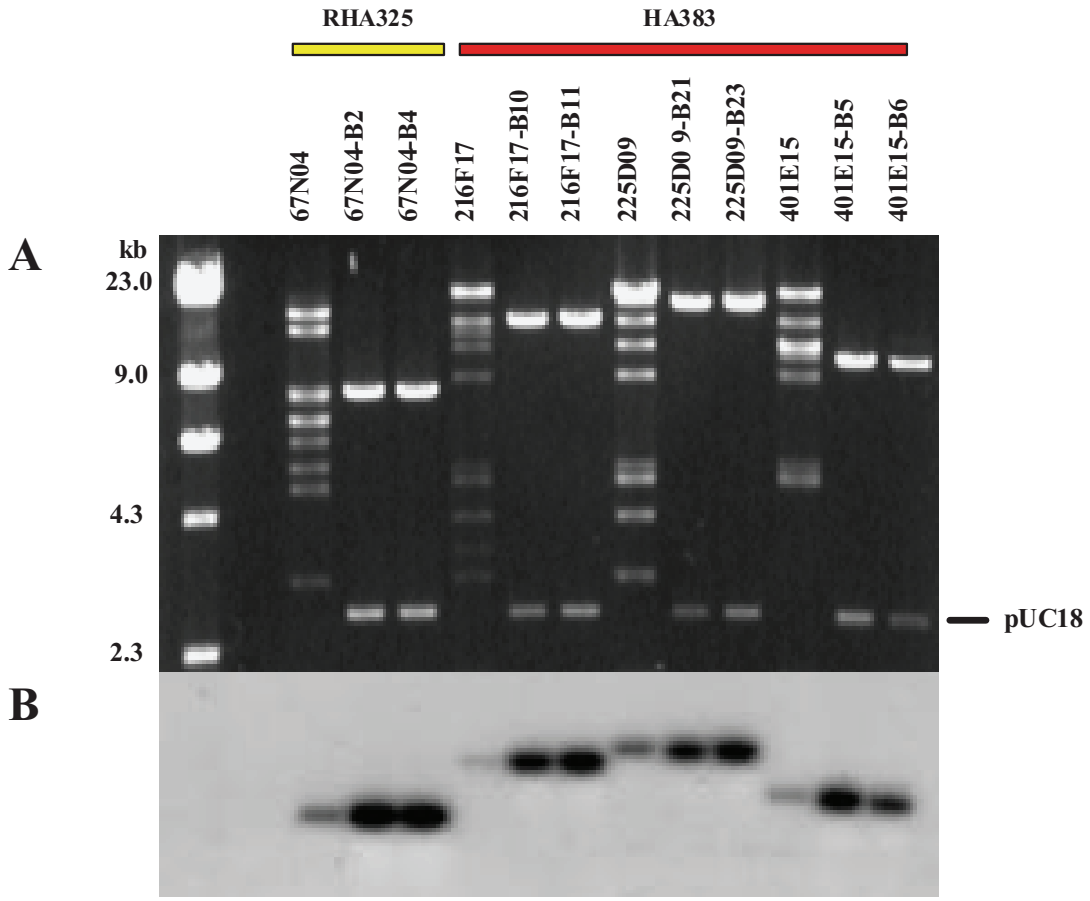


Fig. 21 Verification of cloned BAC-ends of the positive BAC-clones 67N04, 216F17, 225D09, and 401E15, which have been identified using OP-K13-454A as overgo probe. (A) The cloned *Bam*HI-BAC-ends were released from the pUC18 by complete digestion with *Bam*HI enzyme in comparison to the original BAC-clone *Bam*HI digested. Each obtained BAC-end fragment has a larger size than pBeloBAC11 (7.5 kb). (B) Hybridization using pBeloBAC11 vector as probe against the positive BAC-clones and their cloned BAC-ends (minimum two each).

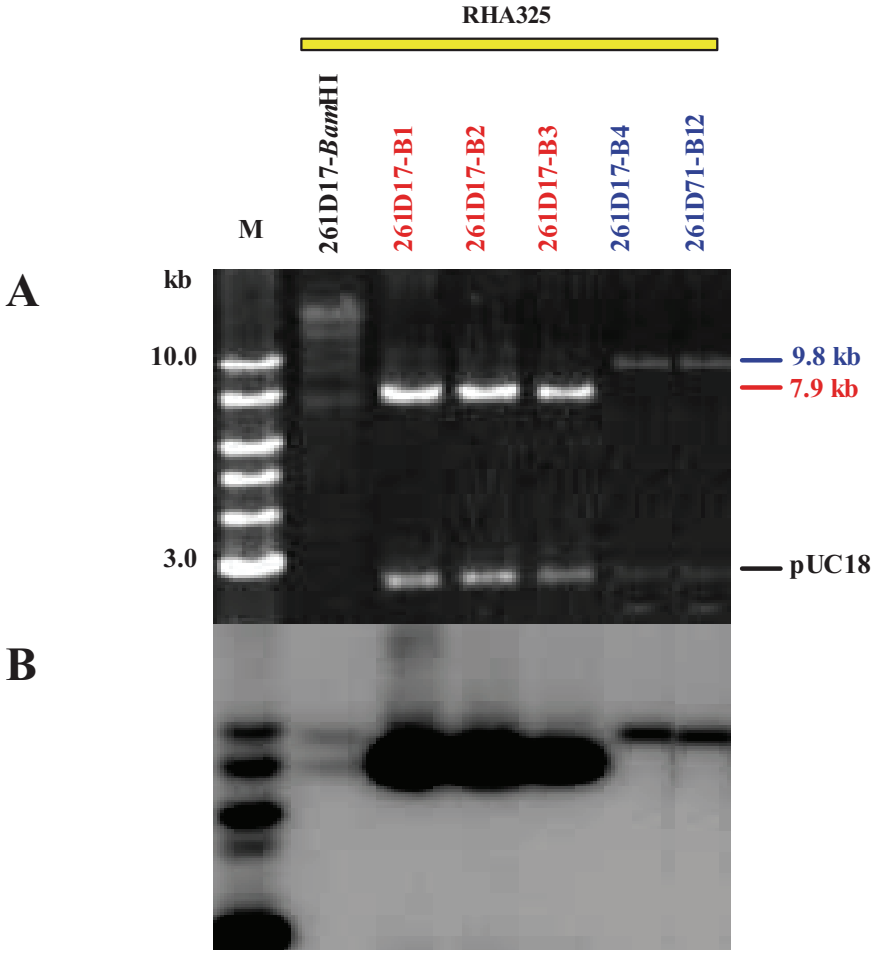


Fig. 22 Verification of cloned BAC-ends of the common positive BAC-clone 261D17 obtained using E42M76-125A and E33M61-136R as overgo probes. (A) The cloned *Bam*HI-BAC-ends were released from the pUC18 by complete digestion with *Bam*HI enzyme in comparison to the original BAC-clones *Bam*HI digested. The two obtained fragments have a larger size than pBeloBAC11 (7.5 kb). (B) Hybridization using pBeloBAC11 vector as probe against the positive BAC-clones and their cloned BAC-ends (minimum two each).

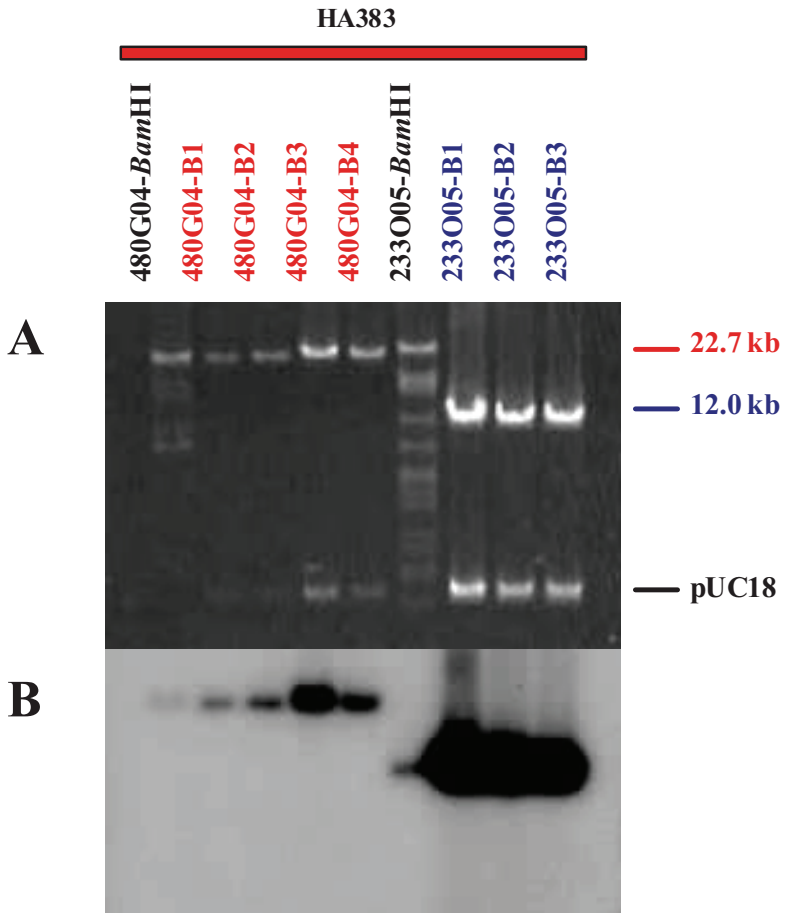


Fig. 23 Verification of cloned BAC-ends of the positive BAC-clones 480G04 and 233O05 identified using E62M52-249A and E41M48-113A as overgo probes, respectively. (A) The cloned *Bam*HI-BAC-ends were released from the pUC18 by complete digestion with *Bam*HI enzyme in comparison to the original BAC-clones *Bam*HI digested. The two obtained fragments have a larger size than pBeloBAC11 (7.5 kb). (B) Hybridization using pBeloBAC11 vector as probe against the positive BAC-clones and their cloned BAC-ends (minimum two each).

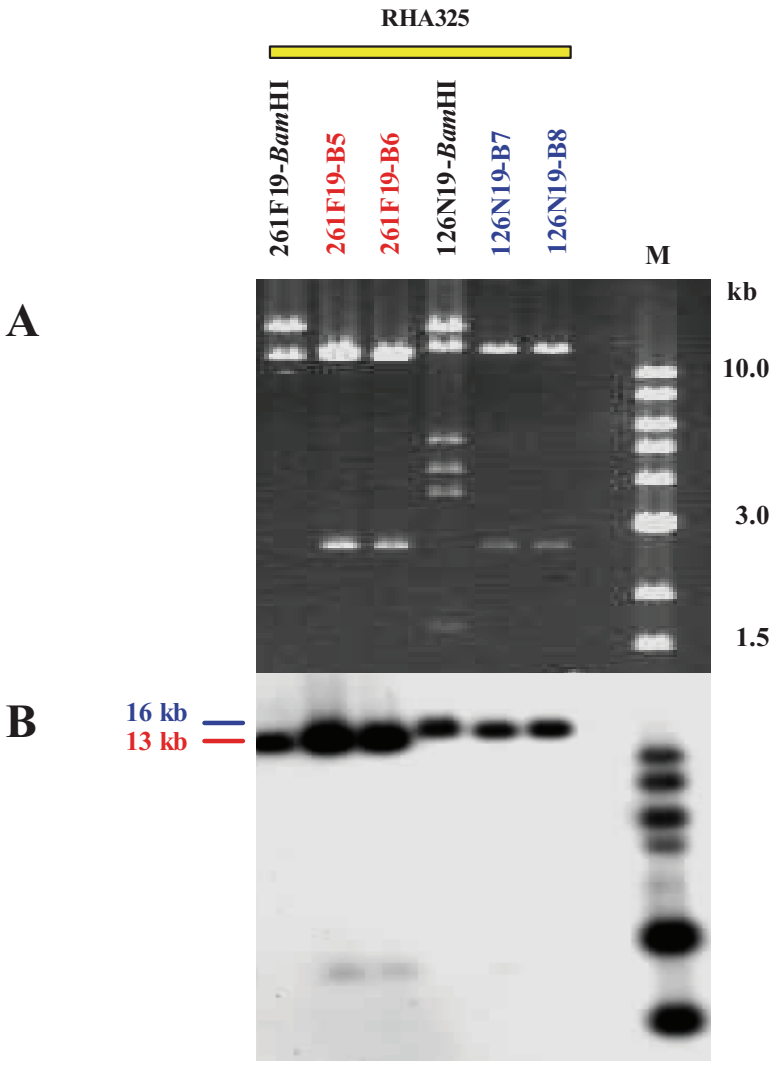


Fig. 24 Verification of cloned BAC-ends of the common positive BAC-clones 261F19 and 126N19 identified using E42M76-125A and E33M61-136R as overgo probes. (A) The cloned *Bam*HI-BAC-ends were released from the pUC18 by complete digestion with *Bam*HI in comparison to the original BAC-clones *Bam*HI digested. The obtained fragments have a larger size than pBeloBAC11 (7.5 kb). (B) Hybridization using pBeloBAC11 vector as probe against the positive BAC-clones and their cloned BAC-ends (two each).

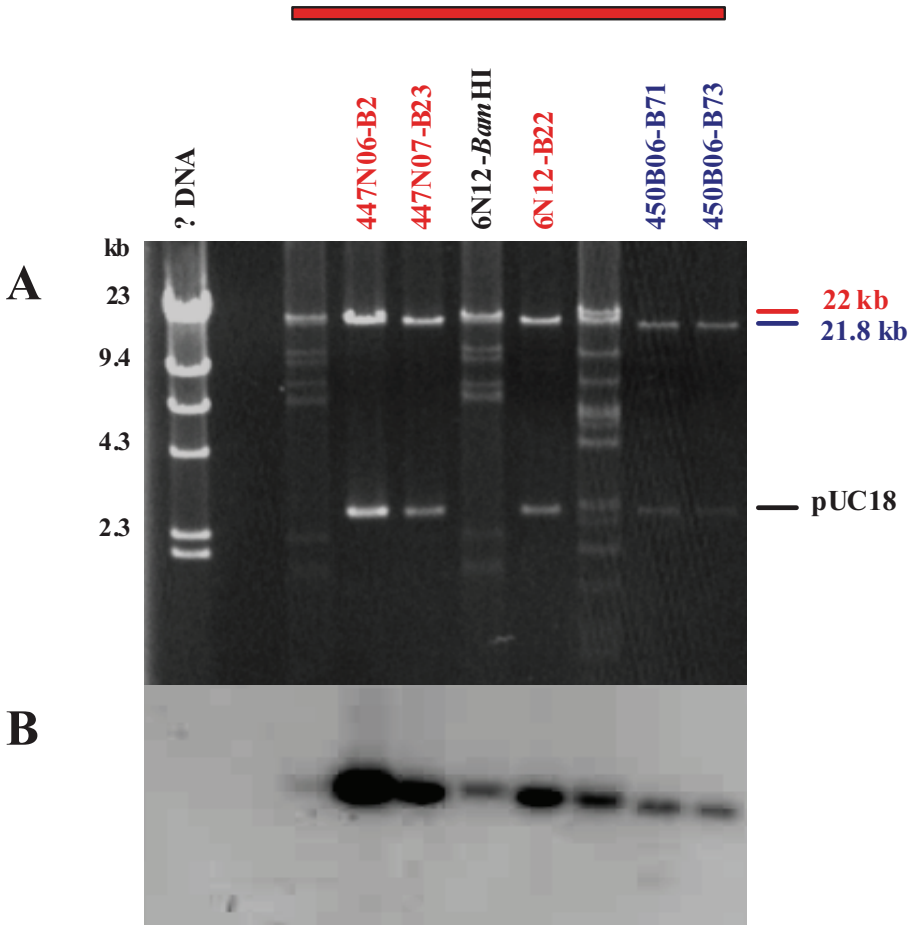


Fig. 25 Verification of cloned BAC-ends of the common positive BAC-clone 450B6 obtained using E44M70-275A and E33M61-136R as overgo probes, and verification of the cloned BAC-ends of the positive BAC-clones 447N06 and 6N12, obtained using E62M52-249A as overgo probe. (A) The cloned *Bam*HI-BAC-ends were released from the pUC18 by complete digestion with *Bam*HI in comparison to the original BAC-clones *Bam*HI digested. The obtained fragments have a larger size than pBeloBAC11 (7.5 kb). (B) Hybridization using pBeloBAC11 vector as probe against the positive BAC-clones and their cloned BAC-ends.

3.6.4 Verification of cloned BAC-ends by restriction digestion

For the BAC-ends, three independent subclones were isolated from each positive BAC-clone. To provide empirical evidence that the three sub-clones were identical prior to sequencing, we verified their authenticity using four different restriction endonucleases as shown here for the cloned BAC-end of 233O05 (Fig. 26). The three cloned BAC-ends had to show an identical restriction pattern for the four enzymes used (*Bam*HI, *Hind*III, *Pst*I, and *Eco*RI).

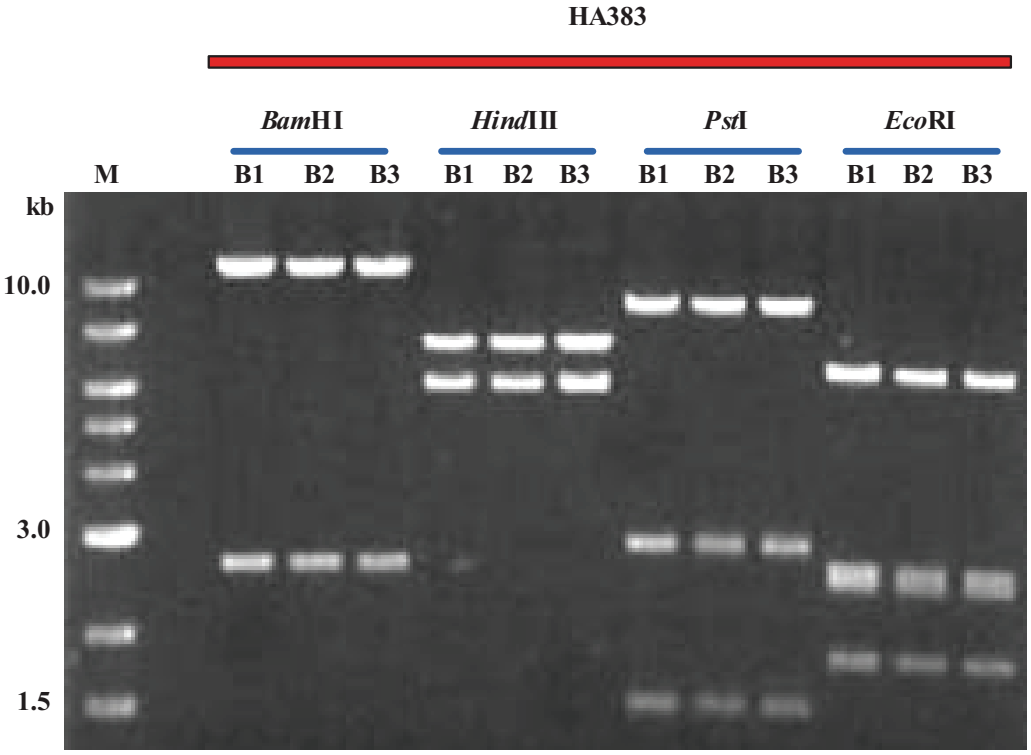


Fig. 26 Verification of three subclones of the cloned BAC-end of 233O05 digested with four restriction enzymes.

3.7 Chromosome walking and BAC contig construction

In order to develop a contig around the restorer gene *Rf1*, the next step was to identify overlapping BACs and to walk towards the gene. The walking was performed using overgo primer pairs, designed from the end sequences of BAC-clones, as probes. These probes were hybridized against the high-density filter sets of the two BAC libraries (RHA325 and HA383) to identify potentially overlapping and extending BAC-clones. Restriction fingerprints were performed to confirm all potential BAC-clone extensions,

or merges, identified with the BAC-end probes. The strategy of restriction fingerprinting is based on the hypothesis that the BAC-clones derived from the same region of a genome share similar restriction patterns: this means a large number of common bands (Zhang and Wu 2001) in addition to overhanging fragments. Nine BAC-ends sequences (Tab. 5) were used to design overgos that could be used as probes to rescreen the BAC libraries by colony hybridization. The remaining BAC-ends (Tab. 7) were not used further because they revealed homology to retrotransposons or repetitive elements.

Further verification of the positive BAC-clones was accomplished by Southern hybridizations using the radioactive labelled overgo probes. Southern hybridization of the fingerprints has two advantages. First, identifying potential joints between contigs, and second ensuring data integrity for chromosome walking to confirm all potential BAC-clone extensions identified with the BAC-ends. Positive BAC-clones identified (Tab. 6), were first subjected to fingerprinting and then overlap analyses, which were performed manually. The expectation is that the BAC-clones belonging to the same map location would produce the same banding patterns when hybridized with the respective overgos. This procedure assists the integration of physical and genetic maps, making chromosome walking relatively feasible (Yüksel and Paterson, 2005).

3.7.1 Identification of overlapping BAC-clones via fingerprinting and hybridization

The overgo derived from the RAPD marker OP-K13-454A was used to screen the BAC library HA383. The probe identified four BAC-clones in HA383 (307N02, 216F17, 225D09, 401E15) in addition to three positive BAC-clones in RHA325 (59J13, 67I05, 67N04). The positive BAC-clones were fingerprinted using *Hind*III and probed with the OP-K13-454A overgo to verify that the hybridising fragments had the expected size. The restriction pattern of the three BAC-clones (67N04, 67I05, 59J13) identified in RHA325 revealed a minimal potential overlap with the four BAC-clones identified in HA383 (Fig. 27A). Southern hybridization showed the same signal (7.2 kb and 8.2 kb) among the four new BAC-clones. However, these signals were different from the BAC-clones of the BAC library RHA325 (15 kb), suggesting a different sub-genomic localization (Fig. 27B). The BAC-ends from 401E15, 216F17, 67N04, 225D09 were cloned for further investigations.

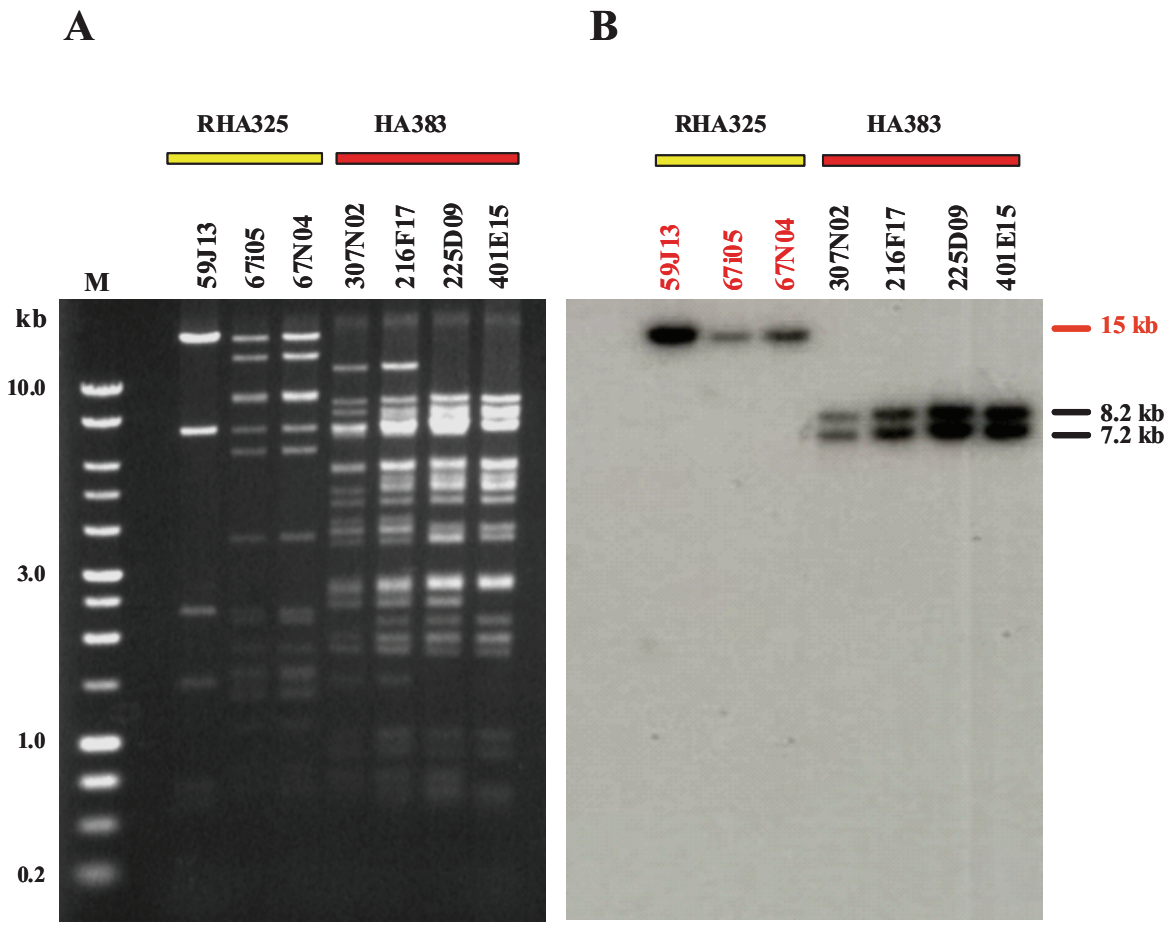


Fig. 27 Characterization of the positive BAC-clones identified by the OP-K13-454A overgo probe in the two BAC libraries (HA383 and RHA325). (A) BAC fingerprinting using *Hind*III restriction enzyme. (B) Southern hybridization using OP-K13-454A overgo against the positive BAC-clones.

To further investigate the obtained results, BAC-end sequences of 401E15, 216F17 and 225D09 were used as overgo radioactively labelled probes against the positive BAC-clones 59J13, 67I05, 67N04, 307N02, 216F17, 225D09, 401E15 (Fig. 28A). Southern hybridization (Fig. 28B), using 401E15-B overgo showed the same banding pattern between 225D05, 307N02, 216F17. The overgo 216F17 showed the same signal of 2.1 kb between the BAC-clones 225D05 and 401E15 but 307N02 did not show any signal. And the 225D09 BAC-end overgo showed a common signal of 0.6 kb only with 401E15.

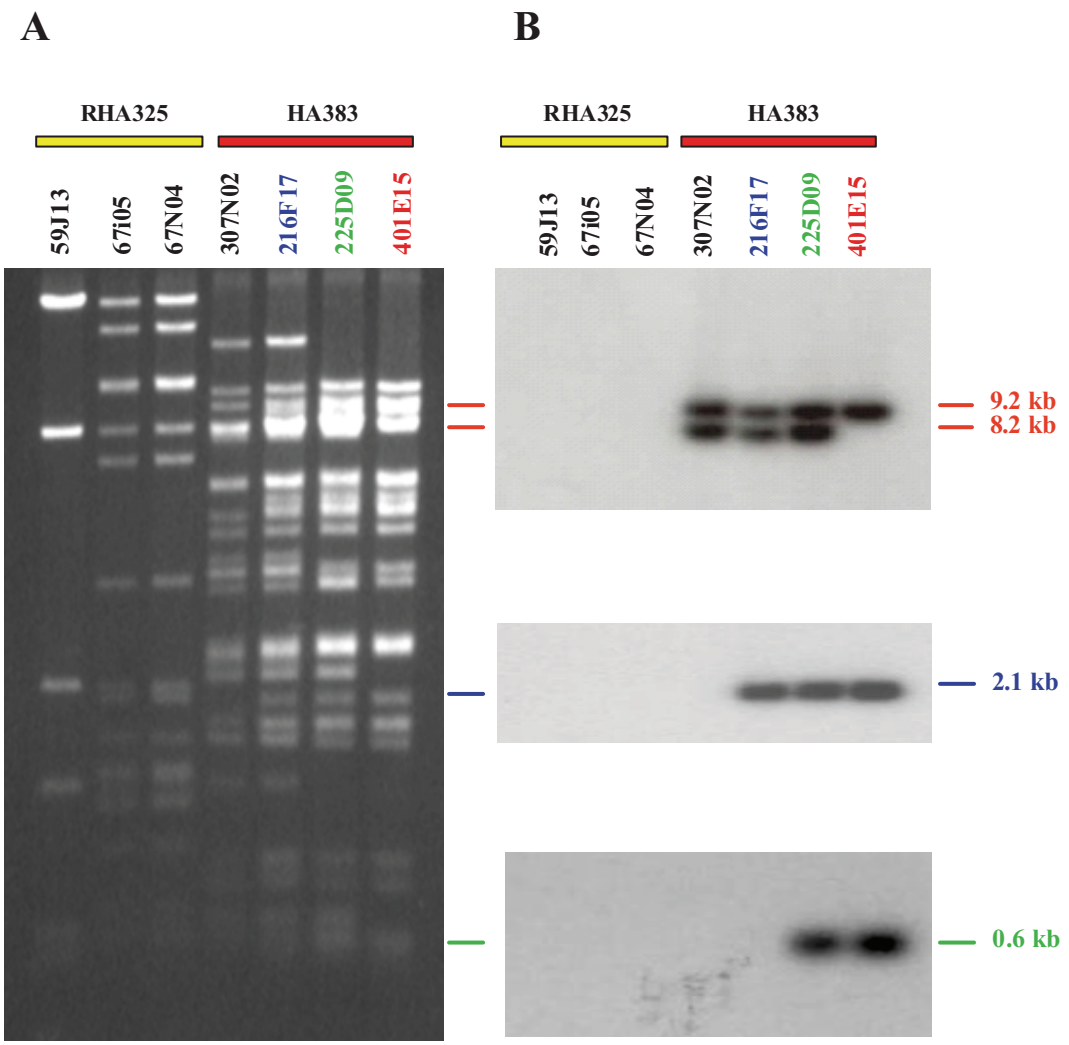


Fig. 28 Further characterization of the positive BAC-clones identified by the OP-K13-454A overgo probe in the two BAC libraries (HA383 and RHA325). (A) BAC fingerprinting using *Hind*III restriction enzyme. (B) Southern hybridization using 401E15, 216F17 and 225D09 as overgo radioactively labelled probes against the positive BAC-clones.

The whole BAC-end (1.5 kb) from the BAC-clone 67N04 was sequenced and an overgo probe was designed to re-screen both BAC libraries. The probe identified two BAC-clones in RHA325 (139A17, 172K02) and 11 BAC-clones in HA383. Fingerprinting with *Hind*III indicated that the 11 positive BAC-clones from HA383 fell into three groups. The first group consists of the BAC-clones 312E12, 179G06, 153F08, 486L01, 180A24, 177K11, 467E16. The second group is represented by the BAC-clones 89P04, 60K10, 192D15, and the third group consists only of a single BAC-clone 177M08. In each of the first two groups, many DNA fragments were shared and the degree of overlapping is shown in Figure 29A. To further confirm the results, Southern

hybridization using 67N04-B radioactively labelled probe were performed and showed among the new BAC-clones the same banding pattern (0.5 kb) apart from BAC-clone 177M08 but different from the BAC-clones 65I05 and 67N04 identified in restorer BAC library RHA325 (Fig. 29B). The single BAC-clone of group III has a unique signal of 1.5 kb. For RHA325, one signal (BAC-clone 179K02) corresponded to the signal obtained for the most of the BAC-clones from HA383 (0.5 kb). However, the BAC-clone 139A17 showed a signal of 1.2 kb which was different from the original BAC-clone 67N04 (1.4 kb) and 177M08 (1.5 kb) from HA383. The hybridization signals obtained by 67N04 as overgo indicated that this genome region seems to be present at least four times in the sunflower genome.

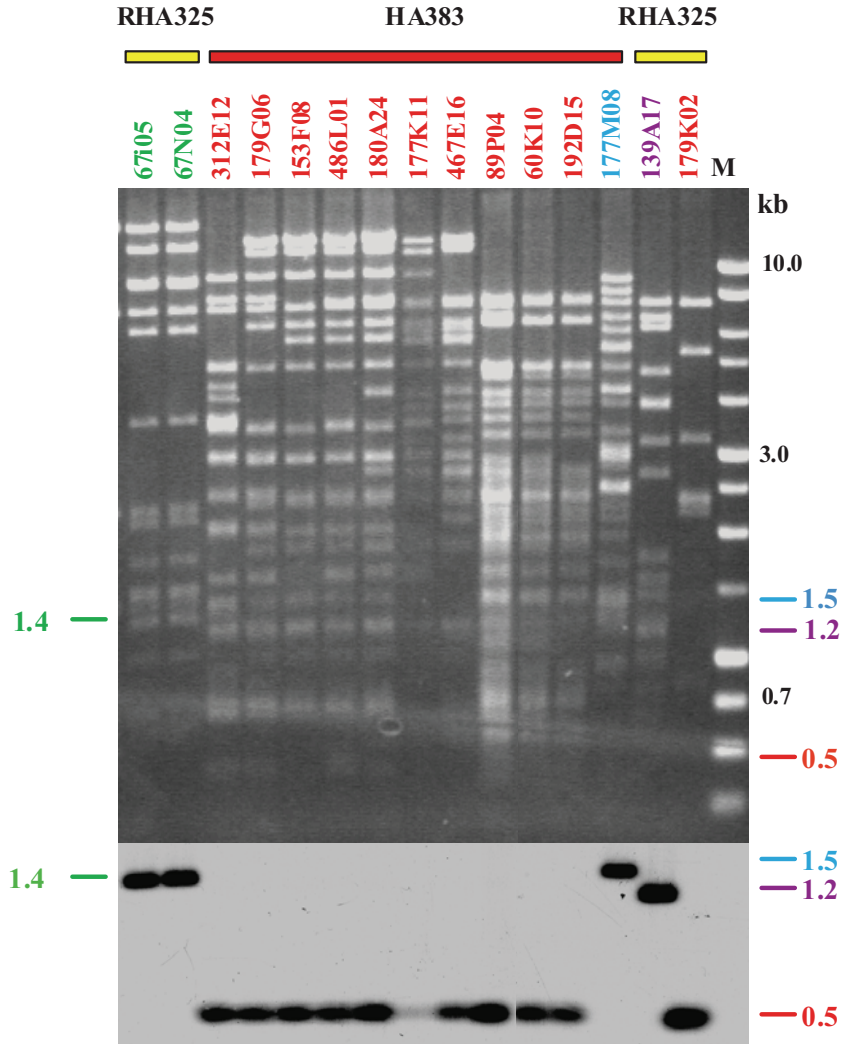


Fig. 29 Characterization of the BAC-clones identified by the 67N04-B-end overgo probe in the two BAC libraries (RHA325 and HA383). (A) BAC fingerprinting using *Hind*III restriction enzyme. (B) Southern hybridization using 67N04-B-end overgo against the new positive BAC-clones.

The BAC-end sequence of 225D09 was used as overgo probe against the BAC libraries. The probe identified one BAC-clone (115P09) in RHA325 and six BAC-clones in HA383, five new ones and the BAC-clone 401E15, which had been in the contig formed by 225D09, 307N02, and 216F17. The BAC-clones were subjected to fingerprinting and Southern hybridization (Fig. 30). The BAC-end overgo 225D09 detected in addition to the expected fragment of 0.6 kb, two other signals of 1.5 kb and 2.2 kb in the newly identified BAC-clones. These fragments must represent duplicated regions in the genome of RHA325 and HA383.

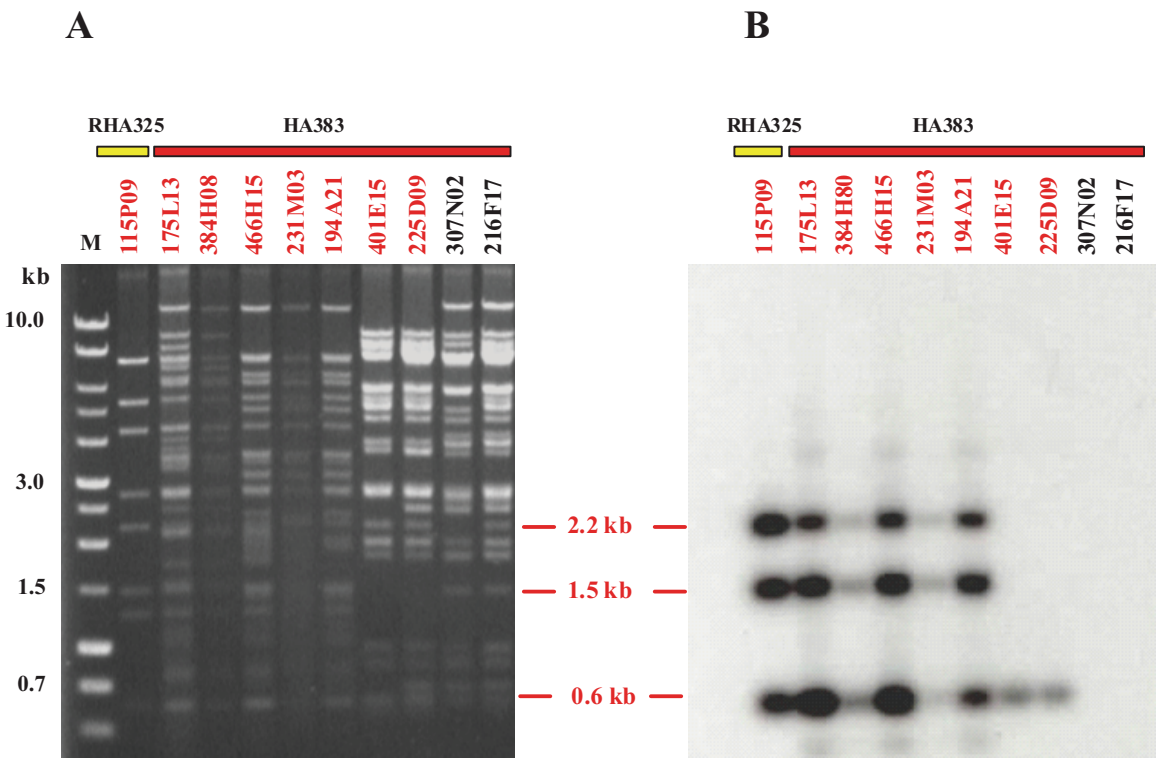


Fig. 30 Characterization of the positive BAC-clones identified by the 225D09-B-end overgo probe in the two BAC libraries (RHA325 and HA383). (A) BAC fingerprinting using *Hind*III restriction enzyme. (B) Southern hybridization using 225D09-B-end overgo against the new positive BAC-clones.

Sequencing of the BAC-end of 225D09 revealed homology to sf_2 (Appendix 8.2) with $2.00E^{-91}$ (E value), an anther specific protein in sunflower (Domon et al. 1990). Design of an overgo, named 225D09-B-sf_2, was used for hybridization against the high density BAC filters of the two BAC libraries. The overgo proved to be repetitive in the BAC

library RHA325 but only hit two BAC-clones (35i08, 454H06) in the BAC library HA383. Restriction digestion of the two BAC-clones identified in HA383 revealed a minimal overlapping with the original BAC-clone 225D09, but both shared several fragments (Fig. 31A). Southern hybridization using 225D09-B-sf$_2$ radioactively labelled probe showed different banding pattern between 225D09, 35i08, and 454H06 (Fig. 31B). The BAC-clones 35i08 and 454H06 showed a signal of 6.2 kb and 6.0 kb, respectively, that were different from the BAC-clone 225D09 (3.9 kb). This may indicate that the corresponding sequence of the BAC-end overgo is at least three times present in the HA383 genome.

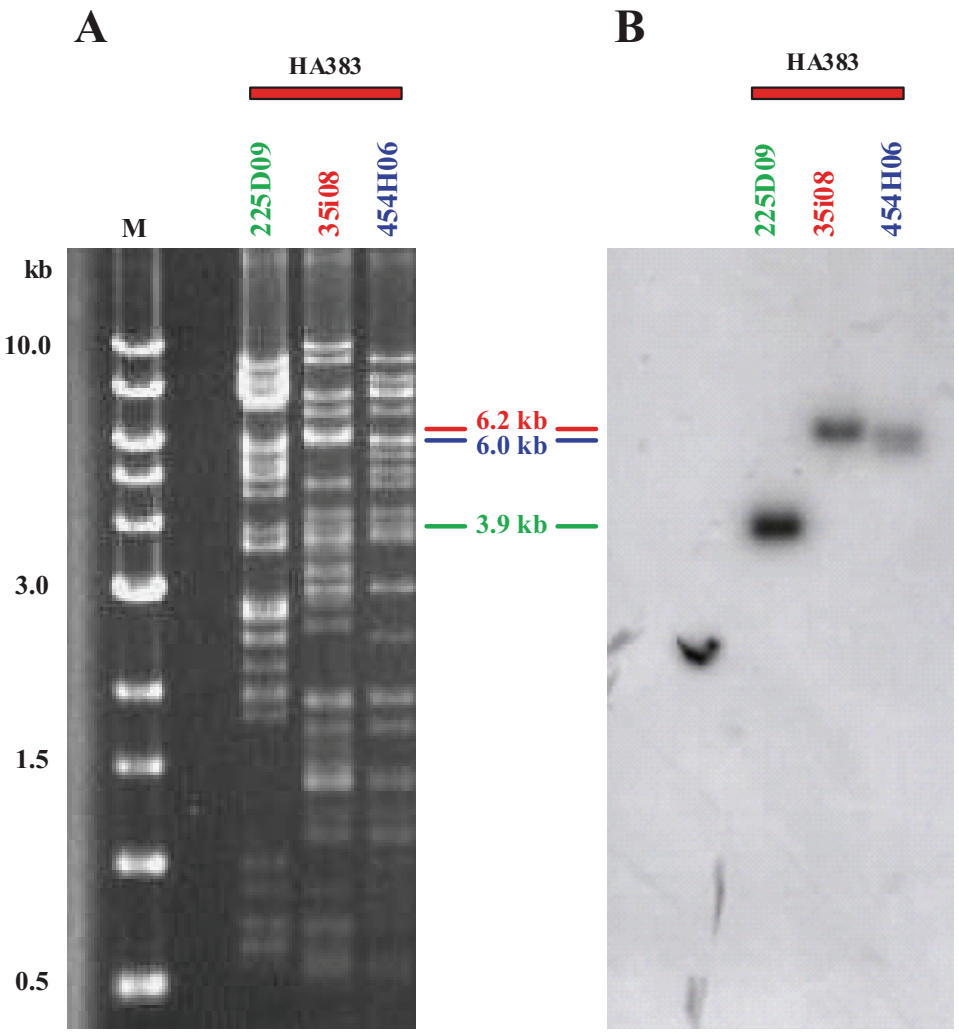


Fig. 31 Characterization of the positive BAC-clones identified by the 225D09-B-sf$_2$ overgo probe in the BAC library HA383. (A) BAC fingerprinting using *Hind*III restriction enzyme. (B) Southern hybridization using 225D09-B-sf$_2$ overgo against the two new positive BAC-clones and 225D09.

Further hybridizations have been done in order to find new positive BAC-clones. The AFLP marker E62M52-249A was used as overgo probe against the two BAC libraries. The probe did not identify BAC-clones in the BAC library RHA325 but identified three positive BAC-clones (6N12, 447N06, 480G04) in the BAC library HA383. *Hind*III fingerprinting indicated that the three BAC-clones 6N12, 447N06, and 480G04 shared several fragments, and the degree of overlapping is shown in Figure 32A. Southern hybridization, using E62M52-249A as radioactively labelled overgo probe, showed among the three new BAC-clones the same signal of 4.2 kb (Fig. 32B). However, the overgo derived from the BAC-end 480G04 did not hit any further positive BAC-clones in both BAC libraries.

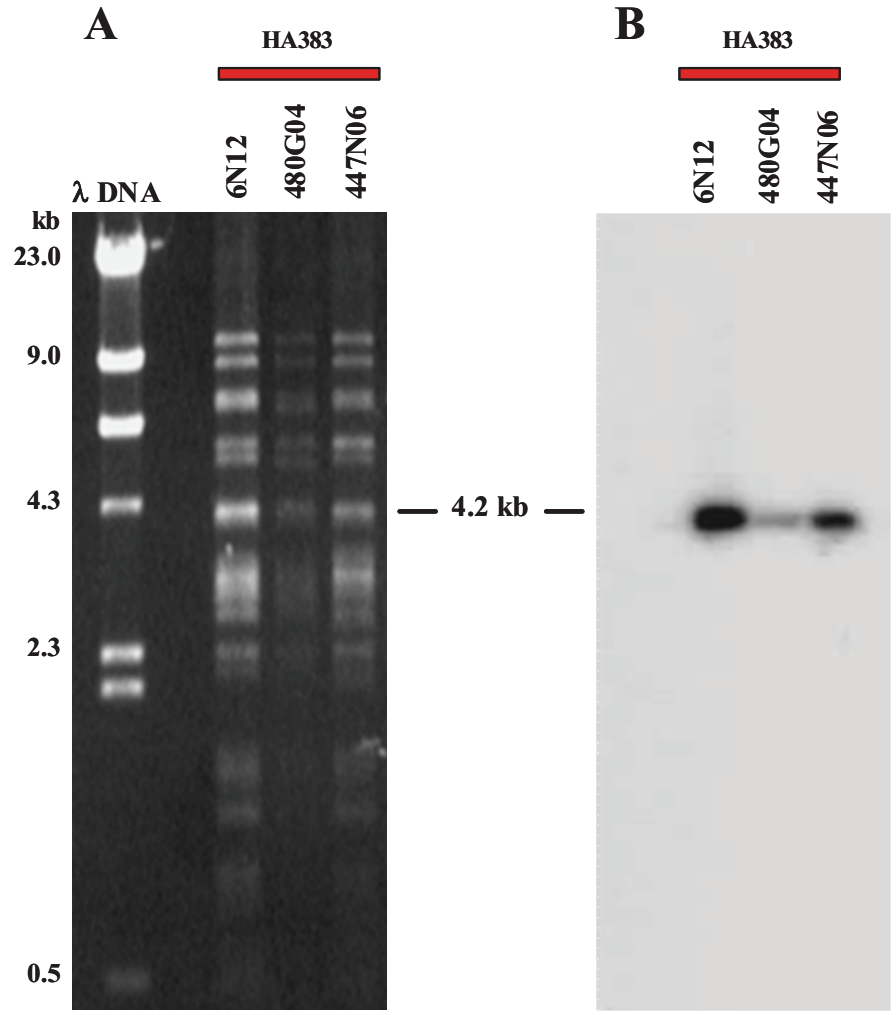


Fig. 32 Characterization of the positive BAC-clones identified by the E62M52-249A overgo probe in the BAC library HA383. (A) BAC fingerprinting using *Hind*III restriction enzyme. (B) Southern hybridization using E62M52-249A overgo against the three new positive BAC-clones. λ: Lambda DNA marker (Promega)

The overgos derived from the AFLP marker E33M61-136R and E42M76-125A were repetitive when used as probes against the BAC library RHA325. The filters were compared and three common BAC-clones were identified (126N12, 261D17, 261F19). These BAC-clones were subjected to fingerprinting and Southern hybridizations (Fig 33, Fig. 34). The E33M61-136R overgo detected three signals, 21.6 kb, 2.9 kb, and 1.7 kb (Fig. 33B), that were different from the signals obtained using E42M76-125A overgo probe (3.9 kb, 1.6 kb, 0.9 kb, and 0.6 kb) (Fig. 34B**),** indicating that each of them represents a different locus.

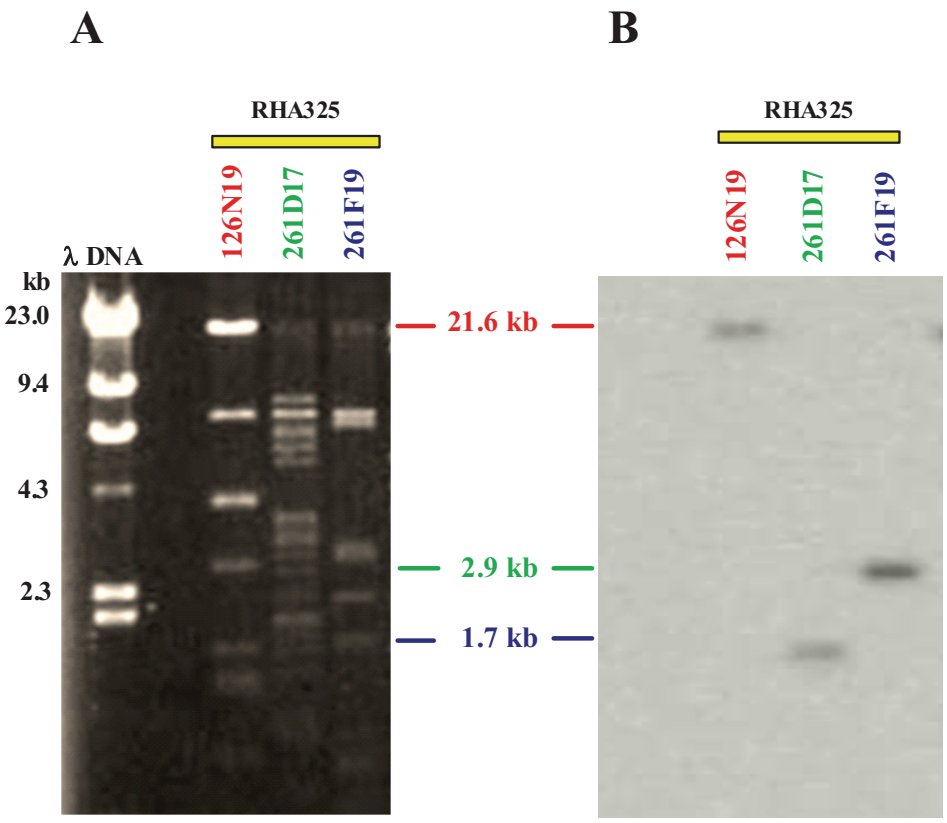


Fig. 33 Characterization of the positive BAC-clones identified by the E33M61-136R and E42M76-125A overgo probes in the BAC library RHA325. (A) BAC fingerprinting using *Hind*III. (B) Southern hybridization using E33M61-136R overgo against the new three positive BAC-clones. λ: Lambda DNA marker.

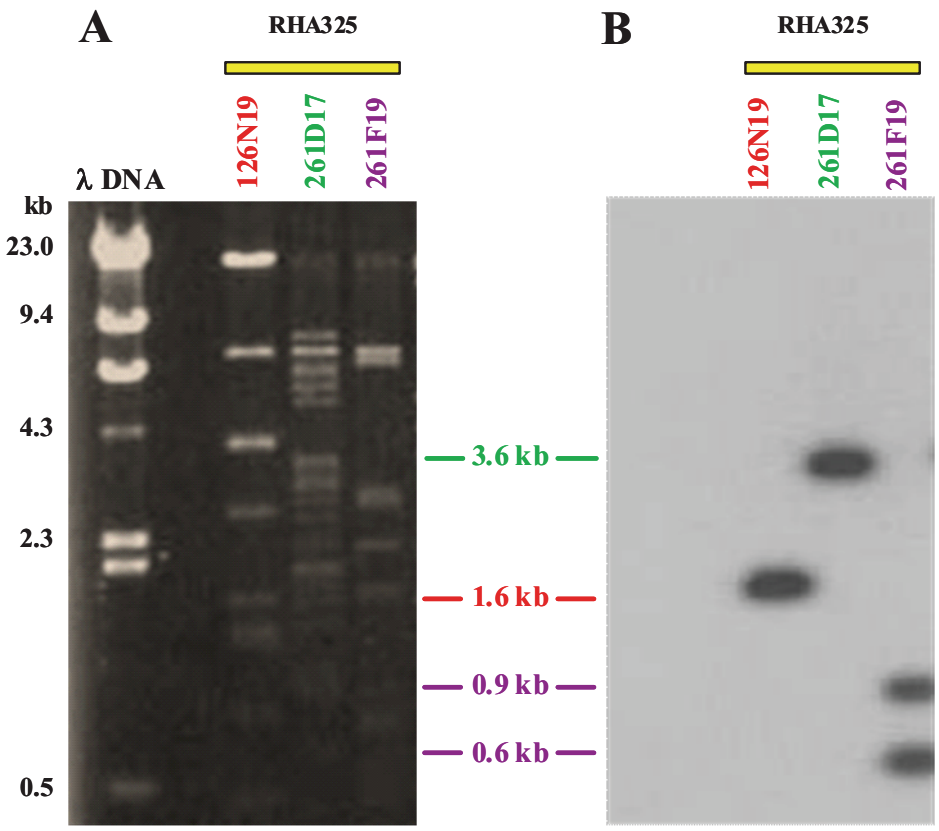


Fig. 34 Characterization of the positive BAC-clones identified by the E42M76-125A and E33M61-136R overgo probes in the BAC library RHA325. (A) BAC fingerprinting using *Hind*III. (B) Southern hybridization using E42M76-125A as overgo against the three positive BAC-clones. λ: Lambda DNA marker.

The BAC-end sequence of 261F19 was used to design an overgo probe for hybridization against the two BAC libraries. The probe identified one BAC-clone (39P12) in RHA325 and nine BAC-clones (131C20, 158M19, 497K04, 390M14, 482H17, 278B16, 269i11, 524G12, 513E12) in HA383. The BAC-clones were fingerprinted and subjected to Southern hybridization. Restriction digestions of the BAC-clone identified in RHA325 revealed a minimal overlapping with the nine BAC-clones identified in HA383 (Fig. 35A). Moreover, Southern hybridization using

261F19-B radioactively labelled probe against *Hind*III digestions of the BAC-clones showed different banding pattern among the new BAC-clones (Fig. 35B).

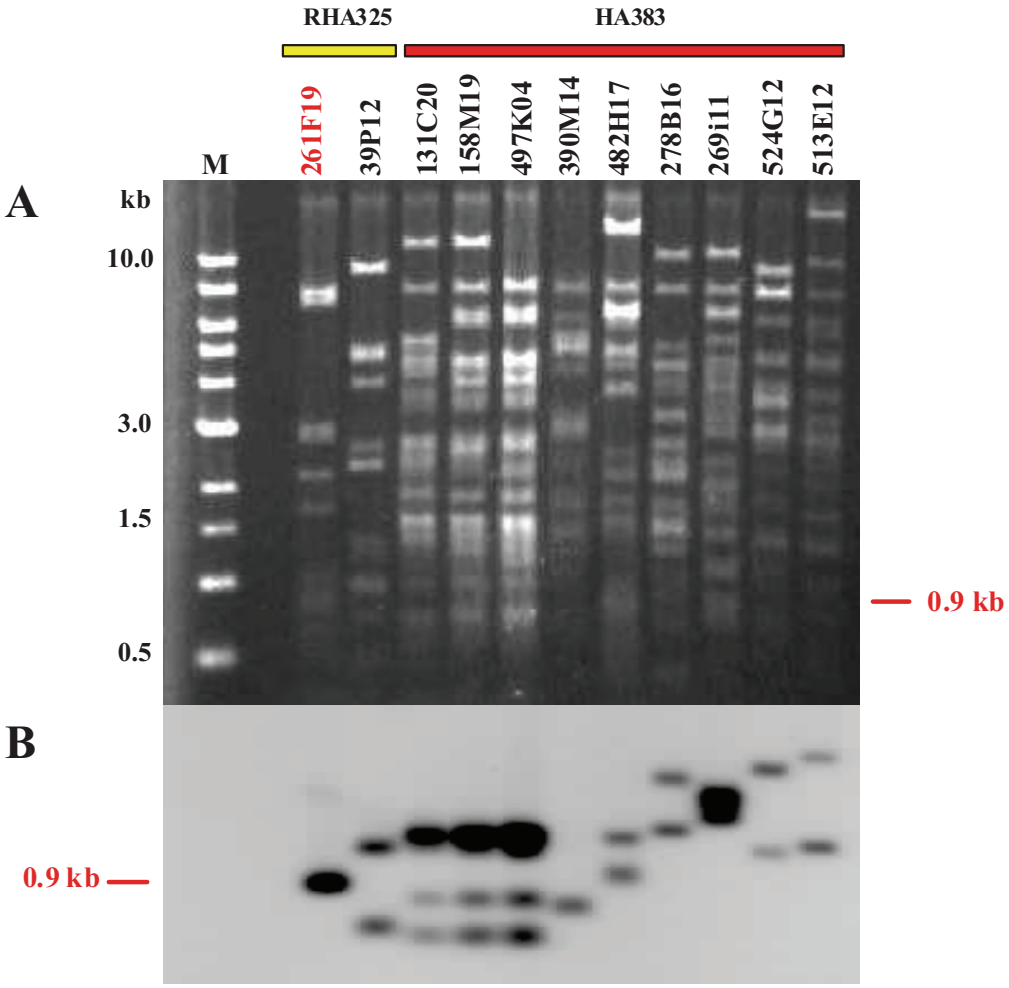


Fig. 35 Characterization of the positive BAC-clones identified by the 261F19-B-end overgo probe in the two BAC libraries (RHA325 and HA383). (A) BAC fingerprinting using *Hind*III restriction enzyme. (B) Southern hybridization using 261F19-BAC-end overgo against the new positive BAC-clones.

The overgo derived from the AFLP marker E41M48-113A was used to screen the two BAC libraries. The probe identified one BAC-clone (94F15) in RHA325 and four BAC-clones (233O05, 100L22, 147A03, 482D10) in HA383. Restriction enzymes of the BAC-clone identified in RHA325 revealed a minimal potential overlap with the four

BAC-clones identified in HA383 (Fig. 36A). Southern hybridization showed the same signal (4.7 kb) for BAC-clones 94F15 and 233O05. The two BAC-clones 147A03 and 482D10 showed also the same signal (4.8 kb), but the BAC-clone 100L22 had a unique signal of 3.2 kb (Fig. 36B).

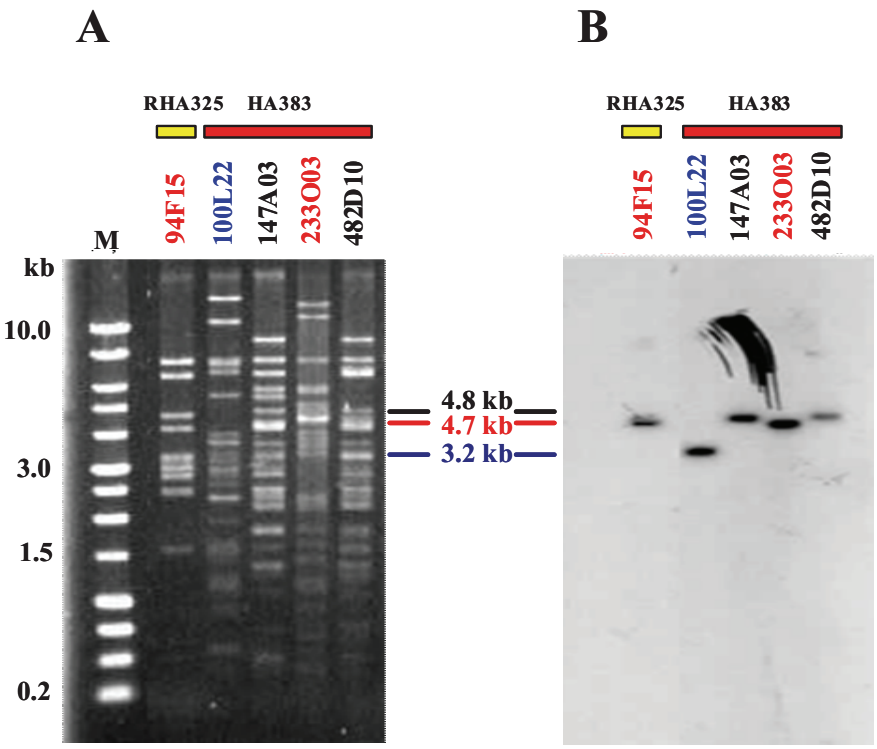


Fig. 36 Characterization of the positive BAC-clones identified by the E41M48-113A overgo probe in the two BAC libraries (HA383 and RHA325). (A) BAC fingerprinting using *Hind*III. (B) Southern hybridization using E41M48-113A overgo against the positive BAC-clones.

3.7.2 Contig assembly

A contig can be assigned unambiguously to a specific genetic location if the contig is associated with a probe (hit by) that is mapped to one genetic location and the probe does not hit BACs in other contigs (Cone et al. 2002). This is a simple concept but in reality a number of parameters could interfere making the contig construction a challenging issue. Overgos derived from seven markers OP-K13-454A, E32M36-155R, E62M52-249A, E44M70-250A, E42M76-125A, E33M61-136R, E41M48-113A and four BAC-end sequence probes (67N04-B, 225D09-B, 225D09-B-sf$_2$, and 261F19-B)

were used to screen the two BAC libraries (RHA325 and HA383) and to construct contigs around the restorer gene *Rf1*. E32M36-155R overgo identified only two BAC-clones in the BAC library RHA325. These two BAC-clones have been excluded because they were too small and they only represented one fragment of 5 kb.

Contigs were constructed manually by fingerprinting of the positive BAC-clones digested with *Hind*III. Southern hybridizations using BAC-end overgo as probes were used to identify overlapping contigs. Cross-hybridization experiments were performed to examine further overlap between the BAC contigs. The positive BAC-clones obtained with the overgos of OP-K13-454A and E62M52-249A were hybridized against the overgo 225D09-B-end, and vice versa. Unfortunately no hybridization showed overlapping between the contigs (data not shown). Consequently, the physical map in the region of the *Rf1* revealed gaps between BAC-contigs and no contiguous region of the genome has been found as each contig or singleton appeared to represent a unique genomic region.

Moreover the sequences of the BAC-ends were subjected to an automated contig assembly using VectorNTI software (Invitrogene) for sequence analyses and data management (Appendix 8.3). BAC-end sequences of 115P09 (from RHA325) and 216F17, 384H08 (from HA383) obtained by overgo derived from the RAPD marker OP-K13-454 were determined to be overlapping (Appendix 8.3). However, the investigation has not been continued because the contig containing 225D09 (Fig. 37) back-mapped in another linkage group than the *Rf1* gene. This can occur when the probe represents a duplicated region in the genome which might be associated with more than one genetic location. This part of the contig was removed from the physical map surrounding the restorer *Rf1* gene (Fig. 38).

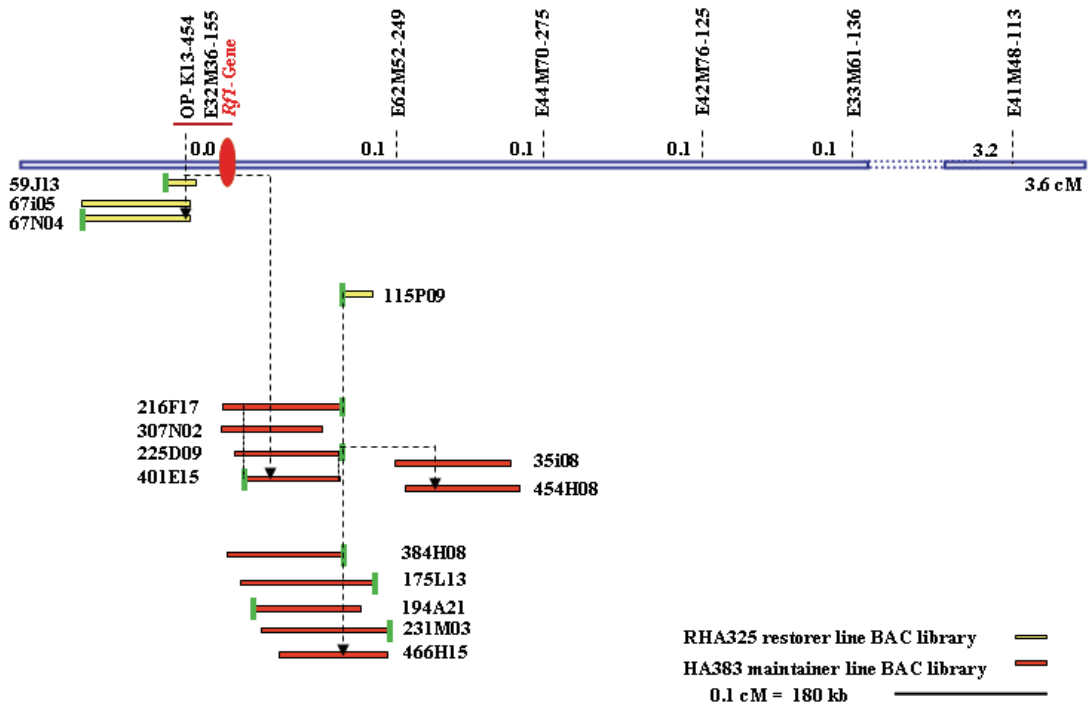


Fig. 37 Hybridization with the overgos derived from the RAPD marker OP-K13-454, 225D09-B-end and 225D09-B-sf$_2$. (A) Genetic map shows the location of the *Rf1* gene on linkage group 13 of sunflower SSR map. This map was constructed based on 91 F_2 individuals generated from the cross between RHA325 and HA342. Numbers above the line represent genetic distances in centiMorgans (cM) between one RAPD marker (OP-K13-454A) and six AFLPs markers. The MAPMAKER 3.0b computer program (Lander et al. 1987) with Kosambi function (Kosambi 1944) was used in this map construction. (B) The physical map consists of 4 BAC-clones (59J13, 67N04, 65i05, 115P09) from restorer RHA325 BAC library and 11 BAC-clones (216F17, 307N02, 225D09, 401E15, 175L13, 384H08, 194A21, 231M03, 466H15) from maintainer BAC library HA383. All are labelled with the name of the BAC-clones. Small boxes (in green) at the extremity of the BACs represent cloned BAC-ends.

The physical map surrounding the fertility restorer locus *Rf1* consists of 10 BAC-clones identified in the BAC library RHA325 and 28 BAC-clones identified in the BAC library HA383 representing seven contigs and 15 singletons (Fig. 38).

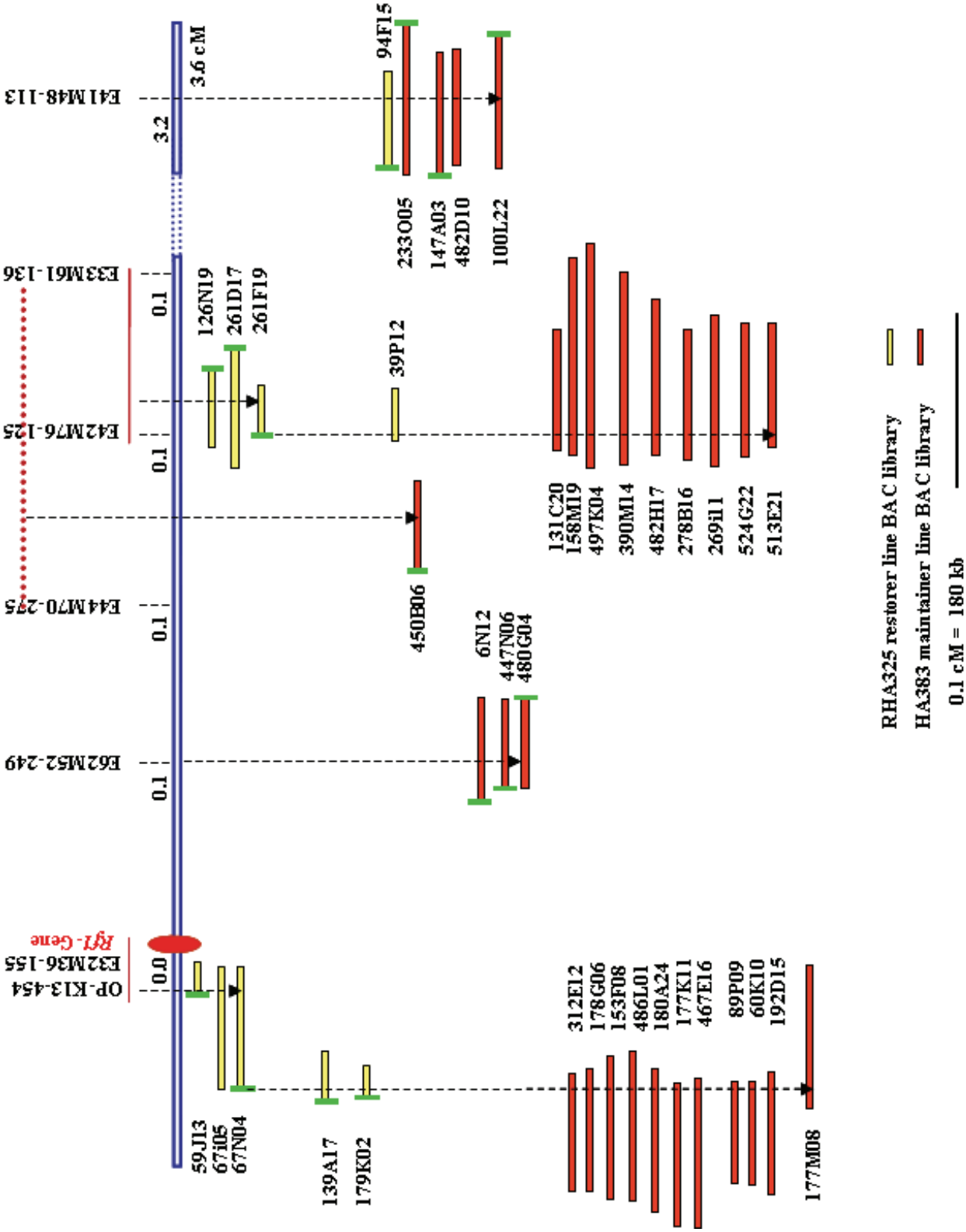

OP-K13-454
E32M36-155
Rf1- Gene
0.0
E62M52-249
0.1
E44M70-285
0.1
E42M76-125
0.1
E33M61-136
0.1
E41M48-113
3.2
3.6 cM
59J13
67i05
67N04
139A17
179K02
312E12
178G06
153F08
486L01
180A24
177K11
467E16
89P09
60K10
192D15
177M08
450B06
6N12
447N06
480G04
126N19
261D17
261F19
39P12
233O05
147A03
482D10
100L22
131C20
158M19
497K04
390M14
482H17
278B16
269i11
524G22
513E21
RHA325 restorer line BAC library
HA383 maintainer line BAC library
0.1 cM = 180 kb

Fig. 38 The genetic and physical map of *Rf1* genomic region. (A) Genetic map shows the location of the *Rf1* gene on linkage group 13 of sunflower SSR map. (B) The physical map represents seven BAC-contigs and 15 singletons consisting of 10 BAC-clones (59J13, 67N04, 65i05, 139A17, 179K02, 126N19, 261D17, 261F19, 39P12, 94F15) from restorer RHA325 BAC library and 28 BAC-clones (312E12, 178G06, 153F08, 486L01, 180A24, 177K11, 467E16, 89P09, 60K10, 192D15, 177M08, 6N12, 447N06, 480G04, 450B06, 131C20, 158M19, 497K04, 390M14, 482H17, 278B16, 269i11, 524G22, 513E21, 233O05, 100L22, 147A03, 482D10) from maintainer BAC library HA383. All are labelled with the name of the BAC-clones. Small boxes (in green) at the extremity of the BACs represent cloned BAC-ends. All singletons and contigs apart from 59J13, 67i5, and 67N04 need still to be back-mapped in the F_2-population to confirm their localisation in the *Rf1* region.

3.8 Locating BAC-ends on the genetic map

Two BAC-ends, 67N04-B-end from the restorer library RHA325 and 225D09-B-end from the maintainer library HA383, were back mapped in the F_2-population of the cross RHA325×HA342 using a PCR-based approach, to integrate the contigs with the genetic linkage map (Horn, personal communication). The end of the BAC-clone 67N04 mapped at 3.9 cM away from the restorer gene *Rf1*, in contrast to the end of 225D09, which mapped on another linkage group than the restorer gene *Rf1*.

3.9 Homology search and sequence alignment

Nucleotide sequences of cloned BAC-ends (401E15-B, 67N04-B, 261F19-B, 447N06-B, 94F15-B, 225D09, 216F17, 100L22, 175L13) (Appendix 8.2) were used to search in the updated version of the non-redundant databases (www.ncbi.nih.gov/blast) applying algorithms for nucleotide sequences and using BLASTN and BLASTX (Altschul et al. 1990) to search homologies on nucleotide and protein level. NCBI (National Centre for Biotechnology Information) is an accurate database, which provides a high level of annotation such a description of protein function, domain structure, etc. Therefore, we used this tool to explore the similarities of the BAC-end sequences with other genomes. The BLASTN resulted in one unique sequence with cutoff (*E* value) of $2e^{-91}$, all other homologies are the result of BLASTX. The summary of the results are shown in Table 8. Most of the shared sequences were similar to retrotransposons and constituted the major component of the data. They were also best matched to *Arabidopsis thaliana*. This is likely due to the fact that *A. thaliana* is dicotyledonous as sunflower and is one of the best studied plant species whose genome is completely sequenced (Tomkins et al. 2001b). Other plant species giving similarities were: *Cucumis melo, Zea mays*, *Oryza sativa*, *Nicotiana tabacum*, *Vitis vinifera*, and *Populus trichocarpa*. The only match with sunflower was with the nucleotide sequences of an anther-specific gene *sf*$_2$ encoding a proline- and glycine-rich polypeptide (Domon and Steinmetz 1994). The sequences of the remaining BAC-ends did not show any homology or the cutoff E value was too high. It is currently unknown what these sequences represent. The identity was based on one part of the sequence of the BAC-ends and not on the whole sequence. It could be necessary to sequence the whole BAC-ends and to repeat the alignments.

Tab. 8 Results from query of sunflower BAC-end sequences against NCBI databases.

BAC-ends	End size kb	Sequence bp	Putative identity	Identities bp - %	BLAST	E-Value
401E15	3.7	1,500	Gag-protease polyprotein *Cucumis melo*	219/828 (26%)	BLASTX	1.00E-15
67N04	1.5	1,500	SocE *Myxococcus xanthus*	153/276 (55%)	BLASTX	1.00E-16
261F19	5.5	3,953	Putative gag-pol polyprotein (*Zea mays*)	198/390 (50%)	BLASTX	8.00E-32
			Jil putative pol protein (*Zea mays*)			
			Copia-type pol polyprotein (*Zea mays*)			
			Gene-Os08g0201800 (*Oryza sativa*). Reverse transcriptase indicative of a mobile element	186/390 (47%)	BLASTX	4.00E-27
447N6	15.2	1,042	Acyltransferase-like protein (*Nicotina tabacum*)	191/612 (32%)	BLASTX	9.00E-30
			Transferase family protein (*Arabidopsis thaliana*)	231/774 (29%)	BLASTX	6.00E-26
94F15	3.7	1,905	Hypothetical protein (*Vitis vinifera*)	537/819 (65%)	BLASTX	8.00E-49
			Glycosyl transferase family 1 protein (*Arabidopsis thaliana*)	495/710 (61%)	BLASTX	3.00E-49
225D09	8.1	5,138	Gene DCT/DIT2 (Dicarboxylate transport) oxoglutarate malate Antiporter (*Arabidopsis thaliana*)	243/453 (53%)	BLASTX	4.00E-50
			Hypothetical protein (*Vitis vinifera*)	237/447 (53%)	BLASTX	7.00E-51
			sf_2 for an anther-specific gene (*Helianthus annuus* L.)	369/452 (81%)	BLASTN	2.00E-91
216F17	6.7	1,482	ATSR (ARABIDOPSIS THALIANA SIGNAL-RESPONSIVE) *Arabidopsis thaliana*.	99/144 (68%)	BLASTX	6.00E-19
100L22	5.8	592	Putative integrase (*Beta vulgaris*)	150/564 (26%)	BLASTX	8.00E-10
175L13	8.6	596	Sodium proton exchanger (*Populus trichocarpa*)	210/447 (46%)	BLASTX	6.00E-28

4 Discussion

Positional cloning is a promising method for isolating and studying genes for which only the locus-derived phenotype is known (Meksem et al. 2000; Pflieger et al. 2001; Jander et al. 2002). It requires the development of a physical map of the genomic region spanning the gene of interest. For this purpose large-insert libraries constructed with YACs and BACs have been the tools of choice for generating high resolution physical maps (Arondel et al. 1992; Tanksley et al. 1995; Song et al. 2003). However, BAC libraries became the system of choice because they have several advantages over the YAC system, like high transformation efficiency, easy isolation of insert from *E. coli* cells, and low frequency of chimeric BAC-clones (Shizuya et al. 1992; Woo et al. 1994). Indeed, the BAC system can be more efficient in generating a high resolution physical map and identifying a number of overlapping BAC-clones covering the maximum length of a genome (Yang et al. 1997). These BAC-clones can be identified by PCR-based screening or by colony hybridisation with markers providing the initial point for chromosome walking and the eventual cloning of the target gene.

In this study, we attempted to clone the *Rf1* gene restoring male fertility in the presence of the PET1 cytoplasm in sunflower. We used as starting point the map previously developed by Kusterer et al. (2004a, 2004b, 2005) and additional new markers mapped tightly to the gene *Rf1* (Horn 2008, personal communication) to screen two sunflower BAC libraries (RHA325 and HA383) in order to construct a contig spanning the gene *Rf1*. However, this study has demonstrated that isolation of a gene using map-based cloning strategy is not as easy as generally assumed. Despite its universal potential, map-based strategy remains a challenge in plant species with large genomes (Ballvora et al. 2001) and it can only be considered as routine in model plants like rice or *A. thaliana*, which have a small genome (Lukowitz et al. 2000).

4.1 Characterization and evaluation of restorer and maintainer BAC libraries used in this study

As a first step towards map-based cloning, two available BAC libraries, the first constructed with a restorer line RHA325 (Özdemir et al. 2004) and the second constructed with a maintainer line HA383 (CUGI) have been screened. The screening

has been accomplished using colony hybridization, which has been successfully applied e.g. in *A. thaliana* (Mozo et al. 1999), potato (Song 2000), maize (O'Sullivan et al. 2001), melon (Luo et al. 2001), rice (Tao et al. 2001), peach (Georgi et al. 2002), and garlic (Lee et al. 2003).

4.1.1 Size of positive BAC-clones

The average DNA fragments indicated by the average size of positive BAC-clones was approximately half (32 kb) than that expected for the restorer library RHA325 (60 kb) and smaller (115 kb) than that for the maintainer library HA383 (125 kb). Similar results were observed in the construction of other BAC libraries like peanut (Wang et al, 1995), soybean (Danesh et al. 1998; Meksem et al. 2000), and rice (Yüksel and Paterson 2005). There is a compromise in transformation efficiency between the size and the number of *E. coli* transformed during the construction of a BAC library. The smaller molecules transform *E. coli* with much higher efficiency. When the average insert size of the BACs is increased, the number of BAC-clones obtained decreases dramatically. Frijters et al. (1997) picked 19,000 BACs from one ligation which had an average insert size of 93 kb, whereas only 4,180 BACs have been isolated from another ligation with an average insert size of 139 kb.

4.1.2 Genome coverage of the libraries

The genome coverage indicated by the number of BAC-clones selected from the restorer RHA325 (1.1) and the maintainer HA383 (4.0) was half of what was expected from estimates of haploid genome size based on measurement of the whole stained nuclei (3000 Mb), the average BAC insert size and the number of clones in each BAC library. Tomkins et al. (2001a) observed a smaller number of hits per overgo, suggesting a lower frequency of *Hind*III restriction site in the targeted *GhMYB* gene regions specific of fiber of the cotton genome. The same results have been obtained by Tao et al. (2001). They tested the true genome coverage of two large-insert rice *Hind*III BAC libraries and there was 15% difference between the theoretical and realized genome coverage. These discrepancies may be due (i) to an underestimation of genome size or (ii) an overestimation of the average insert size and the limited number of BAC-clones used for estimation of the genome coverage.

4.1.3 Stability of the BAC-clones

The insert stability affects the representation of the genome by the library. The possibility remains that BAC-clones may have internal alterations which could affect the use of the libraries for physical mapping and chromosome walking (Kim et al. 1992). Ideally BAC-clones should randomly and unbiased cover the genome without cloning artefacts or rearrangements (Osoegawa et al. 2001). In general, recombination is the major cause of instability of plasmid BAC-clones in *E. coli* and *recA* allele is the most important host gene for stabilizing a plasmid insert (Ishiura et al. 1989). If rearrangements occur during the bacterial transformation or shortly thereafter, independently obtained overlapping BAC-clones would result in an inconsistent genome structure. Insert stability was tested for the positive BAC-clones for several generations, and the DNA was extracted from cell populations (data not shown). For most BAC-clones, the experimentally measured fragments matched the predicted length and no deletions, insertions, or aberrant bands were found. However, two of 71 positive BAC-clones identified showed differences: 139A17 showed a deletion of about 10 kb and 39P12 was missing the pBeloBAC11 vector band. These results indicate that some BAC-clones are unstable when cloned in *E. coli*. Nevertheless, since no deletions have been detected in BAC-clones from human (Shizuya et al. 1992), sorghum (Woo et al. 1994), and bovine (Cai et al. 1995) BAC libraries, and the rate of differences in our study was small we assumed that the large genomic DNA is stably integrated in the pBeloBAC11 and pIndigo536, respectively.

Chimeric BAC-clones are one of the problems encountered in constructing an accurate physical map. Recombination and transformation of multiple large DNA fragments, which are the most plausible mechanisms for creating chimeric BACs, are expected to occur at much lower levels in *E. coli* as compared to yeast (Osoegawa et al. 2001). Generally, it has been proved that BACs have no or rare chimera (Shizuya et al. 1992). More, Woo et al. (1994) and Jiang et al. (1995) did not find any chimera when they investigated the sorghum and the rice BAC libraries by FISH (fluorescence *in situ* hybridisation).

4.1.4 Evaluation of BAC libraries used

An important issue in genome characterization and sequencing is to provide efficient access to the genomic BAC-clones that represent accurately copies of the DNA originated from the region of interest. The two BAC libraries used in this study have total genome coverage of about 10-fold providing a sufficient tool for genomic research. Tomkins et al. (2002) reported a construction and characterization of *Hind*III BAC library. Their results showed that using *Hind*III as restriction enzyme gave a high quality of BAC-clones and resulted in more than 99% probability of recovering any specific sequence of interest. However, our results demonstrate that BAC libraries are biased and that the used two BAC libraries may not completely cover the entire sunflower genome. If this is the case, the problem can be solved by developing more BAC-clones and using one or two new restriction enzymes.

All large-insert BACs libraries developed to date are generated from HMW (high molecular weight) DNA partially digested with restriction enzymes. Studies showed that the distribution of the restriction enzyme sites throughout the genome is uneven (Zhang and Wu 2001; Fang et al. 2004; Akhunov et al. 2005) and results in preferential cloning of some genomic regions and underrepresentation of other regions (Tomkins et al. 2001a). As result, gaps might occur in some regions of the genome if these libraries are used for physical mapping. Therefore, constructing new BAC libraries of a restorer line (RHA325) using different enzyme restrictions will provide enough coverage for future genome research. These new BAC libraries would complement each other and by this would minimize the underrepresentation of certain regions caused by the use of one particular restriction enzyme for BAC library construction (Frijters et al. 1997; Yim et al. 2002). The choice of the enzymes will depend on the complementarity of their nucleotide sequences at the restriction sites, for example, the combination of *Bam*HI being rich in G/C and *Eco*RI being rich in A/T would be an ideal combination for a good coverage of a genome (Chen et al. 2004). Deep libraries based on DNA fragments generated by different restriction enzymes are important and crucial for the development of complete contigs covering large genomic regions (Cao et al. 1999).

4.2 Overgos as hybridization probes

4.2.1 Advantages of overgos

Overgo probes represent a significant improvement for hybridization against high density BAC filters. This is due to several features of overgos: (1) they are short and show a low background in hybridizations because they are easy to be washed off from the filters without causing high backgrounds due to nonspecific binding on the filters (Cai et al. 1998). (2) The sequences for designing overgos can be selected, and thus probes containing repetitive sequences can be avoided. Therefore, the cross hybridization problem that is frequently associated with screening large genomic DNA libraries can be minimized (Feng et al. 2006). (3) Overgos facilitate high throughput procedures and are easy to handle compared to the conventional cloned probes. (4) Overgos can be obtained at reasonable costs and are highly labelled resulting in strong hybridization signals. (5) The major advantage lies in the short sequences (40 bp) needed for an overgo probe design, which increases the probability or the likelihood of finding suitable single copy regions. However, the selection criteria for uniqueness should be quite stringent because we discovered that screening against uniqueness in databases alone is not sufficient for overgo selection.

4.2.2 Characterization and evaluation of overgos

Among the 18 overgos used in screening the two BAC libraries (Tab. 5), 12 hit (66%) positive BAC-clones. The overall rate of 66% of the overgos identifying one or more BACs per probe is lower than the rates obtained in mouse (Cai et al. 1998), human (Han et al. 2000), sunflower (Feng et al. 2006), and maize (Gardiner et al. 2004), with 92%, 91%, 92% and 88%, respectively. The rate of overgos detecting BAC-clones is significantly affected by the genome coverage of the libraries because the probability that a properly designed overgo will detect complementary BAC-clones assumes that this BAC is present among the BAC-clones of the library (Romanov et al. 2003).

The failure of four overgos to identify positive BAC-clones from both BAC libraries may be due to improper overgo design, failed labelling, or improper hybridization stringency. The probes, which hit BAC-clones in one of the BAC library but not in the other (Tab. 5) show that a properly designed overgo will not be able to detect a complementary BAC-clone, because (i) that segments of the sunflower genome might not be represented in the subset of BAC-clones of the corresponding library or (ii) the

probes used were derived from different sunflower genomes (Nakamura et al. 1997; Akhunov et al. 2005).

A comparative sequencing study of the *bz1* region of two North American maize lines revealed extensive differences in the organization and content not only of the intergenic retrotransposons clusters in that region but also in the order of structural genes (Fu and Dooner 2002). Another hypothesis postulates that probes will hybridize to the same locus in different populations and that the gene order in small genome regions is conserved throughout different populations. Some studies demonstrated the conservation of the gene order throughout different populations is maintained so that one probe hybridized with the same locus in a different population (Bohuon et al. 1996; Guimaraes et al. 1997).

Five overgo probes were repetitive: three of them were derived from AFLP markers. This can be explained by the repetitive nature of AFLPs. It has been demonstrated by *in situ* hybridization of cloned AFLPs from *Asparagus officinalis* that AFLPs possess sequences, which are highly repetitive and can be found all over the genome (Reamon-Büttner et al. 1999). Two of the repetitive overgo probes were derived from BAC-ends. Due to the limited information available on the sunflower genome sequence, it cannot be excluded that there were unknown repetitive sequences that cannot be avoided in the process of overgo design (Romanov et al. 2003)

The current study is based on the map of the F_2-population generated by the cross of RHA325 and HA342 and the BAC libraries from RHA325 (avoiding any genome differences) and HA383. So far, BAC-clones identified by probes in RHA325 and HA383 showed very different BAC-fingerprinting patterns indicating that the genomic organization of HA383 might be different from the one of RHA325, especially in the region around the *Rf1* locus.

The false positives, mapping to other genome regions, could arise from overgo sequences that represent duplicated regions in the genome. It is not uncommon for genomic probes that appear to be single copy to identify BACs on high-density filters that correspond to multiple genomic loci (Gardiner et al. 2004). This tendency goes with the findings of Helentjaris et al. (1988) that probes developed from low-copy regions in maize hybridized to more than one fragment in Southern hybridizations. The authors concluded that the genetic mapping of this duplicated loci revealed extensive genome duplication in maize. The presence of a large number of probes hybridizing to different loci in the sunflower genome has already been reported in several other studies (Berry et

al. 1995; Gentzbittel et al. 1995; Jan et al. 1998). And this can be explained by the fact that sunflower has an ancient polyploid origin (Sossey-Alaoui et al. 1998).

Surprisingly, the overgo probe of the RAPD marker OP-K13-454A identified BACs in the restorer library as well as in the maintainer library. The OP-K13-454A marker linked to the restorer gene has been successfully converted into a SCAR marker and has been tested in a set of 20 lines (Horn et al. 2003). Comparison of the PCR pattern of 11 restorer and nine maintainer lines of PET1 demonstrated that this marker was present in all restorer lines but absent in maintainer lines. Duplicated loci for this RAPD fragment could account for this ambiguity which might result in the cloning of different parts of the genome not linked to the gene *Rf1* as it was demonstrated in the current study for the two BAC-end probes 67N04 and 225D09. Other sunflower researchers (Berry et al. 1995; Gentzbittel et al. 1995; Jan et al. 1998) have highlighted that about 30% of the used probes detected duplicated loci throughout the sunflower genome. The proportion of these duplicated loci was almost identical although the three studies were independent and were not related with regard to the source of probes. These duplications might be the consequence of the postulated polyploid origin of *Helianthus annuus* L. and the substantial reorganization the sunflower genome has undergone since hybridization (Jackson and Murray 1983; Gentzbittel et al. 1994; Seiler and Rieseberg 1997; Jan et al. 1998; Sossey-Alaoui et al. 1998; Bert et al. 2001).

4.3 Genetic mapping and recombination frequencies at the *Rf1* locus

Chances for successful positional cloning depends to a large extend on the local relationship between physical and genetic distances (Ballvora et al. 2001). Genetic mapping is a relative measurement for a genome because the relationship between genetic distance and physical distance varies considerably among different genome regions (Wing et al. 1994).

4.3.1 Markers in the vicinity of the *Rf1* gene

The probability to obtain a marker within very close proximity to the target gene region depends on the number of polymorphic primers tested, the genomic size in map units, and the degree of sequence variation around the target region (Janeja et al. 2003). The distance of the target gene decreases as the proportion of detectable polymorphism increases (Michelmore et al. 1991; Tanksley et al. 1992). AFLPs are an efficient way to

obtain markers closely linked to the target gene because of their high number and the fact that no sequence information is needed (Bentolila and Hanson 2001; Touzet et al. 2004; Hagihara et al. 2005).

Among the 256 AFLP primer combinations used to saturate the region surrounding the *Rf1* locus, 55 primer combinations were polymorph. The mapping analyses revealed that four of these 55 AFLP markers selected by bulked segregant analyses, mapped on the same linkage group as *Rf1* locus. The new mapping of *Rf1* locus (Fig. 13) showed that the order of markers in this region is similar to the previous map (Fig. 4), except for the new marker E62M52-249A, which was mapped at 0.1 cM from the restorer gene pushed the other AFLP markers further away. Consequently, the AFLP marker E41M48-113A has been shifted from 1.7 cM to 3.6 cM away from the restorer gene. The markers OP-K13-454A and E32M36-155R, which were separated in the previous map, now cosegregate with the gene *Rf1*.This can be probably explained by two reasons. The first reason is that the map distances are influenced by neighbouring markers that in one case compress and in the other case stretch the corresponding map interval (Liebhard et al. 2003). The second reason is due to the small number of individuals used to develop the map because in the first mapping approach *Rf1* and the markers were mapped using 183 F_2-individuals, in contrast to the second mapping in which only 91 F_2-individuals were used.

Theoretically, the larger the population is the more accurate is the estimation of the genetic distances (Jean et al. 1997; Feng et al. 2005). This assumption has been verified by Feng and Jan (2008). Since they did not detect any recombination between the SSR (ORS1114) marker and *Rf4* gene restoring fertility in sunflower using 113 F_2-plants, they enlarged the population to 1,173 plants. As a result, the SSR marker ORS1114 could be mapped at a distance of 0.9 cM from the *Rf4* locus. Therefore, the genetic distance between the *Rf1* and the surrounding markers might be better tested in a larger population to refine the area map because the population size of 91 F_2-individuals representing 182 gametes is relatively small compared to those used in other map-based cloning experiments. For the map-based cloning of *Rpg1* gene from barley the number of gametes surveyed was 8,518 (Brueggeman et al. 2002). Stein et al. (2000) investigated 6,240 gametes to construct a *Triticum monococcum* BAC contig spanning the *Lr10* leaf rust resistance locus in wheat. Construction of a physical contig map spanning the major domestication locus *Q* in wheat has been performed using 930 gametes (Faris et al. 2003). However Spielmeyer et al. (2000) showed that an *Aegilops*

tauschii F_2-population of 58 individuals was enough to identify recombinants between most of the mapped markers on 1DS.

Moreover, meiotic recombination events are distributed unevenly throughout the eukaryotic genome which leads to distortions of genetic maps that can hinder the ability to isolate genes by a map-based technique (Mesbah et al. 1999; Petes 2001). Stein et al. (2000) studied the recombination frequency in a 230-kb region of the *Lr10* resistance locus in wheat and found that three recombination events occurred within 30-kb while no recombination could be identified in the rest of the region. Civardi et al. (1994) confirmed in the *al-sh2* interval in maize the hypothesis that the recombination events occur at a higher rate in gene-rich regions and no recombination occur in gene-poor regions (Gill et al. 1996a, 1996b; Weng et al. 2000; Sandhu et al. 2001). Sandhu et al. (2002) showed that the average gene density within the *Triticeae* gene-rich regions is about 10 to 20 genes per 100 kb and 15 to 25 in rice.

Another factor, which might affect recombination frequency, is the genetic background of the parents used to generate the F_2-population (Bentolila and Hanson 2001). Paterson et al. (1990) studied genetic intervals in tomato. They noticed a significant reduction in the size of a genetic interval in a BC_2F_2-population (134 cM) as compared to a BC_1-population (209 cM). This reduction in recombination was thought to be caused by reduced reciprocal exchange in heterozygous segments combined with a positive interference from flanking homozygous regions. Consequently, the additional 256 AFLP marker combinations were not sufficient to fill the gap between the markers and to saturate the region of the *Rf1* gene with markers. Failure in finding any molecular marker tightly linked to the *Rf1* gene may be attributed to low polymorphism in the chromosomal region and possibly few restriction/annealing sites for the primers (Janeja et al. 2003). The lack in finding markers displaying perfect co-segregation with *Rf1* in the present study might indicate normal meiotic recombination in the vicinity of *Rf1* (Janeja et al. 2003). It may be possible to find more closely linked markers using additional different marker techniques like SSR markers or retrotransposons tagged markers to target the region. It has been shown that sequence information and genomic resources of *Arabidopsis* could be used for marker development, map-based gene cloning and candidate gene identification in *Brassica* (Schmidt 2002; Snowdon and Friedt 2004). Rapid accumulation of markers represents a challenge for information management and merging of data in order to accumulate more valuable information for further research and better use of genetics.

4.4 Chromosome walking and contig validation

To assign contigs to the genetic map, the general rule is that one contig should have one genetic location. Chromosome walking requires efficient overlapping between the identified BAC-clones in each walking step for correct assembly and identification of new BAC-ends for the next walking step (Deng et al. 2001). Several fingerprinting methods have been developed for construction of physical maps using one up to five enzyme restrictions (Coulson et al. 1986; Gregory et al. 1997; Marra et al. 1997; Zhang and Wing 1997; Luo et al. 2003; Xu et al. 2004) and FPC (FingerPrinted Contig) is the most widely used program to assemble contig maps of several species from the BAC fingerprints generated by the computer simulations (Chang et al. 2001; Tao et al. 2001; Ren et al. 2003; Wu et al. 2004).

Chromosome walking requires the ability to isolate BAC-end fragments, which can be used as probes to rescreen the libraries. A method was developed for the isolation of the BAC-ends for chromosome walking by plasmid rescue which consists of ligation of restriction fragments of BAC-clones digested with *Bam*HI into a universal cloning vector followed by a double antibiotics selection. The ends of the positive BACs from both libraries have been sequenced and used to develop new overgo probes. The BAC-clones picked by colony hybridisation were examined by BAC fingerprinting with *Hind*III digestion to built overlapping BAC-clones spanning the *Rf1* locus. In this study, contigs were assembled using fingerprints and hybridization patterns. Unfortunately, our contig assembly revealed several conflicts.

4.4.1 BAC contig assembly

BAC contigs were assembled manually using one restriction enzyme *Hind*III. The fragments generated by fingerprinting were assembled according to their size. In each case of library screening with the probes, several hits were observed and a number of contigs have been constructed. Unfortunately, BAC-clones detected by the BAC-ends did not show any overlapping between the contigs and no complete contig around the *Rf1* restorer gene could be established. Same problems have been encountered by Cai et al. (1997). They could not construct a contig around the gene $Hs1^{pro\text{-}1}$ resistance against nematode in sugar beet. These gaps may reflect an underrepresentation of probes from this region or, alternatively, an uneven distribution of recombination events (Gentzbittel et al. 1995). The same result has been found by Rajesh et al. (2004). They did not detect

overlaps between the BAC-clone from the end probe and the two identified positive BAC-clones associated with disease resistance in chickpea. For further studying the relationship between the two positive BAC-clones they designed PCR primers from both end sequences (left and right). They found out that the amplification pattern indicated that these two BAC-clones overlapped at the interior sequences.

Chromosome walking requires efficient analysis of the overlap between the BAC-clones identified in each walking step for correct assembly and identification of new BAC-ends for the next walking step. The fingerprinting procedure based on agarose gel electrophoresis and ethidium-bromide staining, as well as Southern hybridization used in our study to assemble BAC contigs were not sufficient to find overlaps between the contigs. To overcome these difficulties, it might be useful to use high-throughput fingerprinting and fingerprint analysis (Zhang and Wing 1997). This procedure is based on sequencing gel electrophoresis and ^{33}P-labelling of DNA fragment ends. When fingerprints are analyzed using IMAGE and FPC software program, it is easier to evaluate the overlaps of numerous positive BAC-clones, to assemble reliable contigs and to select BAC-ends for the next chromosome walking step (Deng et al. 2001).

4.4.2 Development of probes for physical mapping

The development of probes from the BAC-ends still proved to be very difficult and was frequently not successful. In this study 27 BAC-ends were isolated (Tab. 7). Only nine BAC-end sequences (Tab. 5) were used to design overgos to re-screen the BAC libraries by colony hybridization. Two BAC-ends (450B06-B and 261D17-B) were repetitive sequences and could not be used for physical mapping. In most cases, the probes, which did not contain repetitive sequences, did not detect overlapping BACs. Hohmann et al. (2003) could orient BACs by cross-hybridisation with selected BAC-ends, but in this study the cross-hybridization did not show any overlapping between the contigs obtained, so far. This might be due to the fact that the sequence of BAC-ends was not fully sequenced when overgos were constructed or the fact that only one BAC-end was isolated when using our method of BAC-end isolation. These questions remain open because we did not try any other method of BAC-end isolation. Even though the BAC-ends were of limited use as probes in this study, the availability of BAC-end probes proved to be essential, because few cloned AFLP markers could be used as probes for hybridisation against the high-density BAC filters. Problems with the cloned

AFLP markers as probes can be explained by the repetitive nature of most AFLPs, which has been demonstrated by Reamon-Büttner et al. (1999).

To close the physical gaps between the identified contigs, a series of approaches can be applied, which include identification of more AFLP markers immediately adjacent to physical gaps, development of STS-markers from BAC-end sequences, and screening of BAC pools with STS markers designed from BAC-end sequences. AFLP markers are AT rich in the recognition sequence because *Eco*RI recognises the sequence 5'-GAATTC-3', while *Mse*I recognises 5' -TTAA-3'. This means that these enzymes will cut more frequently in AT-rich regions. Using AFLP markers obtained with different enzymes would perhaps be a better strategy to achieve a more balanced coverage of the region (Reamon-Büttner et al. 1999).

Use of several complementary BAC libraries developed with different restriction enzymes is also an efficient strategy for minimizing "gaps" in the physical maps because such libraries are balanced in distribution of BAC-clones in the genome (Tao et al. 2001). A similar strategy was used for the physical mapping of *A. thaliana* (Marra et al. 1999) and soybean (Wu et al. 2004).

4.5 Concluding remarks and future prospects

In retrospect, the use of different genotypes for the genetic linkage map and the construction of BAC libraries could account for some of the difficulties experienced in integrating the physical and genetic maps. The two BAC libraries seem to be too small for complete coverage of the *Rf1* region. The low number of positive BAC-clones obtained from the BAC libraries due to the small coverage resulted in a lack of contig formations. The use of BAC libraries originated from different lines is only feasible if common markers are available.

The resolution and density of markers in the currently available genetic map was not sufficient for the development of a closed contig around the restorer gene. Extensive molecular markers should be developed to further narrow the target gene in a small genomic region. Abundant and highly efficient markers will help to generate information on the organization of *Rf1* in that particular region, and to generate tightly linked additional markers can greatly reduce the problem of physical chromosome walking (Ling et al. 2003) and will largely facilitate cloning of the *Rf1* gene. Incremental time-consuming processes such a new marker development and repeated

library walking as well as clone characterization and comparison are required for contig extension and gap closure. Contig gap closure will be significantly time consuming in a region poorly covered by markers (Cao et al. 1999).

All linkage maps are unique and are a product of the mapping population (derived from specific parents) and the types of markers used. Even if the same set of markers is used to construct linkage maps, there is no guarantee that molecular markers developed in one population might show polymorphism in another population (Collard et al. 2005; Lei et al. 2007). A detailed analysis of the genome structure in the region of *Rf1* is needed to study the relationship between genetic and physical distance (Zhang et al. 2006).

Molecular markers are now integrated into plant breeding programmes to allow breeders to access, transfer and combine genes at a rate with a precision not previously possible. However, one of the major drawbacks is when the linked marker used for selection is at distance away from the gene of interest, leading to crossovers between the marker and the gene. This can produce a high percentage of false-positives/negatives in the screening process and a marker developed for a gene in one cross may not be useful in other crosses unless the marker is derived from the gene itself (Mohan et al. 1997).

The current 9.7 cM sunflower genetic region is of sufficient resolution to localize chromosomal region containing the *Rf1* gene on linkage group 13, but it is not sufficient to allow chromosome walking using BAC libraries. The available AFLP map and the two BAC libraries did not provide a great framework for cloning the restorer gene because many steps are required and the success depends on the size and the complexity of the sunflower genome. More markers and a higher coverage of the genome by the BAC libraries are necessary to proceed in cloning the restorer gene *Rf1*.

5 Summary

Sunflower (*Helianthus annuus* L.) is an economically important oil seed crop. High seed and oil yields are the main objectives of sunflower breeding, and the utilization of heterosis or hybrid vigor has been one of the most important tools for high yield breeding. In sunflower, hybrids are obtained by using cytoplasmic male sterility (CMS). Use of CMS-systems eliminates the need for hand emasculation and ensures large-scale hybrid production. CMS can arise spontaneously, induced by intraspecific and interspecific crosses or by mutagenesis. Whatever its origin, it is generally assumed that CMS is due to a nuclear-cytoplasmic incompatibility.

In sunflower, commercial hybrid breeding is based on a single source of male sterility, PET1, originating from an interspecific cross between *Helianthus petiolaris* Nutt and *Helianthus annuus* L. Alterations in the mitochondrial genome between PET1 and fertile lines are limited to a 17-kb region, which consists of two mutations: a 12-kb-inversion and a 5-kb-insertion. CMS is associated with the expression of a novel open reading frame, *orf*H522, which encodes a 16-kDa polypeptide. Male fertility can be restored by the introduction of dominant nuclear restorer genes (*Rf*). The isolation of *Rf* genes should give insight into its origin, and the mechanism behind its action. The *Rf1* gene, responsible for fertility restoration of the PET1-cytoplasm, was mapped using 183 F_2-individuals based on AFLP-, RADP- and SSR-markers.

The objective of this study was to use the markers linked to the *Rf1* gene as probes in a map-based cloning strategy to isolate the restorer gene. Overgos derived from the markers were used to hybridize against two sunflower BAC libraries (RHA325 and HA383) (i) to identify positive BAC-clones, (ii) to isolate BAC-ends and to walk towards the *Rf1* gene and (iii) to identify overlapping DNA fragments that can be assembled into a contig spanning the *Rf1* region.

Seventy-one positive BAC-clones were obtained by colony hybridization using 20 overgos as probes against the two BAC libraries. The overgo probes were designed from three sources including AFLP-, RAPD-markers and BAC-end sequences.

A wide range in the number of BAC-clones hit by the 20 overgos was observed. Four overgos (two derived from RAPD markers and two derived from BAC-ends) failed to hit any clones. Five overgos (three derived from AFLP markers and two derived from BAC-end sequences) proved to be repetitive. E62M52-249A and 261F19-B-end overgos hit BAC-clones in the maintainer BAC library (HA383) but did not hit any

BAC-clone in the restorer BAC library (RHA325). In contrast E32M36-155R hit two positive BAC-clones in restorer BAC library but did not hit any BAC-clone in maintainer BAC library. This could be due to the intrinsic differences in the chromosomal structure between the two BAC libraries, or a different representation of genome regions. As results 71 BAC-clones were confirmed by fingerprint hybridization, which corresponded to a rate of 73% of positive BACs identified. The average number of BACs that were positive for each non-repetitive overgo probe ranged from one to 17 BACs. The size of isolated BAC inserts were more accurately estimated with PFGE using *Not*I restriction endonuclease than with fingerprinting method using *Hind*III, especially with BACs having sizes higher than 100 kb. Furthermore, 27 BAC-ends were isolated out of 29 positive BAC-clones. For this, a modified plasmid end rescue method was used to obtain the BAC-ends. The method is based on the ligation of restriction fragments of BAC-clones digested with *Bam*HI into a universal cloning vector followed by double antibiotics selection. The size of the subcloned BAC-ends ranged between 1.2 kb to 15.2 kb. In total, six markers (OP-K13-454A, E62M52-249A, E44M70-250A, E42M76-125A, E33M61-136R, E41M48-113A) and four BAC-end sequence probes (67N04-B, 225D09-B, 225D9-sf$_2$, and 261F19-B) were used to construct a contig surrounding the *Rf1* region.

Southern hybridization using BAC-ends as probes did not identify overlaps between BAC contigs. To further examine the overlap of BAC contigs, cross-hybridization experiments were performed. Contigs were constructed manually by fingerprinting of the positive BAC-clones digested with *Hind*III. All in all, 38 BAC-clones encompassing six markers surrounding the *Rf1* region were selected and formed seven contigs and 15 singletons. Contig from RHA325 contained only three BAC-clones whereas contigs from HA383 contained two to 11 BAC-clones. The physical map in the region of the *Rf1* gene revealed gaps between the BAC contigs. No contiguous regions have been found as each contig appeared to represent a unique genomic region. Additionally, the sequences of all BAC-ends were subjected to an automated contig assembly using VectorNTI software (Invitrogene) for sequence analyses and data management. BAC-end sequences of 115P09 (from RHA325) and 216F17, 384H08 (from HA383) overlapped. However, the investigation has not been continued because the 225D09 BAC-end was back-mapped into a different linkage group than the *Rf1* gene. A second BAC-end, 67N04-B was back-mapped in the F_2-population of the cross RHA325×HA342. This BAC-end back-mapped 3.9 cM away from the restorer gene.

The sequences of BAC-ends were subjected to a BLASTN (Nucleotide search) and BLASTX (Protein search). The nucleotide search resulted at $2e^{91}$ to homology to an anther specific gene *sf2*.

Although both BAC libraries (RHA325 and HA383) together represent more than 10-fold genome coverage the number of BAC-clones detected by the overgo probes was too low to be able to form a closed contig around the restorer gene.

6 Zusammenfassung

Die Sonnenblume (*Helianthus annus* L.) zählt zu den ökonomisch wichtigsten Ölpflanzen. Hohe Saatgut- und Ölerträge sind die vorrangigen Ziele in der Sonnenblumenzüchtung. Die systematische Nutzung von Heterosiseffekten in der Hybridzüchtung stellt das wichtigste Mittel zum Erreichen dieser Ziele dar. Bei der Sonnenblume werden Hybride auf der Grundlage der cytoplasmatischen männlichen Sterilität (CMS) erzeugt. CMS-Systeme erleichtern die Saatgutproduktion von Hybriden erheblich, da die Notwendigkeit einer Handkastration entfällt und trotzdem gewährleistet werden kann, dass es sich um Kreuzungssaatgut handelt. CMS kann entweder spontan auftreten oder durch intra-und interspezifische Kreuzungen sowie durch Mutagenese induziert werden. Ungeachtet ihrer Ursachen beruht die CMS auf einer cytoplasmatisch-kerngenischen Inkompatibilität. Bei der Sonnenblume basieren die kommerziellen Hybriden auf einem einzigen CMS-System, dem PET1-Plasma, das auf eine interspezifische Kreuzung *H. petiolaris* Nutt × *H. annuus* L. zurückgeht. Die Mitochondriengenome männlich-steriler und fertiler Linien unterscheiden sich nur durch eine 12-kb groβe Inversion und eine 5-kb groβe Insertion. CMS ist assoziiert mit der Expression eines neuen offenen Leserahmens, *orf*H522, der für ein 16-kDa Protein kodiert. Die Restauration der Pollenfertilität wird über dominante Restorergene (*Rf*) im Kerngenom erreicht. Mittels Isolierung der *Rf*-Gene können Erkenntnisse über deren Herkunft und Wirkungsweise gewonnen werden. Das *Rf1*-Gen, welches für die Fertilitätsrestauration des PET1-Zytoplasmas verantwortlich ist, konnte mittels 183 F_2-Individuen und unter Verwendung von AFLP- RAPD- und SSR-Markern kartiert werden. Ziel der vorliegenden Arbeit war es, mittels markergestützter Klonierung das *Rf1*-Gen zu isolieren. Aus den mit dem *Rf1*-Gen eng gekoppelten AFLP- und RAPD-Markern wurden Overgo-Sonden entwickelt, die für die Kolonie-Hybridisierungen gegen zwei BAC-Bibliotheken (RHA325 und HA383) eingesetzt wurden. Im Weiteren sollten folgende Schritte durchgeführt werden (i) Gewinnung positiver BAC-Klone, (ii) Klonierung der BAC-Enden und „Chromosome Walking“ in Richtung des Gens sowie (iii) Entwicklung eines Contigs um das *Rf1*-Gen.

In der vorliegenden Doktorarbeit wurden 20 Overgos aus Markern und BAC-End-Sequenzen entwickelt.

Die verwendeten 20 Overgos zeigten deutliche Unterschiede hinsichtlich der Anzahl der BAC-Klone, die identifiziert werden konnten. Vier Overgos (zwei abgeleitet von

RAPD-Markern und zwei aus BAC-Enden) erzielten keine Treffer, fünf dagegen (drei abgeleitet von AFLP-Markern und zwei aus BAC-End-Sequenzen) waren repetitiv. Die Overgos, abgeleitet von E62M52-249A und 261F19-B-Ende, detektierten Klone in der Maintainer-Bibliothek (HA383) aber keine in der Restorer-Bibliothek (RHA325). Im Unterschied dazu identifizierte E32M36-155R zwei Klone in der RHA325 Bibliothek und keine in HA383 Bibliothek. Dies könnte auf die den beiden BAC-Bibliotheken innewohnenden unterschiedlichen chromosomalen Strukturen zurückzuführen sein oder auf eine unterschiedliche Repräsentation der Genomanteile in den BAC-Bibliotheken. Insgesamt konnten 71 BAC-Klone durch Fingerprinting und Hybridisierung bestätigt werden. Die Anzahl positiver BAC-Klone variierte zwischen eins und 17, mit einer durchschnittlichen Anzahl von vier positiven BAC-Klonen pro Overgo-Sonde. Die Insertgröβe (*Not*I verdaut) konnte mittels PFGE genauer bestimmt werden als über die Fingerprinting-Methode (*Hind*III verdaut), insbesondere bei BAC-Klonen mit einer Insertgröβe von mehr als 100 kb. Ferner konnten aus 29 positiven BAC-Klonen 27 BAC-Enden isoliert werden. Hierbei wurden die BAC-Enden durch eine modifizierte Plasmid-Ende-Rescue-Methode gewonnen. Diese Methode basiert darauf, dass Restriktions-fragmente von BAC-Klonen mit *Bam*HI verdaut und in einen Universal Vektor ligiert werden mit einer nachfolgenden doppelten Antibiotika-Selektion. Die Gröβe der klonierten BAC-Enden lag zwischen 1.2 kb und 15.2 kb. Zum Aufbau eines geschlossenen Contigs um das *Rf1*-Gen, wurden sechs Marker (OP-K13-454A, E62M52-249A, E44M70-250A, E42M76-125A, E33M61-136R und E41M48-113A) und vier BAC-End-Sonden (67N04-B, 225D09-B, 225D09-sf_2 und 261F19-B) gegen die zwei BAC-Bibliotheken hybridisiert. Mit Hilfe der Southern-Hybridisierung durch BAC-End-Sonden konnten keine Überlappungen von BAC-Klonen nachgewiesen werden. Zur weiteren Untersuchung der Überlappung wurden Kreuz-Hybridisierungsexperimente durchgeführt. Die Contigs wurden manuell durch Fingerprinting der positiven BAC-Klone (*Hind*III verdaut) erzeugt. Alles in allem wurden 38 BAC-Klone selektiert, die sechs Marker und den *Rf1*-Genbereich einschlossen und 15 Singleton sowie 7 Contigs darstellten. Die aus RHA325 gewonnenen Contig enthieltet nur drei Klone, die aus HA383 gewonnenen Contigs hingegen zwei bis 11 Klone. Die physikalische Karte im Bereich des *Rf1*-Gens offenbarte Lücken zwischen den Contigs. Zusammenhängende Bereiche wurden nicht gefunden, da jeder Contig eine einzigartige genomische Region darzustellen schien. Zusätzlich wurden die Sequenzen aller BAC-Enden mit Hilfe des Software Programms

VectorNTI (Invitrogen) auf Contig-Überlappung hin untersucht. Es stellte sich heraus, dass die Sequenzen der BAC-Enden der Klone 115P09 (aus RHA325) und 225D09, 216F17, 384H08 (aus HA383) überlappten. Die Untersuchungen wurden allerdings nicht weitergeführt, weil das BAC-Ende von 225D09 in eine Kopplungsgruppe rückkartierte, die sich von der des *Rf1*-Gens unterschied. Das BAC-Ende von 67N04-B, das in die F_2-Population der Kreuzung RHA325 × HA342 rückkartiert wurde, kartierte mit 3,9 cM vom Restorergen entfernt. Die BAC-End-Sequenzen wurden mit Hilfe des Programms BLAST (Basic Local Alignment Search Tool) auf Sequenzhomologien (BLASTN für den Nukleotidabgleich, BLASTX für den Proteinabgleich) hin untersucht. BLASTN zeigte Homologie auf $2e^{91}$ zu dem antheren-spezifischen *sf2* Gen. Obwohl mit beiden BAC-Bibliotheken zusammen eine mehr als 10-fache Genomabdeckung erreicht wurde, war die Anzahl der identifizierten BAC-Klone zu gering, um damit einen geschlossenen Contig um das Restorergen *Rf1* bilden zu können.

7. References

Adams MD, Soares MB, Kerlavage AR, Fields C, Venter C (1993) Rapid cDNA sequencing (expressed sequence tags) from a directionally cloned human infant brain cDNA library. Nat Genet 4: 73-380

Akagi H, Nakamura A, Yokozeki-Misono Y, Inagaki A, Takahashi H, Mori K, Fujimura T (2004) Positional cloning of the rice *Rf*-1 gene, a restorer of BT-type cytoplasmic male sterility that encodes a mitochondria-targeting PPR protein. Theor Appl Genet 108: 1449-1457

Akhunov ED, Akhunova AR, Dvorak J (2005) BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. Theor Appl Genet 111: 1617-1622

Al-Chaarani GR, Gentzbittel L, Huang XQ, Sarrafi A (2004) Genotypic variation and identification of QTLs for agronomic traits, using AFLP and SSR markers in RILs of sunflower (*Helianthus annuus* L.). Theor Appl Genet 109: 353-360

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215: 403-410

Arondel V, Lemieux B, Hwang I, Gibson S, Goodman HM, Somerville CR (1992) Map-based cloning of a gene controlling *Omega-3* fatty acid desaturation in *Arabidopsis.* Science 258: 1353-1355

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

Aubourg S, Boudet N, Kreis M, Lecharny A (2000) In *Arabidopsis thaliana*, 1% of the genome codes for a novel protein family unique to plants. Plant Mol Biol 42: 603-613

Balk J, Leaver CJ (2001) The PET1-CMS mitochondrial mutation in sunflower is associated with premature programmed cell death and cytochrome *c* release. Plant Cell 13: 1803-1818

Ballvora A, Schornack S, Baker BJ, Ganal M, Bonas U, Lahaye T (2001) Chromosome landing at the tomato *Bs4* locus. Mol Genet Genom 266: 639-645

Bateson W, Gairdner AE (1921) Male-sterility in flax. J Genet 11: 269-275

Ben C, Hewezi T, Jardinaud MF, Bena F, Ladouce N, Moretti S, Tamborindeguy C, Liboz T, Petitprez M, Gentzbittel L (2005) Comparative analysis of early embryonic sunflower cDNA libraries. Plant Mol Biol 57: 255-270

Bentolila S, Alfonso AA, Hanson MR (2002) A pentatricopeptide repeat-containing gene restores fertility to cytoplasmic male-sterile plants. Proc Natl Acad Sci USA 99: 10887-10892

Bentolila S, Hanson MR (2001) Identification of a BIBAC clone that co-segregates with the petunia restorer of fertility (*Rf*) gene. Mol Genet Genom 266: 223-230

Berrios EF, Gentzbittel L, Mokrani L, Alibert G, Sarrafi A (2000a) Genetic control of early events in protoplast division and regeneration pathways in sunflower. Theor Appl Genet 101: 606-612

Berrios EF, Gentzbittel L, Kayyal H, Alibert G, Sarrafi A (2000b) AFLP mapping of QTLs for in vitro organogenesis traits using recombinant inbred lines in sunflower (*Helianthus annuus* L.). Theor Appl Genet 101: 1299-1306

Berrios EF, Sarrafi A, Fabre F, Alibert G, Gentzbittel L (2000c) Genotypic variation and chromosomal location of QTLs for somatic embryogenesis revealed by epidermal layers culture of recombinant inbred lines in the sunflower (*Helianthus annuus* L.) Theor Appl Genet 101: 1307-1312

Berry ST, León AJ, Peerbolte R, Challis C, Livini C, Jones R, Feingold S (1997) Presentation of the Advanta sunflower RFLP linkage map for public research. In: Proc 19th Sunflower Research Workshop, Fargo, ND, National Sunflower Association, Bismarck, ND, pp 113-118

Berry ST, León AJ, Challis P, Livini C, Jones R, Hanfrey CC, Griffiths S, Roberts A (1996) Construction of a high density, composite RFLP linkage map for cultivated sunflower (*Helianthus annuus* L.). In: Proc 14th Int Sunflower Conf, vol. 2, Beijing, China, International Sunflower Association, Paris, pp 1155-1160

Berry ST, León AJ, Hanfrey CC, Challis P, Burkholz A, Barnes SR, Rufener GK, Lee M, Caligari PDS (1995) Molecular marker analysis of *Helianthus annuus* L. 2. Construction of an RFLP linkage map for cultivated sunflower. Theor Appl Genet 91: 195-199

Bert PF, Tourvieille de Labrouhe D, Philippon J, Mouzeyar S, Jouan I, Nicolas P, Vear F (2001) Identification of a second linkage group carrying genes controlling resistance to downy mildew (*Plasmopara halstedii*). Theor Appl Genet 103: 992-997

Birnboim HC, Doly J (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucl Acids Res 7: 1513-1523

Bohuon E, Keith D, Parkin I, Sharpe A, Lydiate D (1996) Alignment of the conserved C genomes of *Brassica oleracea* and *Brassica napus*. Theor Appl Genet 93: 5-6

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

Bouzidi MF, Franchel J, Tao Q, Stormo K, Mraz A, Nicolas P, Mouzeyar S (2006) A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions. Theor Appl Genet 113: 81-89

Brahm L, Röcher T, Friedt W (2000) PCR-Based markers facilitating marker assisted selection in sunflower for resistance to downy mildew. Crop Sci 40: 676-682

Brown GG, Formanova N, Jin H, Wargachuk R, Dendy C, Patil P, Laforest M, Zhang J, Cheung WY, Landry BS (2003) The radish *Rfo* restorer gene of *Ogura* cytoplasmic male sterility encodes a protein with multiple pentatricopeptide repeats. Plant J 35: 262-272

Brueggeman R, Rostoks D, Kudrna A, Killan F, Han F, Chen J, Druka B, Steffenson B, Kleinshofs A (2002) The barley stem rust-resistance gene *Rpg1* is a novel disease resistance gene with homology to receptor kinases. Proc Natl Acad Sci USA 99: 9328-9333

Bruno WJ, Knill E, Balding DJ, Bruce DC, Doggett NA, Sawhill WW, Stallings RL, Whittaker CC, Torney DC (1995) Efficient pooling designs for library screening. Genomics 26: 21-30

Budar F, Touzet P, De Paepe R (2003) The nucleo-mitochondrial conflict in cytoplasmic male sterilities revisited. Genetica 117: 3-16

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

Burke JM, Lai Z, Salmaso M, Nakasato T, Tang S, Heesacker A, Knapp ST, Rieseberg LH (2004) Comparative mapping and rapid karyotypic evolution in the genus *Helianthus*. Genetics 167: 449-457

Burke JM, Tang S, Knapp ST, Rieseberg LH (2002) Genetic analysis of sunflower domestication. Genetics 161: 1257-1267

Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. Science 236: 806-812

Burr B, Burr FA (1991) Recombinant inbreeds for molecular mapping in maize: Theoretical and practical considerations. TIG February 7: 56-60

Cai D, Kleine M, Kifle S, Harloff H, Sandal NN, Marcker KA, Klein-Lankhorst RM, Salentijin ESJ, Lange W, Stieckema W, Wyss U, Grundler F, Jung C (1997) Positional cloning of a gene for nematode resistance in sugar beet. Science 275: 832-834

Cai L, Taylor JF, Wing RA, Gallagher DS, Woo SS, Davis SK (1995) Construction and characterization of a bovine bacterial artificial chromosome library. Genomics 29: 413-425

Cai W-W, Renecker J, Chow C-W, Vaishnav M, Bradley A (1998) An anchored framework BAC map of mouse chromosome 11 assembled using multiplex oligonucleotide hybridization. Genomics 54: 387-397

Cao Y, Kang HL, Xu X, Wang M, Dho SH, Huh JR, Lee BJ, Kalush F, Bocskai D, Ding Y, Tesmer JG, Lee J, Moon E, Jurecic V, Baldini A, Weier H-U, Doggett NA, Simon MI, Adams MD, Kim U-J (1999) A 12-Mb complete coverage BAC contig map in human chromosome 16p13.1-pll.2. Genome Res 9: 763-774

Carrera AD, Pizzaro G, Poverene M, Feingold S, León AJ, Berry ST (2002) Variability among inbred lines and RFLP mapping of sunflower isozymes. Genet Mol Biol 25: 65-72

Cato SA, Gardener RC, Richardson TE (2001) A rapid PCR method for genetically mapping ESTs. Theor Appl Genet 102: 296-306

Cellier F, Conéjéro G, Breitler JC, Casse F (1998) Molecular and physiological responses to water deficit in drought-tolerant and drought-sensitive lines of sunflower. Plant Physiol 116: 319-328

Chang Y-L, Henriquez X, Preuss D, Copenhaver GP, Zhang H-B, (2003) A plant transformation-competent binary BAC library of the *Arabidopsis thaliana* Landsberg ecotype for functional and comparative genomic. Theor Appl Genet 106: 269-276

Chang Y-L, Tao Q, Scheuring C, Ding K, Meksem K, Zhang H-B (2001) An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. Genetics 159: 1231-1242

Chase CD (2006) Cytoplasmic male sterility: a window to the world of plant mitochondrial-nuclear interactions. Trends Genet 23: 81-90

Chen J, Hu J, Vick BA, Jan CC (2006) Molecular mapping of a nuclear male-sterility gene in sunflower (*Helianthus annuus* L.) using TRAP and SSR markers. Theor Appl Genet 113: 122-127

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002) An integrated physical and genetic map of the rice genome. Plant Cell 14: 537-545

Chen Q, Sun S, Ye Q, McCuine S, Huff E, Zhang H-B (2004) Construction of two BAC libraries from the wild Mexican diploid potato, *Solanum pinnatisectum,* and the identification of clones near the late and Colorado potato beetle resistance loci. Theor Appl Genet 108: 1002-1009

Christov M (1993) Sources of cytoplasmic male sterility produced at IWS "Dobroudja". Biotechnol and Biotechnol Eq 7: 132-135

Christov M (1990) A new source of cytoplasmic male sterility in sunflower originating from *Helianthus argophyllus*. Helia 13: 55-61

Civardi L, Xia Y, Edwards KJ, Schnable PS, Nikolau B (1994) The relationship between genetic and physical distances in the cloned *al-sh2* interval of the *Zea mays* L. genome. Proc Natl Acad Sci USA 91: 8286-8272

Clarke B, Lambrecht M, Rhee SY (2003) *Arabidopsis* genomic information for interpreting wheat EST sequences. Funct Integr Genomics 3: 33-38

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. Euphytica 142: 169-196

Collins J, Hohn B (1978) Cosmids: a type of plasmid gene-cloning vector that is packageable *in vitro* in bacteriophage λ heads. Proc Natl Acad Sci USA 75: 4242-4246

Cone KC, McMullen MD, Vroh Bi I, Davis GL, Yim YS, Gardiner JM, Polacco ML, Sancherz-Villeda H, Fang Z, Schroeder SG, Havermann SA, Bowers JE, Paterson AH, Soderlund CA, Engler FW, Wing RA, Coe Jr. EH (2002) Genetic, physical, and informatics resources for maize. On the road to an integrated map. Plant Physiol 130: 1598-1605

Coulson A, Sulston J, Brenner S, Karn J (1986) Toward a physical map of the genome of the nematode *Caenorhabditis elegans.* Proc Natl Acad Sci USA 83: 7821-7825

Crouzillat D, de la Canal L, Perrault A, Ledoigt G, Vear F, Serieys H (1991) cytoplasmic male sterility in sunflower: comparison of molecular biology and genetic studies. Plant Mol Biol 16: 415-426

Cui X, Wise PR, Schnable PS (1996) The *rf2* nuclear restorer gene of male-sterile T-cytoplasm maize. Science 272: 1334-1336

Danesh D, Peñuela S, Mudge J, Denny RL, Nordstrom H, Martinez JP, Young ND (1998) A bacterial artificial chromosome library for soybean and identification of clones near a major cyst nematode resistance gene. Theor Appl Genet 96: 196-202

Deng Z, Huang S, Ling P, Yu C, Tao Q, Chen C, Wendell MK, Zhang H-B, Gmitter Jr. FG (2001) Fine genetic mapping and BAC contig development for the citrus *tristeza* virus resistance gene locus in *Poncirus trifoliate (*Raf). Mol Genet Genom 265: 739-747

Desloire S, Gherbi H, Laloui W, Marhadour S, Clouet V, Cattolico L, Falentin C, Giancola S, Renard M, Budar F, Small I, Caboche M, Delourme R, Bendahmane A (2003) Identification of the fertility restoration locus, *Rfo*, in radish, as a member of the pentatricopeptide-repeat protein family. EMBO Reports 4: 588-594

Domon C, Steinmetz A (1994) Exon shuffling in anther-specific gene in sunflower. Mol Gen Genet 244: 312-317

Domon C, Evrard JL, Herdenberger F, Pillay DTN, Steinmetz A (1990) Nucleotide sequence of two anther-specific cDNAs from sunflower (*Helianthus annuus* L.). Plant Mol Biol 15: 643-646

Dorell DG, Vick BA (1997) Properties and processing of oilseed sunflower. In: Schneiter AA (ed) Sunflower technology and production, Madison, Wisconsin, USA, pp 709-745

Dudareva N, Reddy JT, Kräuter R, Bronner R, Steinmetz A, Pillay DTN (1996) A flower-specific gene family whose expression is regulated temporally and spatially during flower development in sunflower. Plant Sci 120: 161-173

Durante M, Vannozzi GP, Pugliesi C, Bernardi R (2002) The role of biotechnologies in the development of sunflower cultures in the world. Helia 25: 1-28

Durett RT, Chen KY, Tanksley SD (2002) A simple formula useful for positional cloning. Genetics 160: 353-355

Enns H, Dorell DG, Hoes JA, Chubb WO (1970). Sunflower research, a progress report. Proc of the 4th Int Sunflower Conf, ISA, CETIOM Paris (eds), Memphis Tennessee, USA, pp 162-167

Fang X, Gu S, Xu Z, Chen F, Guo D, Zhang H-B, Wu N (2004) Construction of a binary BAC library for an apomictic monosomic addition line of *Beta corolliflora* in sugar beet and identification of the clones derived from the alien chromosome. Theor Appl Genet 108: 1420-1425

Faris JD, Fellers JP, Brooks SA, Gill BS (2003) A bacterial artificial chromosome contig spanning the major domestication locus *Q* in wheat and identification of a candidate gene. Genetics 164: 311-321

Feng CD, Stewart JMcD, Zhang J-F (2005) STS markers linked to the *Rf1* fertility restorer gene of cotton. Theor Appl Genet 110: 237-243

Feng J, Jan CC (2008) Introgression and molecular tagging of Rf_4, a new male fertility restoration gene from wild sunflower *Helianthus maximiliani* L. Theor Appl Genet 117: 241-249

Feng J, Vick BA, Lee MK, Zhang H-B, Jan CC (2006) Construction of BAC and BIBAC libraries from sunflower and identification of linkage group-specific clones by overgo hybridization. Theor Appl Genet 113: 23-32

Fernández P, Paniego N, Lew S, Hopp HE, Heinz RA (2003) Differential representation of sunflower ESTs in enriched organ-specific cDNA libraries in a small scale sequencing project. BMC Genomics 4: 40

Ferreira A, da Siva MF, da Costa e Siva L, Cruz CD (2006) Estimating the effects of population size and type on the accuracy of genetic maps. Genet Mol Biol 29: 187-192

Fick GN, Zimmer DE (1974) Fertility restoration in confectionery sunflowers. Crop Sci 14: 603-604

Frijters ACJ, Zhang Z, van Damme M, Wang G-L, Ronald PC, Michelmore RW (1997) Construction of a bacterial artificial chromosome library containing large *Eco*RI and *Hind*III genomic fragments of lettuce. Theor Appl Genet 94: 390-399

Fu H, Dooner HK (2002) Intraspecific violation of genetic colinearity and its implications in maize. Proc Natl Acad Sci USA 99: 9573-9578

Gagliardi D, Leaver CJ (1999) Polyadenylation accelerates the degradation of the mitochondrial mRNA associated with cytoplasmic male sterility in sunflower. EMBO J 18: 3757-3766

Gardiner J, Schroeder S, Polacco ML, Sanchez-Villeda H, Fang Z, Morgante M, Landewe T, Fengler K, Useche F, Hanefey M, Tingey S, Chou H, Wing R, Soderlund C, Coe EH Jr. (2004) Anchoring 9.371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol 134: 1317-1326

Gedil MA, Wye C, Berry S, Segers B, Peleman J, Jones R, León A, Slabaugh MB Knapp SJ (2003) An integrated restriction fragment length polymorphism amplified fragment length polymorphism linkage map for cultivated sunflower. Genome 44: 213-221

Gedil MA, Slabaugh MB, Berry S, Johnson R, Michelmore R, Miller J, Gulya T, Knapp SJ (2001) Candidate disease resistance genes in sunflower cloned using conserved

nucleotide-binding site motifs: Genetic mapping and linkage to the downy mildew resistance gene *Pl1*. Genome 44: 205-212

Gentzbittel L, Abbott A, Galaud JP, Georgi L, Fabre F, Liboz T, Alibert G (2002) A bacterial artificial chromosome (BAC) library for sunflower, and identification of clones containing genes for putative transmembrane receptors. Mol Genet Genom 266: 979-987

Gentzbittel L, Mestries E, Mouzeyar S, Mazeyrat F, Badaoui S, Vear F, Tourvieille De Labrouhe D, Nicolas P (1999) A composite map of expressed sequences and phenotypic traits of the sunflower (*Helianthus annuus* L.) genome. Theor Appl Genet 99: 218-234

Gentzbittel L, Mouzeyar S, Badaoui S, Mestries E, Vear F, Tourvieille De Labrouche D, Nicolas P (1998) Cloning of molecular markers for disease resistance in sunflower, *Helianthus annuus* L. Theor Appl Genet 96: 519-525

Gentzbittel L, Vear F, Zhang YX, Bervillé A, Nicolas P (1995) Development of a consensus linkage RFLP map of cultivated sunflower (*Helianthus annuus* L.). Theor Appl Genet 90: 1079-1086

Gentzbittel L, Zhang YX, Vear F, Griveau B, Nicolas P (1994) RFLP studies of genetic relationships among inbred lines of the cultivated sunflower, *Helianthus annuus* L.: evidence for distinct restorer and maintainer germplasm pools. Theor Appl Genet 89: 419-425

Georgi LL, Wang Y, Yvergniaux D, Ormsbee T, Iñigo M, Reighard G, Abbott AG (2002) Construction of a BAC library and its application to the identification of simple sequence repeats in peach (*Prunus persical* (L.) Batsch). Theor Appl Genet 105: 1151-1158

Gill KS, Gill BS, Endo TR, Boyko EV (1996a) Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 144: 1883-1891

Gill KS, Gill BS, Endo TR, Taylor T (1996b) Identification and high-density mapping of gene-rich regions in chromosome group 5 of Wheat. Genetics 143: 1001-1012

Gillman JD, Bentolila S, Hanson MR (2007) The petunia restorer of fertility protein is part of a large mitochondrial complex that interacts with transcripts of the CMS-associated locus. Plant J 49: 217-227

Giordani T, Natali L, D´Ercole A, Pugliesi C, Fambrini M, Vernieri P, Vitagliano C, Cavallini A (1999) Expression of a dehydrin gene during embryo development and drought stress in ABA-deficient mutants of sunflower (*Helianthus annuus* L.). Plant Mol Biol 39: 739-748

Green ED (2001) Strategies for the systematic sequencing of complex genomes. Genetics 2: 573-583

Gregory SG, Howell GR, Bentley DR (1997) Genome mapping by fluorescent fingerprinting. Genome Res 7: 1162-1169

Guimaraes CT, Sills GR, Sobral BWS (1997) Comparative mapping of *Andropogoneae Saccharum* L. (sugarcane) and its relation to sorghum and maize. Proc Natl Acad Sci USA 94: 14261-14266

Gupta PK, Rustgi S (2004) Molecular markers from the transcribed/expressed region of the genome in higher plants. Funct Integr Genomics 4: 139-162

Hagihara E, Itchoda N, Habu Y, Mikami T Kubo T (2005) Molecular mapping of fertility restorer gene for Owen cytoplasmic male sterility in sugar beet. Theor Appl Genet 111: 250-255

Hamrit S, Kusterer B, Friedt W, Horn R (2008) Verification of positive BAC clones near the *Rf1* gene restoring pollen fertility in the presence of the PET1 cytoplasm in sunflower (*Helianthus annuus* L.) and direct isolation of BAC ends. In: Proc

17th Int Sunflower Conf. vol.2, Córdoba, Spain, International Sunflower Association, Paris, pp 623-628

Han CS, Sutherland RD, Jewett PB, Campbell ML, Meincke LJ, Tesmer JG, Mundt MO, Fawcett JJ, Kim UJ, Deaven LL, Doggett NA (2000) Construction of a BAC contig map of chromosome 16q by two-dimensional overgo hybridization. Genome Res 10: 714-721

Hanson MR, Bentolila S (2004) Interactions of mitochondrial and nuclear genes that affect male gametophyte development. Plant Cell 16: S154-S169

Hanson MR, Wilson RK, Bentolila S, Köhler RH, Chen H-C (1999) Mitochondrial gene organization and expression in petunia male fertile and sterile plants. J Heredity 90: 362-368

Harter AV, Gardner KA, Falush D, Lentz DL, Bye RA, Rieseberg LH (2004) Origin of extant domesticated sunflower in eastern North America. Nature 430: 201-205

Heiser CB (1982) Registration of Indiana-1 CMS sunflower germplasm. Crop Sci 22: 1089

Helentjaris T, Weber D, Wright S (1988) Identification of the genomic locations of duplicate nucleotide sequences in maize by analysis of restriction fragments length polymorphism. Genetics 118: 353-363

Hernández LF, Green PB (1993) Transductions for the expression of structural pattern: Analysis in sunflower. Plant Cell 5: 1725-1738

Hohmann U, Jacobs G, Telgmann A, Gaafar RM, Alam S, Jung C (2003) A bacterial artificial chromosome (BAC) library of sugar beet and a physical map of the region encompassing the bolting gene *B*. Mol Genet Genom 269: 126-136

Hong G (1997) A rapid and accurate strategy for rice contig map construction by combination of fingerprinting and hybridization. Plant Mol Biol 35: 129-133

Hongtrakul V, Slabaugh MB, Knapp SJ (1998a) A seed specific Δ-12 oleate desaturase gene is duplicated, rearranged, and weakly expressed in high oleic acid sunflower lines. Crop Sci 38: 1245-1249

Hongtrakul V, Slabaugh MB, Knapp SJ (1998b) DFLP, SSCP, and SSR markers for Δ9-stearoyl-acyl carrier protein desaturases strongly expressed in developing seeds of sunflower: intron lengths are polymorphic among elite inbred lines. Mol Breed 4: 195-203

Hongtrakul V, Huestis GM, Knapp SJ (1997) Amplified fragment length polymorphisms as a tool for DNA fingerprinting sunflower germplasm: genetic diversity among oilseed inbred lines. Theor Appl Genet 95: 400-407

Horn R (2006) Cytoplasmic male sterility and fertility restoration in higher plants. Progress in Botany 67: 31-52

Horn R, Hamrit S (2009) Gene cloning and characterization. In: Genetics, genomics and breeding in sunflower. Hu J, Seiler G (eds), in press

Horn R, Friedt W (1999) CMS sources in sunflower: different origin but same mechanism? Theor Appl Genet 98: 195-201

Horn R, Friedt W (1997) Fertility restoration of new CMS sources in sunflower. Plant Breed 116: 317-322

Horn R, Kusterer B, Lazarescu E, Prüfe M, Friedt W (2003) Molecular mapping of the *Rf1* gene restoring pollen fertility in PET1-based F_1 hybrids in sunflower (*Helianthus annuus* L.). Theor Appl Genet 106: 599-606

Horn R, Kusterer B, Lazarescu E, Prüfe M, Özdemir N, Friedt W (2002) Molecular diversity of CMS sources and fertility restoration in the genus *Helianthus*. Helia 25: 29-40

Horn R, Hustedt JEG, Horstmeyer A, Hahnen J, Zetsche K, Friedt W (1996) The CMS-associated 16kDa protein encoded by *orfH522* is also present in other male sterile cytoplasms of sunflower. Plant Mol Biol 30: 523-538

Horn R, Köhler RH, Zetsche K (1991) A mitochondrial 16 kDa protein is associated with cytoplasmic male sterility in sunflower. Plant Mol Biol 7: 29-36

Hu J (2006) Defining the sunflower (*Helianthus annuus* L.) linkage ends with the *Arabidobsis*-type telomere sequence repeat-derived markers. Chromosome Res 14: 535-548

Hu J, Yue B, Vick BA (2007) Integration of TRAP markers onto a sunflower SSR marker linkage map constructed from 92 recombinant inbred lines. Helia 30: 25-36

Hu J, Bervillé A, Vick BA (2004) High potential of TRAP markers in sunflower genome mapping. In: Proc 16th Int Sunflower Conf, Fargo, ND, USA, pp 665-671

Imai R, Koizuka N, Fujimoto H, Hayakawa T, Sakai T, Imamura J (2003) Delimitation of the fertility restorer locus *Rfk1* to a 43-kb contig in kosena radish (*Raphanus sativus* L.). Mol Gen Genom 269: 388-394

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage *P1*-derived vector for the propagation of large human DNA fragments. Nat Genet 6: 84-89

Ishiura M, Hazumi N, Koide T, Okada Y (1989) A *recB recC sbcb recJ* host prevents r*ecA*-independent deletions in recombinant cosmid DNA propagated in *Escherichia coli*. J Bacteriol 171: 1068-1074

Iwabuchi M, Koizuka N, Fujimoto H, Sakai T, Imamura J (1999) Identification and expression of the kosena radish (*Raphanus sativus* cv. Kosena) homologue of the ogura radish CMS-associated gene, *orf138*. Plant Mol Biol 39: 183-188

Jackson RC, Murray BG (1983) Colchicine induced quadrivalent formation in *Helianthus*: evidence for ancient polyploidy. Theor Appl Genet 64: 219-222

Jan CC (2000) Cytoplasmic male sterility in two wild *Helianthus annuus* L. accessions and their fertility restoration. Crop Sci 40: 1535-1538

Jan CC (1997) Cytology and interspecific hybridization. In: Schneiter AA (ed) Sunflower technology and production, Madison, Wisconsin, USA, pp 497-558

Jan CC (1992) Inheritance and allelism of mitomycin C- and streptomycin-induced recessive genes for male sterility in cultivated sunflower. Crop Sci 32: 317-320

Jan CC (1988) Two new sources of cytoplasmic male sterility from wild *H. annuus* L. In Proc Sunflower Res 10th workshop, Natl Sunflower Assoc, Bismarck, ND, pp 7

Jan CC, Vick BA (2007) Inheritance and allelic relationships of fertility restoration genes for seven new sources of male-sterile cytoplasm in sunflower. Plant Breed 126: 213-217

Jan CC, Vick BA (2006) Registration of seven cytoplasmic male sterile and four fertility restoration sunflower germplasms. Crop Sci 46: 1829-1830

Jan CC, Rutger JN (1988) mitomycin C- and streptomycin-induced male sterility in cultivated sunflower. Crop Sci 28: 792-795

Jan CC, Vick BA, Miller JF, Kahler AL, Butler III ET (1998) Construction of an RFLP linkage map for cultivated sunflower. Theor Appl Genet 96: 15-22

Jander G, Norris SR, Rounsley SD, Bush DF, Levin IM, Last RL (2002) *Arabidopsis* map-based cloning in the post-genome era. Plant Physiol 129: 440-450

Janeja HS, Banga SS, Lakshmikumaran M (2003) Identification of AFLP markers to linked to fertility restorer genes for *tournefortii* cytoplasmic male-sterility system in *Brassica napus.* Theor Appl Genet 107: 148-154

Jean M, Brown GG, Landry BS (1997) Genetic mapping of nuclear fertility restorer genes for the `Polima´ cytoplasmic male sterility in canola (*Brassica napus* L.) using DNA markers. Theor Appl Genet 95: 321-328

Jiang J, Gill B, Wang G-L, Ronald PC, Ward D (1995) Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci USA 92: 4487-4491

Kaufmann H, Mattiesch L, Lörz H, Debener T (2003) Construction of a BAC library of *Rosa rugosa* Thumb, and assembly of a contig spanning *Rdr1*, a gene that confers resistance to black spot. Mol Genet Genom 268: 666-674

Kazama T, Toriyama K (2003) A pentatricopeptide repeat-containing gene that promotes the processing of aberrant *atp6* RNA of cytoplasmic male-sterile rice. FEBS letters 544: 99-102

Kelley JM, Field CE, Brook Craven M, Bocskai D, Kim UJ, Rounsley SD, Adams MD (1999) High throughput direct end sequencing of BAC clones. Nucl Acids Res 27: 1539-1546

Kim SC, Rieseberg LH (1999) Genetic architecture of Species differences in annual sunflower: implication for adaptive trait introgression. Genetics 153: 965-977

Kim UJ, Shizuya H, de Jong PJ, Birren B, Simon MI (1992) Stable propagation of cosmid sized human DNA inserts in an F factor based vector. Nucl Acids Res 20: 1083-1085

Kinney AJ (1994) Genetic modification of the storage lipids of plants. Curr Opin Biotech 5: 144-151

Klein RR, Klein PE, Mullet JE, Minx P, Rooney WL, Schertz KF (2005) Fertility restorer locus *Rf1* of sorghum (*Sorghum bicolour* L.) encodes a pentatricopeptide repeat protein not present in the collinear region of rice chromosome 12. Theor Appl Genet 111: 994-1012

Knapp SJ, Berry S, Rieseberg LH (2001) Genetic mapping in sunflower. In: Phillips RL, Vasil IK (eds) DNA markers in plants, Kluwer, Dordrecht, The Netherlands, pp 379-403

Köhler RH, Horn R, Lössl A, Zetsche K (1991) Cytoplasmic male sterility in sunflower is correlated with the co-transcription of a new open reading frame with the *atpA* gene. Mol Gen Genet 227: 369-376

Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J (2003) Genetic characterization of a pentatricopeptide repeat protein gene, *orf687,* that restores fertility in the cytoplasmic male-sterile Kosena radish. Plant J 34: 407-415

Koizuka N, Imai R, Iwabuchi M, Sakai T, Imamura J (2000) Genetic analysis of fertility restoration and accumulation of ORF125 mitochondrial protein in the Kosena radish (*Raphanus sativus* cv. Kosena) and a *Brassica napus* restorer line. Theor Appl Genet 100: 949-955

Kolkman JM, Berry ST, León AJ, Slabaugh MB, Tang S, Gao W, Shintani DK, Burke JM, Knapp SJ (2007) Single nucleotide polymorphisms and linkage disequilibrium in sunflower. Genetics 177: 457-468

Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N (2004) Map-based cloning of a fertility restorer gene, *Rf-1,* in rice (*Oryza sativa* L.). Plant J 37: 315-325

Korell M, Mösges G, Friedt W (1992) Construction of a sunflower pedigree map. Helia 15: 7-6

Kosambi DD (1944) The estimation map distances from recombination values. Annals of Eugenics 112: 172-175

Kräuter-Canham R, Bronner R, Evrard JL, Hahn V, Friedt W, Steinmetz A (1997) A transmitting tissue- and pollen-expressed protein from sunflower with sequence similarity to the human RTP protein. Plant Sci 129: 191-202

Kural A, Miller JF (1992) The inheritance of male fertility restoration of the PET2, GIG1, and MAX1 sunflower cytoplasmic male sterility sources. In Proc. 13th Int Sunflower Conf, Pisa, Italy, Sunflower Assoc, Paris, France

Kusterer B (2003) Genomkartierung und Markerentwicklung bei der Sonnenblume (*Helianthus annuus* L.) unter besonderer Berücksichtigung des Restorergens *Rf1* und des Mehltau-Resistenzgens Pl_2. Inaugural Dissertation, Justus-Liebig University of Gießen. Cuvillier Verlag Göttingen.

Kusterer B, Horn R, Friedt W (2005) Molecular mapping of the fertility restoration locus *Rf1* in sunflower and development of diagnostic markers for the restorer gene. Euphytica 143: 35-43

Kusterer B, Friedt W, Lazarescu L, Prüfe M, Özdemir N, Tzigos S, Horn R (2004a) Map-based cloning strategy for isolating the restorer gene *Rf1* of the PET1 cytoplasm in sunflower (*Helianthus annuus* L.). Helia 27: 1-14

Kusterer B, Rozynek B, Brahm L, Prüfe M, Tzigos S, Horn R, Friedt W (2004b) Construction of a genetic map and localization of major traits in sunflower (*Helianthus annuus* L.). Helia 27: 15-23

Kusterer B, Prüfe M, Lazarescu E, Özdemir N, Friedt W, Horn R (2002) Mapping of the restorer gene *Rf1* in sunflower (*Helianthus annuus* L.). Helia 25: 41-46

Lacombe S, Bervillé A (2001) A dominant mutation for high oleic acid content in sunflower (*Helianthus annuus* L.) seed oil is genetically linked to a single oleate-dasaturase RFLP locus. Mol Breed 8: 129-137

Lai Z, Livingstone K, Zou Y, Church SA, Knapp SJ, Andrews J, Rieseberg LH (2005) Identification and mapping of SNPs from ESTs in sunflower. Theor Appl Genet 111: 1532-1544

Langar K, Lorieux M, Desmarais E, Griveau Y, Gentzbittel L, Bervillé A (2003) Combined mapping of DALP and AFLP markers in cultivated sunflower using F_9 recombinant inbred lines. Theor Appl Genet 106: 1068-1074

Laughnan JR, Gabay-Laughnan S (1983) Cytoplasmic male sterility in maize. Ann Rev Genet 17: 27-48

Laver HK, Reynolds SJ, Monéger F, Leaver CJ (1991) Mitochondrial genome organization and expression associated with cytoplasmic male sterility in sunflower (*Helianthus annuus*). Plant J 1: 185-193

Lawson WR, Goulter KC, Henry RJ, Kong GA, Kochman JK (1998) Marker-assisted selection for two rust resistance genes in sunflower. Mol Breed 4: 227-234

Lazarescu E, Friedt W, Horn R (2006) Expression analysis of the sunflower *SF21* gene family reveals multiple alternative and organ-specific splicing of transcripts. Gene 374: 77-86

Leclercq P (1969) Une stérilité mâle chez le tournesol. Ann Amélior Plantes 19: 99-106

Lee H-R, Eom E-M, Lim Y-P, Bang J-W, Lee D-H (2003) Construction of a garlic BAC library and chromosomal assignment of BAC clones using the FISH technique. Genome 46: 514-520

Lei S, Yao X, Yi B, Chen W, Ma C, Tu J, Fu T (2007) Towards map-based cloning: fine mapping of a recessive genic male-sterile gene (*BnMs2*) in *Brassica napus* L. and syntenic region identification based on the *Arabidopsis thaliana* genome sequences. Theor Appl Genet 115: 643-651

León AJ, Andrade FH, Lee M (2003) Genetic analysis of seed-oil concentration across generations and environments in sunflower. Crop Sci 43: 135-140

León AJ, Lee M, Rufener GK, Berry ST, Mowers RP (1996) Genetic mapping of a locus (*hyp*) affecting seed hypodermis colour in sunflower. Crop Sci 36: 1666-1168

Leónardo SL, Bazzalo ME, Abratti G, Cimmino CJ, Galella MT Grondona M, Giolitti F, León AJ (2005) Screening sunflower for resistance to sunflower chlorotic mottle virus and mapping the *Rcmo-1* resistance gene. Crop Sci 45: 735-739

Lexer C, Welch M, Durphy JL, Rieseberg LH (2003) Natural selection for salt tolerance QTL in wild sunflower hybrids: implications for the origin of *Helianthus paradoxus*, a diploid hybrid species. Mol Ecol 12: 1225-1235

Li YC, Korol AB, Fahima T, Beiles A, Nevo E (2002) Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. Mol Ecol 11: 2453-2465

Libert F, Lefort A, Okimoto R, Womack J, Georges M (1993) Construction of a bovine genomic library of large yeast artificial chromosome clones. Genomics 18: 270-276

Liebhard R, Koller B, Gianfranceschi L, Gessler C (2003) Creating a saturated reference map for the apple (*Malus* × *domestica Borkh*.) genome. Theor Appl Genet 106: 1497-1508

Lijavetzky D, Muzzi G, Wicker T, Keller B, Wing R, Dubcovsky J (1999) Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 42: 1176-1182

Lin JJ, Kuo J, Ma J, Saunders JA, Beard HS, MacDonald MH, Kenworthy W, Ude GN, Matthews BF (1996) Identification of Molecular markers in Soybean comparing RFLP, RAPD, and AFLP DNA mapping techniques. Plant Mol Biol Rep 14: 156-169

Ling H-Q, Zhu Y, Keller B (2003) High resolution mapping of the leaf rust disease resistance gene *Lr1* in wheat and characterization of BAC clones from the *Lr1* locus. Theor Appl Genet 106: 875-882

Liu F, Cui X, Horner HT, Weiner H, Schnable PS (2001) Mitochondrial aldehyde dehydrogenase activity is required for male fertility in maize. Plant Cell 13: 1063-1078

Lu YH, Melero-Vara JM, Garcia-Tejada JA (2000) Development of SCAR markers linked to the gene *Or5* conferring resistance to broomrape (*Orobanche cumana Wallr.*) in sunflower. Theor Appl Genet 100: 625-632

Lu YH, Gagne G, Grezes-Besset B, Blanchard P (1999) Integration of molecular linkage group containing the broomrape resistance *Or5* into an RFLP map in sunflower. Genome 42: 453-456

Lukowitz W, Gillmor SC, Scheible W-R (2000) Positional cloning in *Arabidopsis*. Why it feels good to have a genome initiative working for you. Plant Physiol 123: 795-805

Luo M, Wang Y-H, Frisch D, Joobeur T, Wing RA, Dean RA (2001) Melon bacterial artificial chromosome (BAC) library construction using improved methods and identification of clones linked to the locus conferring resistance to melon *Fusarium* wilt (*Fom-2*). Genome 44: 154-162

Luo MC, Thomas C, You FM, Hsiao J, Onyand S, Buell R, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosome using the SNaPshot labelling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82: 378-389

Lurin C, Andrés C, Aubourg S, Bellaoui M, Bitton F, Bruyère C, Caboche M, Debast C, Gualberto J, Hoffman B, Lecharny A, Le Ret M, Martin-Magniette ML, Mireau

H, Peeters N, Renou JP, Szurek B, Taconnat L, Small I (2004) Genome-Wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16: 2089-2103

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997) High throughput fingerprinting analysis of large-insert clones. Genome Res 7: 1072-1084

Marra MA, Kugaba T, Sekhon M, Hiller L, Martienssen R (1999) A map for sequence analysis of the *Arabidopsis thaliana* genome. Nat Genet 22: 265-270

Martinez-Rivas JM, Sperling P, Lühs W, Heinz E (2001) Spatial and temporal regulation of three different microsomal oleate desaturase genes (*FAD2*) from normal-type and high-oleic varieties of sunflower (*Helianthus annuus* L.) Mol Breed 8: 159-168

Mazeyrat F, Mouzeyar S, Courbou I, Badaoui S, Roeckel-Drevet P, Tourvieille de Labrouhe D, Ledoigt G (1999) Accumulation of defence related transcripts in sunflower hypocotyls (*Helianthus annuus* L.) infected with *Plasmopara halstedii*. Europ J Plant Pathol 105: 333-340

Mazeyrat F, Mouzeyar S, Nicolas P, Tourvieille de Labrouhe D, Ledoigt G (1998) Cloning, sequence and characterization of a sunflower (*Helianthus annuus* L.) pathogen-induced gene showing sequence homology with auxin-induced genes from plants. Plant Mol Biol 38: 899-903

Meksem K, Zobrist K, Ruben E, Hyten D, Quanzhou T, Zhang H-B, Lightfoot DA (2000) Two large-insert soybean genomic libraries constructed in a binary vector: applications in chromosome walking and genome wide physical mapping. Theor Appl Genet 101: 747-755

Mesbah LA, Kneppers TJA, Takken FLW, Laurent P, Hille J, Nijkamp HJJ (1999) Genetic and physical analysis of a YAC contig spanning the fungal disease resistance locus *Asc* of tomato (*Lycopersicon esculentum*). Mol Genet Genom 261: 50-57

Mestries E, Gentzbittel L, Tourvieille de Labrouhe D, Nicolas P, Vear F (1998) Analyses of quantitative trait loci associated with resistance to *Sclerotinia sclerotiorum* in sunflowers (*Helianthus annuus* L.) using molecular markers. Mol Breed 4: 215-226

Michelmore RW, Paran I, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. Proc Natl Acad Sci USA 88: 9228-9832

Micic Z, Hahn V, Bauer E, Schön CC, Knapp SJ, Tang S, Melchinger AE (2004) QTL mapping of *Sclerotinia midstalk-rot* resistance in sunflower. Theor Appl Genet 109: 1474-1484

Miller JF (1996) Inheritance of restoration of the *Helianthus petiolaris* ssp. *fallax* (PEF1) cytoplasmic male sterility. Crop Sci 36: 83-86

Miller JF, Fick GN (1997) The genetics of sunflower. In: Schneiter AA (ed) Sunflower technology and production, Madison, Wisconsin, USA, pp 441-495

Miller JF, Wolf SL (1991) Registration of three cytoplasmic male sterile and three restorer sunflower germplasm lines. Crop Sci 31: 500

Mohan M, Nair S, Bhagwat A, Krishna TG, Yano M, Bhatia CR, Sasaki T (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. Mol Breed 3: 87-103

Monéger F, Smart CJ, Leaver CJ (1994) Nuclear restoration of cytoplasmic male sterility in sunflower is associated with the tissue-specific regulation of a novel mitochondrial gene. EMBO 13: 8-17

Mouzeyar S, Roeckel-Drevet P, Gentzbittel L, Tourvielle de Labrouhe D, Vear, F Nicolas P (1995) RFLP and RAPD mapping of the sunflower *Pl1* locus for resistance to *Plasmopara halstedii* race 1. Theor Appl Genet 91: 733-737

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrfach H, Marra M, Martienssen R, Meir-Ewert S, Altmann T (1999) A complete BAC-based physical map of the *Arabidopsis thaliana* genome. Nat Genet 22: 271-275

Nakamura S, Asakawa S, Ohmido N, Fukui K, Shimizu N, Kawasaki S (1997) Construction of an 800-kb contig in the near-centromeric region of the rice blast resistance gene *Pi-ta*^{2} using a highly representative rice BAC library. Mol Gen Genet 254: 611-620

O'Sullivan DM, Ripoll PJ, Rodgers M, Edwards KJ (2001) A maize bacterial artificial chromosome (BAC) library from the European flint inbred lines F_2. Theor Appl Genet 103: 425-432

Osoegawa K, Mammoser AG, Wu C, Frengen E, Zeng C, Catanese JJ, de Jong PJ (2001) A bacterial artificial chromosome library for sequencing the complete human genome. Genome Res 11: 483-496

Ouvrard O, Cellier F, Ferrare K, Tousch D, Lamaze T, Dupuis JM, Casse-Delbart F (1996) Identification and expression of water stress- and abscisic acid-regulated genes in a drought-tolerant sunflower genotype. Plant Mol Biol 31: 819-829

Özdemir N, Horn R, Friedt W (2004) Construction and characterization of a BAC library for sunflower (*Helianthus annuus* L.). Euphytica 138: 177-183

Özdemir N, Horn R, Friedt W (2002) Isolation of HMW DNA from sunflower (*Helianthus annuus* L.) for BAC cloning. Plant Biol Mol Rep 20: 239-250

Paniego N, Echaide M, Muñoz M, Fernández L, Torales S, Faccio P, Fuxan I, Carrera M, Zandomeni R, Suárez EY, Hopp HE (2002) Microsatellites isolation and characterization in sunflower (*Helianthus annuus* L.) Genome 45: 34-43

Pashley CH, Ellis JR, McCauley DE, Burke JM (2006) EST database as a source for molecular markers: lessons from *Helianthus*. J Heredity 97: 381-388

Paterson AH, DeVerna JW, Lanini B, Tanksley SD (1990) Fine mapping of quantitative trait loci using selected overlapping recombinant chromosome in an interspecies cross of tomato. Genetics 124: 735-742

Peerbolte RP, Peleman J (1996) The CARTISOL sunflower RFLP map (146 loci) extended with 291 AFLP markers. In: Proc of the 18th sunflower research forum, Fargo, ND. National Sunflower Association, Bismarck, ND pp 174-178

Pérez-Vich B, Velasco L, Muñoz-Ruz J, Fernández-Martinez JM (2006) Inheritance of high stearic acid content in the sunflower mutant CAS-14. Crop Sci 46: 22-29

Pérez-Vich B, Berry ST, Velasco L, Fernández-Martinez JM, Gandhi S, Freeman C, Heesacker A, Knapp SJ, León AJ (2005) Molecular mapping of nuclear male sterility genes in sunflower. Crop Sci 45: 1851-1857

Pérez-Vich B, Fernández-Martinez JM, Grondona M, Knapp SJ, Berry ST (2002) Stearoyl-ACP and oleol-PC desaturase genes cosegregate with quantitative trait loci underlying stearic and high oleic acid mutant phenotypes in sunflower. Theor Appl Genet 104: 338-349

Pérez-Vich B, Garcés R, Fernández-Martinez JM (2000) Genetic relationships between loci controlling the high stearic and the high oleic acid traits in sunflower. Crop Sci 40: 990-995

Petes TD (2001) Meiotic recombination hot spots and cold spots. Nat Rev Genet 2: 360-369

Pflieger S, Lefebre V, Causse M (2001) The candidate gene approach in plant genetics: a review. Mol Breed 7: 275-291

Powell W, Machray GC, Provan J (1996) Polymorphism revealed by simple sequence repeats. Trends Plant Sci 1: 215-222

Putt ED (1997) Early History of Sunflower. In: Schneiter AA (ed) Sunflower technology and production, Madison, Wisconsin, USA, pp 1-19

Quagliaro G, Vischi M, Tyrka M, Olivieri AM (2001) Identification of wild and cultivated sunflower for breeding purposes by AFLP markers. J Heredity 92: 38-42

Rajesh PN, Coyne C, Meksem K, Dev Sharma K, Gupta V, Muehlbauer FJ (2004) Construction of a *Hind*III bacterial artificial chromosome library and its use in identification of clones associated with disease resistance in chickpea. Theor Appl Genet 108: 663-669

Reamon-Büttner SM, Schmidt T, Jung C (1999) AFLPs represent highly repetitive sequences in *Asparagus officinalis* L. Chromosome Res 7: 297-304

Ren C, Lee M-K, Yan B, Ding K, Cox B, Romanov MN, Price JA, Dodgson JB, Zhang H-B (2003) A BAC-based physical map of the chicken genome. Genome Res 13: 2754-2758

Rieseberg LH, Randal Linder C, Seiler GJ (1995) Chromosomal and genic barriers to introgression in *Helianthus.* Genetics 141: 1163-1171

Romanov MN, Price JA, Dodgson JB (2003) Integration of animal linkage and BAC contig maps overgo hybridization. Cytogenet Genome Res 101: 277-281

Ronning CM, Stegalkina SS, Ascenzi RA, Bougri O, Hart AL, Utterbach TR, Vanaken SE, Riedmuller SB, White JA, Cho J, Pertea GM, Lee Y, Karamycheva S, Sultana R, Tsai J, Quackenbush J, Griffiths HM, Restrepo S, Smart CD, Fry WE, van der Hoeven R, Tanksley S, Zhang P, Jin H, Yamamoto ML, Baker BJ, Buell R (2003) Comparative analyses of potato expressed sequence tag libraries. Plant Physiol 131: 419-429

Sambrook J, Russell DW (2001) Molecular cloning: A laboratory manual, 3d edition Cold Spring Harbor laboratory Press, Cold Spring Harbor, NY

Sandhu D, Champoux JA, Bondareva SN, Gill KS (2001) Identification and physical localization of useful genes and markers to a major gene-rich on wheat group *1S* chromosomes. Genetics 157: 1735-1747

Sandhu D, Gill KS (2002) Gene-containing regions of wheat and the other grass genomes. Plant Physiol 128: 803-811

Scheible WR, Törjek O, Altmann T (2004) From markers to cloned genes: Map-based cloning. In: Lörz H and Wenzel G (eds) Biotechnology in Agriculture and Forestry. Vol. 55 Molecular marker systems, Springer-Verlag, Berlin, Heidelberg, pp 55-86

Schilling EE (1997) Phylogenetic analysis of *Helianthus* (Asteraceae) based on chloroplast DNA restriction site data. Theor Appl Genet 94: 925-933

Schmidt R (2002) Plant genome evolution: lessons from comparative genomics at the DNA level. Plant Mol Biol 48: 21-37

Schnabel U, Engelman U, Horn R (2008) Development of markers for the use of the PEF1 cytoplasm in sunflower hybrid breeding. Plant Breed 127: 587-591

Schnable PS, Wise RP (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. Trends Plant Sci 3: 175-180

Schuppert GF, Tang X, Slabaugh MB, Knapp SJ (2006) The sunflower high-oleic mutant *Ol* carries variable tandem repeats of *FAD2-1*, a seed-specific oleoyl-phosphatidyl choline desaturase. Mol Breed 17: 241-256

Seiler GJ (1992) utilization of wild sunflower species for the improvement of cultivated sunflower. Field Crops Res 30: 195-230

Seiler GJ, Rieseberg LH (1997) Systematic, origin, and germplasm resources of the wild and domesticated sunflower. In: Schneiter AA (ed) Sunflower technology and production, Madison, Wisconsin, USA, pp 21-66

Serieys H (2005) Identification, study and utilization in breeding programs of new CMS sources, in the FAO Subnetwork. In: Proc. 2005 sunflower Subnetwork progress report, 17-20 July 2005, FAO, Rome, Italy, pp 47-53

Serieys H (1996) Identification, study and utilization in breeding programs on new CMS sources. Helia 19: 144-160

Serieys H, Vincourt P (1987) Charactérisation de nouvelles sources de stérilité mâle chez le tournesol. Les Colloques de l'INRA Paris. 45: 53-64

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 89: 8794-8797

Siculella L, Palmer JD (1988) Physical and gene organization of mitochondrial DNA in fertile and male sterile sunflower. CMS associated alterations in structure and transcription of the atpA gene. Nucl Acids Res 16: 3787-3799

Small ID, Peeters N (2000) The PPR motif a TPR-related motif prevalent in plant organellar proteins. Trends Biochem Sci 25: 46-47

Smart CJ, Monéger F, Leaver CJ (1994) Cell-specific regulation of gene expression in mitochondria during anther development in sunflower. Plant Cell 6: 811-825

Smith D, Smyth AP, Moir D (1990) Amplification of large artificial chromosomes. Proc Natl Acad Sci USA 87: 8242-8246

Snowdon RJ, Friedt W (2004) Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. Plant Breed 123: 1-8

Song J, Bradeen JM, Naess SK, Helgeson JP, Jiang J (2003) BIBAC and TAC clones containing potato genomic DNA fragments larger than 100 kb are not stable in *Agrobacterium*. Theor Appl Genet 107: 958-964

Song J, Dong F, Jiang J (2000) Construction of a bacterial artificial chromosome (BAC) library for potato molecular cytogenetics research. Genome 43: 199-204

Sossey-Alaoui K, Serieys H, Tersac M, Lambert P, Schilling E, Griveau Y, Kaan F, Bervillé A (1998) Evidence for several genomes in *Helianthus*. Theor Appl Genet 97: 422-430

Spielmeyer W, Moullet O, Laroche A, Lagudah E (2000) Highly recombinogenic regions at seed storage protein loci on chromosome 1DS of *Aegilops tauschii*, the D-genome donor of wheat. Genetics 155: 361-367

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000) Subgenome chromosome walking in wheat: A 450-kb physical contig in *Triticum monococcum* L. spans the *Lr10* resistance locus in hexaploid wheat (*Triticum aestivum* L.). Proc Natl Acad Sci USA 97: 13436-13441

Sternberg N (1990) Bacteriophage *P1* cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc Natl Acad Sci USA 87: 103-107

Tamborindeguy C, Ben C, Liboz T, Gentzbittel L (2004) Sequence evaluation of four specific cDNA libraries for developmental genomics of sunflower. Mol Genet Genom 271: 367-375

Tang S, Yu J-K, Slabaugh MB, Shintani DK, Knapp SJ (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105: 1124-1136

Tanksley SD, Ganal MW, Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes. Trends Genet 11: 63-68

Tanksley SD, Ganal MW, Prince JP, de Vicente MC, Bonierbale MW, Broun P, Fulton TM, Giovannoni JJ, Grandillo S, Martin GB, Messeguer R, Miller JC, Miller L, Paterson AH, Pineda O, Röder MS, Wing RA, Wu W, Young ND (1992) High density molecular linkage maps of the tomato and potato genomes. Genetics 132: 1141-1160

Tao Q, Wang A, Zhang H-B (2002) One large-insert plant transformation-competent BIBAC library and three BAC libraries of japonica rice for genome research in rice and other grasses. Theor Appl Genet 105: 1058-1066

Tao Q, Chang Y-L, Wang J, Chen H, Islam-Faridi MN, Scheuring C, wang B, Stelly DM, Zhang H-B (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprinting analysis. Genetics 158: 1711-1724

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

Tomkins JP, Davis G, Main D, Yim YS, Duru N, Musket T, Goicoechea JL, Frisch DA, Coe EH Jr, Wing RA (2002) Construction and characterization of a deep-coverage bacterial artificial chromosome library for maize. Crop Sci 42: 928-933

Tomkins JP, Peterson DG, Yang TJ, Main D, Wilkins TA, Peterson AH, Wing RA (2001a) Development of genomic resources for cotton (*Gossypium hirsutum* L.): BAC library construction, preliminary STC analysis, and identification of clones associated with fiber development. Mol Breed 8: 255-261

Tomkins JP, Wood TC, Stacey MG, Loh JT, Judd A, Goicoechea JL, Stacey G, Sadowsky MJ, Wing RA (2001b) A marker-dense physical map of the *Bradyrhizobium japonicum* genome. Genome Res 11: 1434-1440

Tomkins JP, Mahalingam R, Smith H, Goicoechea JL, Knap HT, Wing RA (1999a) A bacterial artificial chromosome library for soybean PI 437654 and identification of clones associated with *cyst* nematode resistance. Plant Mol Biol 41: 25-32

Tomkins JP, Yu Y, Miller-Smith H, Frisch DA, Woo SS, Wing RA (1999b) A bacterial artificial chromosome library for sugarcane. Theor Appl Genet 99: 419-424

Touzet P, Hueber N, Bürkholz A, Barnes S, Cuguen J (2004) Genetic analysis of male fertility restoration in wild cytoplasmic male sterility G of beet. Theor Appl Genet 109: 240-247

Vear F, Gentzbittel L, Philippon J, Mouzeyar S, Mestries E, Roeckel-Drevet P, Tourvielle de Labrouhe D, Nicolas P (1997) The genetic of resistance to five races of downy mildew (*Plasmopara halstedii*) in sunflower (*Helianthus annuus* L.). Theor Appl Genet 95: 584-589

Venter JC, Smith HO, Hood L (1996) A new strategy for genome sequencing. Nature 381: 364-366

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Horenes 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

Vranceanu VA (1970) Advances in sunflower breeding in Romania. In: Proc of the 4th Int Sunflower Conf, ISA, CETIOM Paris (eds), Memphis Tennessee, USA, pp 136-148

Vranceanu VA, Stoenescu FM (1971) Pollen fertility restorer gene from cultivated sunflower (*Helianthus annuus* L.). Euphytica 20: 536-541

Vranceanu VA, Iuoras M, Stoenescu FM (1986) A contribution to the diversification of the CMS sources in sunflower. Helia 4: 23-27

Wang G-L, Holsten TE, Song W-H, Wang H-P, Ronald PC (1995) Construction of a rice bacterial artificial chromosome library and identification of clones linked to the *Xa-21* disease resistance locus. Plant J 7: 525-533

Wang Z, Zhang YJ, Xiang CP, Mei SY, Zhou Y, Chen GP, Wang T (2008) A new fertility restorer locus linked closely to the *Rfo* locus for cytoplasmic male sterility in radish. Theor Appl Genet 117: 313-320

Wang Z, Zou Y, Li X, Zhang Q, Chen L, Wu H, Su D, Chen Y, Guo J, Luo D, Long Y, Zhong Y, Liu YG (2006) Cytoplasmic male sterility of rice with boro II cytoplasm is caused by a cytotoxic peptide and is restored by two related PPR motif genes via distinct modes of mRNA silencing. Plant Cell 18: 676-687

Weng Y, Tuleen NA, Hart GE (2000) Extended physical maps and a consensus physical map of the homoeologous group-6 chromosomes of wheat (*Triticum aestivum* L. Em Thell). Theor Appl Genet 100: 519-527

Whelan EDP (1981) Cytoplasmic male sterility in Helianthus giganteus L. × H. annuus L. interspecific hybrids. Crop Sci 21: 855-858

Whelan EDP (1980) A new source of cytoplasmic male sterility in sunflower. Euphytica 29: 33-46

Wing RA, Zhang H-B, Tanksley SD (1994) Map-based cloning in crop plant: Tomato as a model system: I. Genetic and physical mapping of jointless. Mol Gen Genet 242: 681-688

Winter P, Kahl G (1995) Molecular marker technologies for plant improvement. World Journal of Microbiology and Biotechnology 11: 438-448

Woo SS, Jiang J, Gill BS, Paterson AH, Wing RA (1994) Construction and characterization of a bacterial artificial chromosome library of sorghum bicolour. Nucl Acids Res 22: 4922-4931

Wu CC, Nimmakayala P, Santos FA, Springman R, Scheuring C, Meksem K, Lightfoot DA, Zhang H-B (2004) Construction and characterization of a soybean bacterial artificial chromosome library and use of multiple complementary libraries for genome physical mapping. Theor Appl Genet 109: 1041-1050

Xu Z, Sun S, Covaleda L, Ding K, Zhang A, Wu C, Scheuring C, Zhang H-B (2004) Genome physical mapping with large-insert bacterial clones by fingerprint analysis: methodologies, source clone genome coverage, and contig map quality. Genomics 84: 941-951

Yang D, Parco A, Nandi S, Subudhi P, Zhu Y, Wang G, Huang N (1997) Construction of a bacterial artificial chromosome (BAC) library and identification of overlapping BAC clones with chromosome 4-specific RFLP markers in rice. Theor Appl Genet 95: 1147-1154

Yim Y-S-, Davis GL, Duru NA, Musket TA, Linton EW, Messing JW, McMullen MD, Soderlund CA, Polacco ML, Gardiner JM, Coe Jr. EH (2002) Characterization of three maize bacterial artificial chromosome libraries toward anchoring of the physical map to the genetic map using high-density bacterial artificial chromosome filter hybridization. Plant Physiol 130: 1686-1696

Yin J, Guo W, Yang L, Liu L, Zhang T (2006) Physical mapping of the *Rf1* fertility-restoring gene to a 100 kb region in cotton. Theor Appl Genet 112: 1318-1325

Yokobata K, Trenchak B, De Jong PJ (1990) Rescue of unstable Cosmids by *in vitro* packaging. Nucl Acids Res 19: 403-404

Yu JK, Tang SX, Slabaugh MB, Heesacker A, Cole G, Herring M, Soper J, Han F, Chu WC, Webb DM, Thompson L, Edwards KJ, Berry S, León AJ, Grondona M, Olungo C, Maes N, Knapp SJ (2003) Towards a saturated molecular genetic linkage map for cultivated sunflower. Crop Sci 43: 367-387

Yu JK, Mangor J, Thompson L, Edwards KJ, Slabaugh MB, Knapp SJ (2002) Allelic diversity of simple sequence repeats among elite inbred lines of cultivated sunflower. Genome 45: 652-660

Yüksel B, Paterson AH (2005) Construction and characterization of a peanut *Hind*III BAC library. Theor Appl Genet 111: 630-639

Zhang H-B, Wu C (2001) BAC as tools for genome sequencing. Plant Physiol Biochem 39: 195-209

Zhang H-B, Wing RA (1997) Physical mapping of the rice genome with BACs. Plant Mol Biol 35: 115-127

Zhang H-B, Choi S, W SS, Li Z, Wing RA (1996) Construction and characterization of two bacterial artificial chromosome libraries from the parents of a permanent recombinant inbred mapping population. Mol Breed 2: 11-24

Zhang ZF, Wang Y, Zheng YL (2006) AFLP and PCR-based markers linked to *Rf3,* a fertility restorer gene for S cytoplasmic male sterility in maize. Mol Genet Genom 276: 162-169

8 Appendix

8.1 Sequence homology of *sf2*

H. annuus sf2 gene for an anther-specific protein

```
Score =  348 bits (188),  Expect = 2e 91
Identities = 369/452 (81%), Gaps = 31/452 (6%)
Strand=Plus/Minus

Query  2084  TTTTAATTTAAACGGTATAGTATGGAAACGGTATAATATATTCGAACCGTACTCAAGTAT  2143
              ||||||||||||| ||| |||||  |||||||||| |||||||||| ||| | |||||
Sbjct  2576  TTTTAATTTAAACCGTAGAGTATTCAAACGGTATAACATATTCGAACCCTACACGAGTAT  2517

Query  2144  TTAAAAATAATTAACGGTATGATATGGAAACGTTTATAGCATTTGAACCGCACTAGA TG  2202
              |||||||| |||||||||| ||||  ||| || ||||  |||||||||| |||||| |
Sbjct  2516  TTAAAAAT ATTAACGGTATAATATTTAAAAGTATATAATATTTGAACCGTACTAGACT   2459

Query  2203  TGGTATAATAATTTAATTTAAACGATAACAGTATAATATATTCGAACCGTATAT ACGTT  2261
              |||||||||||||||||||||||||||| |||||||||||||||||||| | | |||
Sbjct  2458  TGGTATAATAATTTAATTTAAACGATAACAATATAATATATTCGAACCGTATCTGA GTT  2400

Query  2262  TTT aaaaaaaTAATTAACGGTATATAATATTAATGGTATAAGATTAACGTTTATATATT  2320
              ||| ||||||| ||||||||||  |||| ||||| ||| ||   | |   |  || | ||
Sbjct  2399  TTTCAAAAAAA AATTAACGGTC ATAACATTAACGGT TA   T A   T  AT T TT  2354

Query  2321  TTAAGCGATTTA TGTTGGAACCGGGTGAGAATTTATACCACAAATTCATCAAAATAACC  2379
                |||| ||||| |||| |||||||||| ||||||||| ||||||| || |||| ||||
Sbjct  2353    AAGCAATTTAATGTTTGAACCGGGTGGGAATTTATAACACAAATACACCAAACTAACA  2296

Query  2380  AAACAAAATTCTGATGGGTTATATACGTTGTGTATAGCTCTATACGCAGCGTATATAAAC  2439
              ||||||||||||||| |||| |||||| || |||||   ||||||| |||||||| ||||
Sbjct  2295  AAACAAAATTCTGATTGGTTCTATACGCTGCGTATAAAGCTATACGTAGCGTATACAAAC  2236

Query  2440  TCTGGTTGTCTGTGCC AATGATTGCTGGG CAGNG CTCTGAA TTCTATGCAACTTAT  2495
               || ||| ||| || | || ||||||| || |||   | || || ||| ||||||| |||
Sbjct  2235  CCTAGTTATCTATGTCTAA GATTGCTTGGTCAGACACGCTAAAATTC ATGCAACCTAT  2178

Query  2496  ACGCTGCGTATAGCTCTATACGC GTCGTATA  2526
              ||||||| ||||||| | ||||| | ||||||
Sbjct  2177  ACGCTGCATATAGCTGTGTACGCAG CGTATA  2147
```

- Query : 225D09-B-sf2 sequence

8.2 Partial sequences of the cloned BAC ends

16M16- B1- B4

Forward: 492 bp

AAGACACATAAACCTGAAAATCATGACAGATACATGAATTGAACATGATATGAGTGTTCACATGTTATCA
CAAACACTACCCGTTTAAACCCGAACTTTATTTTTGTTCTTTGTGTATTAATACTTTATAGTAACATCTT
ACTATACTAATAAATGAAAATATTTTTTGGACACGTGTCATCTCCTGGTGTTTTCTCACCTTATTTTCCT
ATTTATCTCTCATATTAAATAATAATAATAATAATTTATTAGATAAGAATAAATATTTAGATAAGAATAA
ATATTTTGATAAGGATAAATGTTTGTTTTTATTATGATTATATTAAATAATAATAACAATAATTTTAAAC
AAAAAATTAGATAAGAATATAAAAGTTTGTTTTTATTATGATCATAGTAAATAATAAAAATTATAATAAA
AAATATAGATAAGAATAAATATTTGTTTTTCTATAAATAAATCTAAACAAAAATTTCGATAAGCATAAAC
GT

139A17 – B1- B6

Forward: 1376 bp

CAATAAGAGTAACGACTAGCAGGCAATAACATTTAATTGCTATTTATTCTTGTTTATTAGTATTACTTGG
TCAAGAATAAAGATATATTATGGAAATACAAATTTTCTTGATTCTGGATTAAAGCACTAATAAAAAAAAG
AGGCACTTTTTGAAAAGATACTATTTATGGGAAATAGGAAATTCTCTCCGGCAAACCTGGGTATAGAGTA
GCAGACTCTCCAGCTTGTCCGGATATACTGAAATTACCGCTCCGGCACGAACCGGATAAGACGGGATAAA
TGGTATGTCAAACCAGTAACAAGATTTCCCTAAATAGTATCATGACCAAATATCTGACTGGTTTTTGGAA
ATATGAATCTATTATATACATTTAACCCTATGAAAGGTATGTATATAACAACATGTGAAATATATAATAT
AGATATGTATATCTATAAGGAATTCCAAAAGTCAATTAAGAATAAAGCACAAGCATATGTATATAGATTT
CTTTATCAGGAATCTCTCGCATATATCTTTAATTCCGATATCAAACATTATTTCATTTGATCATCTGTCT

CGTAACTCATGCGATAATAATAAATGGAAATTCTATAAGTTGGAACACCATCAAAAATATAAGAATTTCC
AATACACTACCATAAACTATTAACGCAAGTAAGAAACATCTAAAGTTGTGCTAATGCACTAGAAAACACA
TGAAACCTAAAATTTCACGTCCGCAAGACAACAAAGTGTATCACATACAATTTCCAAGGCAAGACCTTTG
AAAAATATAAATTTGAGCACGATGATACCAATACTAAAAGAACTCCTAATCAAAAAGCGGCACTTTGCCA
CATGATCTTACAAACGATCAAAATCTCGCTGAGGTACGTTAGAACAATATTTCACAAACATCGGTGCATA
GTCTAACCTGAATCATAATTATTAGCACTACAAAAGAAGTCATATATTTTTGCAAATCCCCGATTTCATT
TCAATTCATTCGACATTCTCTGATAATGTCATTTAACAAATAGTTATCACACGAAATTCCAAATAAGAAG
TTCTTATATTTCTTTCGTTGCAAATCTGACTTCTGCCTATAAAGAGTTGACTTTCGTGTTTCTACTTCCT
CTAATAGTCATCATACCATGAGGATTTCTTTCATAATAGCAAATATTGATCTATTTCATTTCTTGGTAAA
TCCACACTTTACACATGCACTGTTTGTTGCGCATGTGGTTCGTTTGATTTATGAAGACCACTGGAGGGCT
TTATTCCAATTTGTTATCAAAAGTTCAAATATCGTATCACTATACTTGATCTTTGCATTTATAAATCACG
TTTATTGCAAAAAGTTGACTCTTTGATATCGAGTCAATAAATTTCT

Reverse: 1011 bp

GGTCATGATACTATTTAGGGAAATCTTGTTACTGGTTTGACATACCATTTATCCCGTCTTATCCGGTTCG
TGCCGGAGCGGTAATTTCAGTATATCCGGACAAGCTGGAGAGTCTGCTACTCTATACCCAGGTTTGCCGG
AGAGAATTTCCTATTTCCCATAAATAGTATCTTTTCAAAAAGTGCCTCTTTTTTTTATTAGTGCTTTAAT
CCAGAATCAAGAAAATTTGTATTTCCATAATATATCTTTATTCTTGACCAAGTAATACTAATAAACAAGA
ATAAATAGCAATTAAATGTTATTGCCTGCTAGTCGTTACTCTTATTGAATACCAATTATACTTATTTATA
CTTATATAAGTAATTTTATCTTGAAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCAT
GGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCAT
AAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCT
TTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGTACTTATATAAGTAATTTTATCTTGAA
GCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAA
TTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAA
TGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCC
AGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTTT
GACAGCTTATCATCGAATTTCTGCCATTCATCCGCTTATTATCACTTATTCAGGCGTAGCAACCAGGCGT
TTAAGGGCACCAATAACTGCCTTAAAAAAAT

126N19 – B7- B8

Forward: 3486 BP

GTTTCGACAGCCGTACGCTGTGGCAATAAGACACACTATCTTTTTACAGAACCATTTAGTGTAAAATATT
CAATTGACAAGCCTAGCAAATTAACTCAATTATTTGGACAAATATACAGAATCAAGCAAAATTAACAAGA
GTAGCAAATTTGGGGGCAAGAAACTACACACAATTTAGGGCTTAAATGATTACCTTCTACGATTGACGAC
TTGGAATATATCAGAACTGATATTTGATACTTGGAACACAAAATATCGATCAATTGACCGGGACGAGGAA
GACCCAAGGACACAAGTTTATTCCAGGGTCGATGATTTTGACTCCAATTTGTCAATTACATGCTTCTTTA
TTCGCTTTTCACTTCTCTAACCTAATTTCGGTAACCACGCTTACCTATGGTCGGCCCTAGTTTTCAATCA
AAACCCTAAATAATGAGCGAGATATTAACCCGCCTGGTTAAATTTGGGCTTAAAATTTCATTTTGTTACC
GCTAGACGGTAGCAATTAAAAAAATACGCATATATTTGGGATAATCTAGGGAAGTTATTCTTATTTTTCC
ACGAATCTGTCACCAAAAATAGATGGTGGCAAAATTTTCCACCGCTTTGCTACCAAACGATGCTGAGTAC
ACTTAATTTACATATTCATGGGAAATAGGTGGCTAAATTTGCCACCGATTTTTTCGATAACAAATTTTGG
TGACAATTGCAAAGTTTTCTAGTAGTGAATCTTCAGTCACCAGCACTACACTTACATATTGGTTACCTAA
TCTTAGAGCATTCTCATCGGAATCTCTAAACTTCGTTTATTTTTTAAAGATCTCTCACAATTTATTATTT
ATTTGTATCTTTATCTTACTAACAAATAAAACACTTATTTTATTGTTTTTTATCGCTTTTATACGCTTAC
AAAATTATATTTAGAGATGAATTGTAATCTCTATATCTAGGGTGTAGGGTATGATTCAAACACTTTAGAG
TGTTTGAGTTTTCTCTATCAAATAATCTCATGCCATGTTAACTCCCCTCTTTCACTCCACATTTTCTACT
CAAACACTAGAAATGATTCAAACACCCACTTAATTAAAAAAAAATTGATTGGTTCTCTATCTCCTCTCTC
CTTGACTCTCTCCTCTCTCTCTCCTCCCTTCCTTCTTCTTCACCTTCGGTGACTATACACCGATTTTG
CATGTCCCCACTCACCGAAATCTGGCTGGCTTTCAATCACCGAGCGATGATTCTAAATCCCCGAAGAGTT
TACCGCATTCATACCAACCCCCTTATAGACGTAGTATCAACTTTCTTAATGTTTTCATCATTATAGAGAC
TCTAATGTATGTAAAAAAATTTACAAAACTATATATCTTTCTGGAATGTGGTTTTAGAGAAGCACATGGG
AATTGTAAGTTATTCCATAACCATCAGGTACTTGATTACATATTGTCCATCACATATTAGTTATACACAT
TAGTTTATTAAGAATACACATTCGGATTAGCAATTTTACATTTATATATCATTTATTACTATTCTAATTT
TTATATAGTTACTAGGGAGAAAGCCCGCGCGATGCACGGCATGCATAATAGCTATTGTTTGTTTGTGGTA
AGATGTCTGACACGGCCGGTGCATAACATGTAAGACAATACGCCTACAAAGGGGCGTGCTTAAACTTTAT
TAAACCGTGGACCGTAAGTTGTTTCGTTGGCAAAACTTACCGGCCAAAAAGAAGTGAAAAAAGATGTCTT
CCAAGAAAAAAAATGAAACAGAAAAAGGAAAAACGCATAAACTCCCGCACGCCTTCCTATCTTATTGGAT
AACACATAGACCAAACTTTAGCTAACGCATGCCACCGTCATCCCCAAACCCATCACCAGCCAACTGACTG
CAGTGAAATAACGAAAAGAGAAGTACTAAAAATGGTATACTATGTTTACGTATATATGCATCGAAAAGTA

GTATGGGTTTTCATTACTAACACTAAGTTCAGACGGTATACCTTTCATGTGGGTCACTGTCAACTGTTTC
TTCTGAATTCGAATCTGTATAAGCAGGCCTGGGAAGGGAAAGCATGAAAAATTTCTGTAGACTGTTATCC
AAAAATAAATAGGCACAATTACATTGTTTATGGAAATCAATCAATTATAGTTGAAAGTTTTTTTGTTTCC
TTTCCATTTATAACGACACCAAGTTTATTTTTTTCAAAAGCCAACTTCTCCCCTTATATGTCTTTACACA
TTCGAAGTGATGCGTTACCAAGTTTTTTTATTCCTATTTTTTTTAATGTAGAACAATGCACTTTTATATT
TACGAACTCAAGTTTCCGACTGTTGTAGCCTCCTAATTTTAAAATAGAGTAAACATATACCAGGTATGTA
AATAACATTTCAGATCGAGTAGAAAGGTAAATAGCACATCTTTAATTAAAATGGTTTATTAAACAGGGCA
CGAATAAAATGTAAGCCGATATATACCTTACGATCGCTTCAACGCGCTTAACTGGTGTCGGTATGGGTAG
CACCCCCTTGTCTTTAGCGGGGATCGATGGAGGTGTGGAAGAACCAAAGGGATTACAACCATCCCCATCT
AGCAGGACATTCTTGTAGTTGAAGGATTCAAAATGGCCATGATCATAATACGTACACGAATCAAGCTGGA
GGGTCTGAGTGCTGCCCATAAGAGAGTGCAAACAATTCGGCAGTGTTGAAATGCTTCCATCAGTGTTCCT
TACCGTCACCGGAGTGAACTTATCCTGTATGTCAGTGGAATAAAGGCTGTAGACGCGTGACAGGCCAGAC
TCTGCTGTGAGTATACTATGAGCAGTGGTCTTTGTCAAGCGTTGGGCGGTGTCTTCAAAGCAGACAATGA
CAGCTTCGGCTGTTGAGTCAAACACCTCGAGTTCTAGCCTGAAACTGTGCAAGGTTTGTAATTTATTTTC
AGTTATAAGCCTGCATGAGAATGGCTACAAATTAATACAACTAAGCTAAAAGGAATCATATACATATTAG
GTAATAACCATATATAGACTTAATTAACTTAAGTTGCACACATAATAAGGATAGGCTTCACATTCACTAC
TTCCACCCCACAAGGGTCATCTTCCACCATTGCCATCGCCATCGCAGTCCCCGCACTTTTGTATCGCTCC
ACAACCTCTCCACCATCACCCTCCGTCGTCTTAGTTTCAATTTCTTCCTCTTTTATCTCGCTGCCGCTGC
TTCGGCTAGAATCAACCGGATACACTTCCGAGTTGGACGCCTTTATGTTTGGCATCACCGGAGTGAAGCA
CAACATTCAGGTTTCTTTGACACCGATTATTTTTTACTACAATTAACTAATCTACTTTCAACTATTTTCC
ACGCCATTTGAGTAAAAAGAAAAAAAGGACATTTTCTTGATTCTCTTAAAAAGATTGTTTTTATTCATTT
TGAACTGGGGATTTTTATGTGTCAAAGTGGTTAGAGACATGCAACTTATAAGCCTT

Reverse: 3353 bp

TAATTAGTAGCCATTCATCATGCAGGCTTATATACTGAAAATCAAATTACAAACCTTGCACAGTTTCAGG
CTAGAACTCGAGGTGTTTGACTCAACAGCCGAAGCTGTCATTGTCTGCTTTGAAGACACCGCCCAACGCT
TGACAAAGACCACTGCTCATAGTATACTCACAGCAGAGTCTGGCCTGTCACGCGTCTACAGCCTTTATTC
CACTGACATACAGGATAAGTTCACTCCGGTGACGGTAAGGAACACTGATGGAAGCATTTCAACACTGCCG
AATTGTTTGCACTCTCTTATGGGCAGCACTCAGACCCTCCAGCTTGATTCGTGTACGTATTATGATCATG
GCCATTTTGAATCCTTCAACTACAAGAATGTCCTGCTAGATGGGGATGGTTGTAATCCCTTTGGTTCTTC
CACACCTCCATCGATCCCCGCTAAAGACAAGGGGGTGCTACCCATACCGACACCAGTTAAGCGCGTGAAG
CGATCGTAAGGTATATATCGGCTTACATTTTATTCGTGCCCTGTTTAATAAACCATTTTAATTAAAGATG
TGCTATTTACCTTTCTACTCGATCTGAAATGTTATTTACATACCTGGTATATGTTTACTCTATTTTAAAA
TTAGGAGGCTACAACAGTCGGAAACTTGAGTTCGTAAATATAAAAGTGCATTGTTCTACATTAAAAAAAA
TAGGAATAAAAAAACTTGGTAACGCATCACTTCGAATGTGTAAAGACATATAAGGGGAGAAGTTGGCTTT
TGAAAAAAATAAACTTGGTGTCGTTATAAATGGAAAGGAAACAAAAAAACTTTCAACTATAATTGATTGA
TTTCCATAAACAATGTAATTGTGCCTATTTATTTTTGGATAACAGTCTACAGAAATTTTTCATGCTTTCC
CTTCCCAGGCCTGCTTATACAGATTCGAATTCAGAAGAAACAGTTGACAGTGACCCACATGAAAGGTATA
CCGTCTGAACTTAGTGTTAGTAATGAAAACCCATACTACTTTTCGATGCATATATACGTAAACATAGTAT
ACCATTTTTAGTACTTCTCTTTTCGTTATTTCACTGCAGTCAGTTGGCTGGTGATGGGTTTGGGGATGAC
GGTGGCATGCGTTAGCTAAAGTTTGGTCTATGTGTTATCCAATAAGATAGGAAGGCGTGCGGGAGTTTAT
GCGTTTTTCCTTTTTCTGTTTCATTTTTTTTCTTGGAAGACATCTTTTTTCACTTCTTTTTGGCCGGTAA
GTTTTGCCAACGAAACAACTTACGGTCCACGGTTTAATAAAGTTTAAGCACGCCCCTTTGTAGGCGTATT
GTCTTACATGTTATGCACCGGCCGTGTCAGACATCTTACCACAAACAAACAATAGCTATTATGCATGCCG
TGCATCGCGCGGGCTTTCTCCCTAGTATACTATATAAAAATTAGAATAGTAATAAATGATATATAAATGT
AAAATTGCTAATCCGAATGTGTATTCTTAATAAACTAATGTGTATAACTAATATGTGATGGACAATATGT
AATCAAGTACCTGATGGTTATGGAATAACTTACAATTCCCATGTGCTTCTCTAAAACCACATTCCAGAAA
GATATATAGTTTTGTAAATTTTTTTACTACATTAGAGTCTCTATAATGATGAAAACATTAAGAAAGTTGA
TACTACGTCTATAAGGGGGTTGGTATGAATGCGGTAAACTCTTCGGGGATTTAGAATCATCGCTCGGTGA
TTGAAAGCCAGCCAGATTTCGGTGAGTGGGGACATGCAAAATCGGTGTATAGTCACCGAAGGTGAAGAAG
AAGGAAGGGAGGAGAGAGAGAGAGGAGAGAGAAAGGAGAGAGGAGATAGAGAACCAATCAATTTTTTTTT
AATTAAGTGGGTGTTTGAATCATTTCTAGTGTTTGAGTAGAAAATGTGGAGTGAAAGAGGGGAGTTAACA
TGGCATGAGATTATTTGATAGAGAAAACTCAAACACTCTAAAGTGTAAGAATCATACCCTACACCCTAGA
TATAGAGATTACAATTCATCTCTAAATATAATTTTGTAAGCGTATAAAAGCGATAAAAAACAATAAAATA
AGTGTTTTATTTGTTAGTAAGATAAAGATACAAATAAATAATAAATTGTGAGAGATCTTTAAAAAATAAA
CGAAGTTTAGAGATTCCGATGAGAATGCTCTAAGATTAGGTAACCAATATGTAAGTGTAGTGCTGGTGAC
TGAAGATTCACTACTAGAAAACTTTGCAATTGTCACCAAAATTTGTTATCGAAAAAATCGGTGGCAAATT
TAGCCACCTATTTCCCATGAATATGTAAATTAAGTGTACTCAGCATCGTTTGGTAGCAAAGCGGTGGAAA
ATTTTGCCACCATCTATTTTTGGTGACAGATTCGTGGAAAAATAAGAATAACTTCCCTAGATTATCCCAA
ATATATGCGTATTTTTTTAATTGCTACCGTCTAGCGGTAACAAAATGAAATTTTAAGCCCAAATTTAACC
AGGCGGGTTAATATCTCGCTCATTATTTAGGGTTTTGATTGAAAACTAGGGCCGACCATAGGTAAGCGTG

GTTACCGAAATTAGGTTAGAGAAGTGAAAAGCGAATAAAGAAGCATGTAATTGACAAATTGGAGTCAAAA
TCATCGACCCTGGAATAAACTTGTGTCCTTGGGTCTTCCTCGTCCCGGTCAATTGATCGATATTTTGTGT
TCCAAGTATCAAATATCAGTTCTGATATATTCCAAGTCGTCAATCGTAGAAGGTAATCATTTAAGCCCTA
AATTGTGTGTAGTTTCTTGCCCCCAAATTTGCTACTCTTGTTAATTTTGCTTGATTCTGTATATTTGTCC
AAATAATTGAGTTAATTTGCTAGGCTTGTCAATTGAATATTTTACACTAAATGGTTCTGTAAAAAGATAG
TGTGTCTTATTGCCACAGCGTACGGCTGTCGAAACGTAGAGCGCGGTTGTAGAAACGCAGCGGCCAAACA
CAACGGCACGGCTGCAGGCAGACGCAACGGACACGACTGTCGTAGCGCTGTAGCGCGCAGCTTCAGAAAC
GCAGCGTGCGGCTACCGAAACACAGCGCGCAGTTGCAGATGCGCGTACGTTTGTATGTTTGTTGCAGTAT
GTTATAAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTG
TGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGG
TGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGG

261D17 – B1- B2

Forward: 350 bp

GAACAGGTACCAAGAGAGGATAATGCTGAAGCTGATGCCTTGGCAAATCTAGGATCATCAGTCAGGATAC
CGGAAGGAACCCGAATACCGATATTACATATCCTATATCCTGTAATGGATCCCCGGGTACCGAGCTCGAA
TTCGTAATCATGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAACATACGAG
CCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTC
ACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCCAACGCGCGGGGAGA

261D17 – B4- B12

Forward: 2128 bp

AGTCTTGTTAAGCTTCCAATCTGTAACCTCAACGGGTTCTTCGGTAAAGTGGAGTGTGTCGTCAATGTGC
ACTTCATCCGCAGGGATGACCACTATCTCTTGTGTTGGACTCTTCTTCAGATTAGATACATGAAATGTAT
CATGAACACCATTAAGTTCTGCTGGTAGCTCCAACTTGTATGCTACAGAACCAACCCTTTCCAAAATTCT
AAACGGTCCGATAAATCGTGGGTTTAACTTACCACGCTTCCGAAAGCGTGCCACACCTTTCCAAGGTGCA
ACTTTTAAATTAACCATATCACCCACTTCAAATTCCAAGGGTTTTCGTCTTTGGTCCGCATAACTCTTCT
GTCGATCACGAGCCGCCTTGATGCGCTCTCGAATCTGCATGATCTTATCAGTTGTCTCTTGGACCAGTTT
GGGACTAACCATCTGTCTGTCACCTGCATCTGCCTAGCATAACGGCTATTGACACTTGCGTCTGTAAAGA
GCCTAGAACGGTGCAGCTTGGATGCTTGTATGGTAGCTGTTGTTATAGGAGAACTCAACCAAAGGCAAAT
GAGTATCCCAACTGCCACCTAAATCCATTACACAAGCACATAGCATATCTTCTAGTGTCTGAATCATTCT
CTCGCTTTGACCGTCCGTCTGCGGGTGAAAAGTTGTACTCAGATTCAACTGAGAACCAAAAGCTTCTTGC
AAGGATTGCGAAATCCTTGACACAAACCGTCCATCTTTGTCCGATATAATCGAGATAGGCACTCCATGAC
GTGCTACTATCTCCCTTAGGTATAGCTCAGCTAGCTTACCAGTGTTATCCTTCTCTCTAATCGGCAAGAA
GTGCGTAAACTTTGTCAGTCGGTCCACGATTATCCAAATCGTATCATGACCTCTGGGAGTCTTGGGCAAC
TTGGTTATGAAATCCATAGAAATCTGCTCCCATTTCCACATGGGTATTTCTGGCTGTTGCAGAAGACCTG
AAGGCTTCTGATACTCGGCCTTCACCTTGGCACAAGTCAAACACTTTCCAACGTAAGTAGCAACATCACC
TTTAAGTCTTGGCCACCAGTAATAATCCTTCAAATCTTGGTACATCTTGTCCGACCCTGGATGGATCGAG
TACCTTGATTTATGAGCTTCATCGAAAATAACCTCTCTAAGACCTCCAAAGAGCGGAACCTAAATACGTC
CCATAAAATACAAGGTTCCTTCCTCGTTGGGTACCAACTTTTTCTCCATCCCACGGAGATATTCATCCTT
AAGGTTCTCCTTCTTAAGAGCTTCTCTTTCTGCCGCACGAATACGCAACGAGAGATCTGTCTGAATAATC
ATCTCTAAAGCCCTAACCCTTATGGGCTTAATCCTTTCTTTTCGACTTAAGGCGTTTGCCACCACATTCG
CGTTTCCTGGATGATACTTAATCTCGCAGTCATAATAGTTTAACAACTCAACCCATCGTTGTTGTCTCAT
GTTAAGCTCTTTCTGATCAAATATGCGCTGAAGACTCTTGTGATTTGTAAAGATCGTACATCGGGTACCA
TACAAATAATGTCGCCAAATCTTCAAAGTAAACACCACTGCGCCTAACTCCAAATCGTGCGTGGTGTAGT
TCTTCTCATGGATCTTCAATTAGCGCGAAGCATAAGCGATAACTTTTTGTCGCTGCATTAGCACACAGCC
CAATCCCTGACGCGAGGCATCGCAGTATACCACGAAGTCGTCGGTGCCTTCGAGTAGGGATAATATAGGT
GCATCGCAAAGCTTATTCTTAAGAAGCTGAAATGCTTCTTCTTGTTTGTCTCCCATTCATACTTCTTATC
CTTTTGAGTAAGGGAGGTCAGCGGTTGAGCGATCTTGGAGAAATCCTCAATGAATCTGCGACAGTACCCT
ACTAAACCCAAAAATTGTCGGATCTCGGTTGGCGTCTTCGGAGTTTCCCAATTCTTAATCGCCTCGATCT
TGGTGGGATCCTCTAGAGTCGACCTGCAGGCATGCAAGCTTGGCACTGGCCGTCGTTTTACAACGTCGTG
ACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAA
TAGCGAAGAGGCCCGCACCGATCGCCCT

Reverse: 1026 bp

CGGATGAAGCTCATAAATCAAGGTACTTCGATCCATCCAGGGTCGGACAAGATGTACCAAGATTTGAAGG
ATTATTACTGGTGGCCAAGACTTAAAGGTGATGTTGCTACTTACGTTGGAAAGTGTTTGACTTGTGCCAA
GGTGAAGGCCGAGTATCAGAAGCCTTCAGGTCTTCTGCAACAGCCAGAAATACCCATGTGGAAATGGGAG
CAGATTTCTATGGATTTCATAACCAAGTTGCCCAAGACTCCCAGAGGTCATGATACGATTTGGATAATCG
TGGACCGACTGACAAAGTTTACGCACTTCTTGCCGATTAGAGAGAAGGATAACACTGGTAAGCTAGCTGA

GCTATACCTAAGGGAGATAGTAGCACGTCATGGAGTGCCTATCTCGATTATATCGGACAAAGATGGACGG
TTTGTGTCAAGGATTTCGCAATCCTTGCAAGAAGCTTTTGGTTCTCAGTTGAATCTGAGTACAACTTTTC
ACCCGCAGACGGACGGTCAAAGCGAGAGAATGATTCAGACACTAGAAGATATGCTATGTGCTTGTGTAAT
GGATTTAGGTGGCAGTTGGGATACTCATTTGCCTTTGGTTGAGTTCTCCTATAACAACAGCTACCATACA
AGCATCCAAGCTGCACCGTTCTAGGCTCTTTACAGACGCAAGTGTCAATAGCCGTTATGCTAGGCAGATG
CAGGTGACAGACAGATGGTTAGTCCCAAACTGGTCCAAGAGACAACTGATAAGATCATGCAGATTCGAGA
GCGCATCAAGGCGGCTCGTGATCGACAGAAGAGTTATGCGGACCAAAGACGAAAACCCTTGGAATTTGAA
GTGGGTGATATGGTTAATTTAAAAGTTGCACCTTGGAAAGGTGTGGCACGCTTTCGGAAGCGTGGTAAGT
TAAACCCACGATTTATCGGACCGTTTAGAATTTTGGAAAGGGTTGGTTCTGTAGCATACAAGTTGGAGCT
ACCAGCAGAACTTAATGGTGTTCATGATACATTTCATGTATCTAAT

261F19 – B5- B6

Forward: 3953 bp

CTTACCAAAAGGATCTAGAACTGTACAGGGAGACCCAAACGGTTCTAAATACTTTAGGTTCGACTTGCGA
TTATTGATCAGCTCGAAACATGTTTTGTTGAATTTCTTGACAGTGAGAACCCTGTTGAGTGTATAGCAAG
CGGCTGAAACTGCTTCTGCCCAAAAATTGACTGGTAACTTTGAATCAGCAAGCATTCTTCTGGCCGTTTC
AATTAGTGTCATGTTCTTTCGTTCTGCTACTCCATTTTGCTGCGGAGTGTACGGAGCACTAAATTCTTGC
AATATACCTCTCTCATCACAGTATTCGTCCATTCGGTTGTTCTTAAACTCAGTACCATTGTCACTTCTGA
TTCTTCTGATTCGTGCTTGGTACAAGTTTTCAATCTTTCTGAACAATGCCATAAACCTATTTTATCATTA
ATATAACTTGGTAATTAATGTAAGAACATGTCTAAACATGTTCACCCCGACAATTTTCAGTTTAAGCTCG
GTTCGGAACCGAAAATCGCAAAAGTAGACTTTTGCTTTGAGTTTCAGTTCTGACCCATTTGAGCATAGTT
TAGAAATTCCTTAGAACTTATTTAGGACTAGGTTTCATGTTAGTATAACCCTCTGTGATTACACAACTTG
ATTCATTAGTTATCCGATTCACATGCATGTTCCCATTAATTGCTTAAATGTTGACTATTATGCCCATATG
ACCTTAAAATGTGATTTTTGAAAATGTAAAAGAGTGGACACCTTAGTTACTGATTTATAAGCTTGTATCT
AAAATTTGACACTAGTTTAAGTTCTAGGTTAAGAGTTATACTCATTAGCGTAATTAGAAAGTTTTTAGTA
ATTAAACGACGTAATTAGCATATAGCCTATCTAAACCCAATTTTTGATACCAAACTTTTTACCCACTGAT
GTAAAATAATATTTTAGGATTTTTGTAGATTGTTATTTATTTTTAGGCTGAGCATAACCTAGAGTTCTAA
GCTTAATTCGGTAGTTGTCGGTTTTACCCTTCTAGGCTATAAAAATGAGTTTTACAAATCCTTTTGGTAC
CAAACCTTTTTCTACTGATCTAATATGATAAATAAAATGTTTTGAGCCTTCTGGAATAATAAAAACATCC
GCTTTCCTTTAAAAACCGGAAATTGCTCCAAATCGCCTTTTTAAGCGTTTTTAACGCATAGTGTGTATTA
AAACTATTTTAAACATATAAGGTTTGATGCCTACTGATATTTTCAGTAAATTTTTATATTTTAACACTAA
GCAAAAGTTTTAAACTCAGATTTCCAGTTTTAACCTTTAGGCTTATGTGAAATTACCAAAATGCCCCTAC
GGTGCATAGTATGGTTTTAATCAATAAATTTCACATATATATGATACCCTACTGTTATAACTTATTAAAT
TAAGTATATTTACTCATTAAATCAGACCTGTAACTCAGATTACCATTTAAACTCTTTTATAACCTTTTAA
ATGACCAAAATGCCCTTACGGGGCATAGTTTGAGTTTAAAATCATTTTGGGCATAATAGAAGATATCTTA
CTGATGTCACAACATATTTAAGTCATATTAACTTAGGAAACCTGTATATGACTCTTATGGTTACCCGGTT
ACGCATTATTCGCGTTCGGTTCGGTTTCTGTAACTAGTTTGCATAAATTGACCAAAACGGGTCAAACCTT
ATCATTTTTGTCTCAAAATCCAGAATGTGTTTAGTTTACCCATGTTATATAAGTATCCAAACTTGTCGGG
TCTAAATCACATTCTATTCCGGTTTTCGCTTAATCATGCGTTTGAACCGTATCTATCTTTAAAACTAACC
GGTCTAAGATTAGGCAAAATTTAAGACCGTTAGGAATGTAATAGGTTATTATAAACCTTCGTTCCAGAAT
AGGAGACCCGATAAAAGCTACGTGCACTTGCTTATTGTGATTACTACTTGCAAAGGTAAATACTTTTAAC
TTATTTTCCCTTATACGGGCTTGGGGTACGGTATATAAAATACCGCTTGGTCGGGCATTTGACTTTAATC
TTTTGGTGGTTAAATATTATAAAGATGACCCGTTTAAAAATTTGGTTTTGTTTTCTTAACGCCTTTGGGG
GTTAATGACCGTGTCCCGGATATCCTTGGCATCTTTCATAGAATGGCCACGACCTTAGCATACGAGTGTA
GGCGTACACCCGTCAGTGCATTTATAATTAATAAGGTATAACCGCCGGTTACGAGAATAACTCGCCGCGG
GACTATACCATGTGGTGTGTCTATTAATCTTTAACCCGGAATGAACCGGGCAACTGAACGCATAACAAAC
ATGTAATTCTTTTACAAGATTATATACAAGTAATTATCCCAAGTTATAAAAAAATCTCTTGTGCCACGTG
CATTCAAATCAAATTTTAAACATTTTCAAAATGAGTCAATTAAATTGTATTTACCAGTGTAAACTGTTGG
ATCGTCGTGTGACCCTAACGAGTCAATCTGAAAAGTTCAACATCAAAATCAAAATCAAAGGCGGAATCAA
TGAAATTGGCGTAACACAGCTTGTTTTCCTCTAATTATCTTATTATTAACCGATTCTGAGCTTTTACAGA
TTCAAATGACAAACGGACAGCACCTCGGCTCTGAATCACACACACTGATTCCGTTTAAAATGAAACATCA
TTTTCAGTATTTATAGGCATTGTAATTCCGCTTGAAACCATCTCAAACGAAATTACCAGTTCATGCGGAA
TTACTTAATCCCGTTCGGAATTAGTAATCTCGCTTGAACTCACTAATTCCGTTTGAAATGTCATTAAGCA
ACTTTCAAGCGGAAGTACCTATATATTTGATTTCTCGTTTTGTGTGCCTTGATCTATCTAATACAATACA
AGACTTTATAAAGACGAAGTTAACAGACATCAGTGCACCAACAGACTCCCCCTTGGATGTTGACGAAGTC
TTTAGCATCAAGTCTTTAGTCTTTAGACTTGATCTATCGCTCTTTGCCCTAAATTGTCTTTGACATTGAA
AAGCATCCTTCAATCTTCTATCTTCAATCTCCCCTTTGAATGTTGATCCAGAATTTTGACTCCTCCTCTC
GTAGACTCCCTCTCTCGTAGACTCCCTCTCTCGATAAGCTGGGACTTAAATTCTTGTTCAAACTTTGACA
TTTTCTTTACCATGGGAATCTTGACCCCTGTTTCCTGTTTCTTAAACATATAACATTATAATCTATTCAA
TCTTAAACTTCTCTTAATTGGTCATCAGTTTTAGAAATCCACTTAAGAATTCAGGTCCAAGTTTATGACC

CTGATTACTCAATTTATCCACCAAGTCCAACTTAATGACCTTGATTGATATGTGACAAAAACAAACTGAT
TTATGTTCAACAATTTTCTAACAATTACAACTTTCAAGTTAATGAACTTGTTTTGACTTTCCATTTTCTC
AATTAACATGTTAAGCTCCCAACTCATCATCCAGTTTAGAATCTTTCAGCATCTGCTCCAACCTTCTGAA
TCTGAAGATTAACTCTCCACTTTGGTTTGTCGGAAAATAAAGCAGAAATGAAAAATCCTTTTGGATTTGG
AATTAAAGTTTCTAAACTAGAAAGCTATAAAATCTTTTTGAGTTTTGGAATTTTTAACAGAAATGAAATA
CAGAATGTAAACTATTTACAAACATATTTTGGTGAGCGTGTTAGGGAATCATATCAGTCAATGACAAGTT
ACTAGCACCGTTAAGCTTGATACCATACTAAGTCTTAAACAATTCACTTAGATTGTCGATATGCTGATCC
TCTTAAATTTTCATACAAACTTCAATCTATTCAGGATACGATTTAGGTGAATTAAAGACTTAAACTTATA
TGTGTGTCCCACCTCTTGAATATACTCCCGGTCCAGATCCCCATATTCAATCTTACAGGGAGTATACCAC
AGATGATATCTGTAAGGGGTTAAATGCGAAACC

Reverse: 3247 bp

AGAAAATTGTTGAACATAAATCAGTTTGTTTTTGTCACATATCAATCAAGGTCATTAAGTTGGACTTGGT
GGATAAATTGAGTAATCAGGGTCATAAACTTGGACCTGAATTCTTAAGTGGATTTCTAAAACTGATGACC
AATTAAGAGAAGTTTAAGATTGAATAGATTATAATGTTATATGTTTAAGAAACAGGAAACAGGGGTCAAG
ATTCCCATGGTAAAGAAAATGTCAAAGTTTGAACAAGAATTTAAGTCCCAGCTTATCGAGAGAGGGAGTC
TACGAGAGAGGGAGTCTACGAGAGGAGGAGTCAAAATTCTGGATCAACATTCAAAGGGGAGATTGAAGAT
AGAAGATTGAAGGATGCTTTTCAATGTCAAAGACAATTTAGGGCAAAGAGCGATAGATCAAGTCTAAAGA
CTAAAGACTTGATGCTAAAGACTTCGTCAACATCCAAGGGGGAGTCTGTTGGTGCACTGATGTCTGTTAA
CTTCGTCTTTATAAAGTCTTGTATTGTATTAGATAGATCAAGGCACACAAAACGAGAAATCAAATATATA
GGTACTTCCGCTTGAAAGTTGCTTAATGACATTTCAAACGGAATTAGTGAGTTCAAGCGAGATTACTAAT
TCCGAACGGGATTAAGTAATTCCGCATGAACTGGTAATTTCGTTTGAGATGGTTTCAAGCGGAATTACAA
TGCCTATAAATACTGAAAATGATGTTTCATTTTAAACGGAATCAGTGTGTGTGATTCAGAGCCGAGGTGC
TGTCCGTTTGTCATTTGAATCTGTAAAAGCTCAGAATCGGTTAATAATAAGATAATTAGAGGAAAACAAG
CTGTGTTACGCCAATTTCATTGATTCCGCCTTTGATTTTGATTTTGATGTTGAACTTTTCAGATTGACTC
GTTAGGGTCACACGACGATCCAACAGTTTACACTGGTAAATACAATTTAATTGACTCATTTTGAAAATGT
TTAAAATTTGATTTGAATGCACGTGGCACAAGAGATTTTTTTATAACTTGGGATAATTACTTGTATATAA
TCTTGTAAAAGAATTACATGTTTGTTATGCGTTCAGTTGCCCGGTTCATTCCGGGTTAAAGATTAATAGA
CACACCACATGGTATAGTCCCGCGGCGAGTTATTCTCGTAACCGGCGGTTATACCTTATTAATTATAAAT
GCACTGACGGGTGTACGCCTACACTCGTATGCTAAGGTCGTGGCCATTCTATGAAAGATGCCAAGGATAT
CCGGGACACGGTCATTAACCCCCAAAGGCGTTAAGAAAACAAAACCAAATTTTTAAACGGGTCATCTTTA
TAATATTTAACCACCAAAAGATTAAAGTCAAATGCCCGACCAAGCGGTATTTTATATACCGTACCCCAAG
CCCGTATAAGGGAAAATAAGTTAAAAGTATTTACCTTTGCAAGTAGTAATCACAATAAGCAAGTGCACGT
AGCTTTTATCGGGTCTCCTATTCTGGAACGAAGGTTTATAATAACCTATTACATTCCTAACGGTCTTAAA
TTTTGCCTAATCTTAGACCGGTTAGTTTTAAAGATAGATACGGTTCAAACGCATGATTAAGCGAAAACCG
GAATAGAATGTGATTTAGACCCGACAAGTTTAAATGGTAATCTGAGTTACAGGTCTGATTTAATGAGTAA
ATATACTTAATTTAATAAGTTATAACAGTAGGGTATCATATATATGTGAAATTTATTGATTAAAACCATA
CTATGCACCGTAGGGGCATTTTGGTAATTTCACATAAGCCTAAAGGTTAAAACTGGAAATCTGAGTTTAA
AACTTTTGCTTAGTGTTAAAATATAAAAATTTACTGAAAATATCAGTAGGCATCAAACCTTATATGTTTA
AAATAGTTTTAATACACACTATGCGTTAAAAACGCTTAAAAAGGCGATTTGGAGCAATTTCCGGTTTTTA
AAGGAAAGCGGATGTTTTTATTATTCCAGAAGGCTCAAAACATTTTATTTATCATATTAGATCAGTAGAA
AAAGGTTTGGTACCAAAAGGATTTGTAAAACTCATTTTTATAGCCTAGAAGGGTAAAACCGACAACTACC
GAATTAAGCTTAGAACTCTAGGTTATGCTCAGCCTAAAAATAAATAACAATCTACAAAAATCCTAAAATA
TTATTTTACATCAGTGGGTAAAAAGTTTGGTATCAAAAATTGGGTTTAGATAGGCTATATGCTAATTACG
TCGTTTAATTACTAAAAACTTTCTAATTACGCTAATGAGTATAACTCTTAACCTAGAACTTAAACTAGTG
TCAAATTTTAGATACAAGCTTATAAATCAGTAACTAAGGTGTCCACTCTTTTACATTTTCAAAAATCACA
TTTTAAGGTCATATGGGCATAATAGTCAACATTTAAGCAATTAATGGGAACATGCATGTGAATCGGATAA
CTAATGAATCAAGTTGTGTAATCACAGAGGGTTATACTAACATGAAACCTAGTCCTAAATAAGTTCTAAG
GAATTTCTAAACTATGCTCAAATGGGTCAGAACTGAAACTCAAAGCAAAAGTCTACTTTTGCGATTTTCG
GTTCCGAACCGAGCTTAAACTGAAAATTGTCGGGGTGAACATGTTTAGACATGTTCTTACATTAATTACC
AAGTTATATTAATGATAAAATAGGTTTATGGCATTGTTCAGAAAGATTGAAAACTTGTACCAAGCACGAA
TCAGAAGAATCAGAAGTGACAATGGTACTGAGTTTAAGAACAACCGAATGGACGAATACTGTGATGAGAG
AGGTATATTGCAAGAATTTAGTGCTCCGTACACTCCGCAGCAAAATGGAGTAGCAGAACGAAAGAACATG
ACACTAATTGAAACGGCCAGAAGAATGCTTGCTGATTCAAAGTTACCAGTCAATTTTTGGGCAGAAGCAG
TTTCAGCCGCTTGCTATACACTCAACAGGGTTCTCACTGTCAAGAAATTCAACAAAACATGTTTCGAGCT
GATCAATAATCGCAAGTCGAACCTAAAGTATTTAGAACCGTTTGGGTCTCCCTGTACAGTTCTAGATCCT
TTTGGTAAGTTTGCTCCAAAGAGTATCGAGGGAATATTTGTGGGTTACTCAAGTCCACTAAGACGAGTCT
TTGTTCCAAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTGGGCGTAATCATGGTCATAGCTGTTTCC
TGTGTGAAATTGTTATCCGCTCACAAT

67N4 – B2- B4

Forward: 1509 bp

GTCTTTCTTGTGTTTTACGATGCCTTTTTCTTCTTTGAGGGCTCATCTCCAGTTACTTGCTTTCCTTTTC
TTTTTCCCTTCAAACGTTCTTCCTTTTCAGCAGCCTCTGCTTGAGCTTGTGAAGCTCTTTGTTGTTTCTC
TCGAAGTGCCTCTCTGTATTTTCTGTAAGCCTCCTTCACTAATGGCAGCCTATCTGCAGAATACATTGAA
TCATGCAGCATTACCCATTCATTTGTCTTGATCTCTGTGTACTGACGACTTGTTGGAACATGAAACGTGT
AGGGATCTCCATCATTTGGCGGATTTTCATATTGGTCGTTGATCAGCGTTTGGATAAATCTGGGATACAT
TGCCCAAGTCGGCCCTCTCACATTCTCGATCATATAATTGAAGATTAGACCAGAAAGGTTGAACTTATGA
TTCAAACACAAGTTTAACATTGCTGCAGACCATTCTAGATTCAGTCCATCATAACCACCCTTACGATGGG
CTAAGCTCATTGCAATGACGTGAACTATGAATCTCCAGTCTCTTGTTAGCCCGTTCCTCTTTATCTCTTT
CTTTTGACTAAAGTCTCCTGCATAACCCATCCCGCGAAAACCATCCAAAATATCGTCCTTGCTCAAGCTG
ATCGGATCTTGAGCGTTATCTTTCAGCTTAATAACTTTGCGGGTTTTTTCTTCCGTTACAACTATCTTTT
TGTTCCGGACGGTTACAGATATTCCTCTGACATTGTCGACAGTTTTGGGTTTTGCAGTTTGCCAGAAGTC
TTGTATATGGATTCTGTAAATGCGCACTTCAGTTTCAACCGCGTAGCCAATCCTGCTTGTTCTTACCCAT
CTAGCTGCATCCTAGAAATCCTGTGGAACTTGCTCGAGATTGATACTTTGATTGTGCCTTAACGTCTTAT
CCCACTCAATGTTCTTCACCATTCTGTGTTGGACACGAAAAACAAGGCAAGTTAGAAAGTTGTCGAGATA
CAAGTATATTTTTTTTACAGGTGAAGCACTATATAAAAAAAACTTTACAGAAATTTACATTTTTATAAAA
GAAAGTGAAAATAGAGAATATCTTCAAAAGTATAAGTTATAATAAAAAGTTAAATTAGGTATTTACAAAA
ATTTAAGACTTAAATTCGCTGACAGCCCTATCGACTGTCTTTGCTGATGTTTGTGATCGGGATCCCCGGG
TACCGAGCTCGAATTCGTAATCATGTTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCC
ACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCACATTA
ATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAGCTGCATTAATGAATCGGCC
AACGCGCGGGGAGAGGCGGTTTGCGTATTGGGCGCTCTTCCGCTTCCTCGCTCACTGACTCGCTGCGCTC
GGTCGTTCGGCTGCGGCGAGCGGTATCAGCTCACTCAAA

Forward: 1275 bp

TTCTAACTTGCCTTGTTTTTCGTGTCCAACACAGAATGGTGAAGAACATTGAGTGGGATAAGACGTTAAG
GCACAATCAAAGTATCAATCTCGAGCAAGTTCCACAGGATTTCTAGGATGCAGCTAGATGGGTAAGAACA
AGCAGGATTGGCTACGCGGTTGAAACTGAAGTGCGCATTTACAGAATCCATATACAAGACTTCTGGCAAA
CTGCAAAACCCAAAACTGTCGACAATGTCAGAGGAATATCTGTAACCGTCCGGAACAAAAAGATAGTTGT
AACGGAAGAAAAAACCCGCAAAGTTATTAAGCTGAAAGATAACGCTCAAGATCCGATCAGCTTGAGCAAG
GACGATATTTTGGATGGTTTTCGCGGGATGGGTTATGCAGGAGACTTTAGTCAAAAGAAAGAGATAAAGA
GGAACGGGCTAACAAGAGACTGGAGATTCATAGTTCACGTCATTGCAATGAGCTTAGCCCATCGTAAGGG
TGGTTATGATGGACTGAATCTAGAATGGTCTGCAGCAATGTTAAACTTGTGTTTGAATCATAAGTTCAAC
CTTTCTGGTCTAATCTTCAATTATATGATCGAGAATGTGAGAGGGCCGACTTGGGCAATGTATCCCAGAT
TTATCCAAACGCTGATCAACGACCAATATGAAAATCCGCCAAATGATGGAGATCCCTACACGTTTCATGT
TCCAACAAGTCGTCAGTACACAGAGATCAAGACAAATGAATGGGTAATGCTGCATGATTCAATGTATTCT
GCAGATAGGCTGCCATTAGTGAAGGAGGCTTACAGAAAATACAGAGAGGCACTTCGAGAGAAACAACAAA
GAGCTTCACAAGCTCAAGCAGAGGCTGCTGAAAAGGAAGAACGTTTGAAGGGAAAAAGAAAAGGAAAGCA
AGTAACTGGAGATGAGCCCTCAAAGAAGAAAAAGGCATCTGAAAACCAAGAAAGACTCATGGGTGCAAGC
TTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATT
GTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATG
AGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTGCCAG
CTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGTTTGA
CAGCTTATCATCGAA

216F17 – B10- B11

Forward: 1482 bp

ATGCTTAAATGAGTGATTTTGGATTAGAGATGGTGTTTTTCCATTTAGATCTGCACAATTTGCCCAAAAT
CAATAATGCTTACGTACGTTATTGTCTTGCTATCTCCAAAAGATTGATGGTTAGTTTGCAAAAACGTGTT
AATTTAAAAAGTCAATACGATAAGCATGTCAAACTATATATATATATCTTCAGGTTTTAAATGTTAACAA
AATGGTTATAAAAGTGTTCTAATTATGTCACTCGTATCATTTTTGTCCCAATTGGATAACTTAACATTTA
AGTCGAGCCATATCTTTTTTTCCATGTGTCAAGAAAAAAATGAAGCATACGACGCCAGGGTCGGATTATT
TTAATAATAGAAATTATATTTATTAAAAAGAAAAACACTTTAATTACATTAAAACTTAAAAAAAACAAGT
TAAAATCCACGTTATCACTCGACGTTAGCATCAAATAAAAGGAAAAAAAGAAACAAAAAACTACGTAACC
ACGGTTACCTATGAAATGAATGAAAAAAACTCTTAATTTTAATGAAAATGAATGTTGAGTGACATGATTG
AAACATTTTTAAAATTTGAGTGACCTATTGTAACACTTTTAAAACTTAGTTACTATTATACAAAGTTACC
CCTATACTATTAACTATAAACCCATGTTGATCCTTTAGTATGTCTGACCTCGATTGGAAAAGTACACCTA
TCTTACCTGCTGTGTGTTACAACTCCTTTTGCTGCTTCTTCTTCGTCAAGAGACAAGACAGCGTTTAATG

CACCTCTGAATTTAAAAGCTCCACTGATAACCATCAACAAAATACGTTAAAAATTAACATGTTCATTTTA
CCACCTCTAGATCGATAAAGATAACGCAGAAGACCTCTCTTACCTCTTTTGAAAGATCTCACACTTGAAA
TATAGTGATTTCCCAGCAATGGAATTTAGGGACTCTGAAGACAAGACCGGAGTTACGTGAATATGGGCCT
TGATTACATCTTGCGCTTTGGTTATGGAGTCGACATCAGCGGCATATTTGTAACCTCCCATGGTGAGTTT
TGTTTAACACAGGATCAATAACTTCAGTTACTGCAACCAAAAAAAGATATAATATGAGAAGCTAACGTTA
TACTCAAAACACTTCTCTTCAATTTTGATATCGGTACCGTAATCTCATCCCATTTTCCCACTAAAACACT
AAAAGATTTTTATTACAAAAGGAGAAGGTTTCAAATAATTAGTGGATTACCTTAGGTTGTAGAAGGGAAT
CACAAAATAGGTCGAAGAAGGTCAGTGGTGAGCAAAAACGTGTAGGGAGTAACCATATATATAAGGGATT
TAAATTTAAATTTCCATATAAAAAAATATATTGACACATGTTAAACGTTTATAAAGGGAAGGTGAATAAA
TTCAATATACAGGCATTCTAAGATTAATTCTTCATGCGTTGGGTTCATATGGGCTTTAGATTAGTGGATT
ACCTTAGGTTG

Reverse: 712 bp

CATTCATTTCATAGGTAACCGTGGTTACGTAGTTTTTTGTTTCTTTTTTTCCTTTTATTTGATGCTAACG
TCGAGTGATAACGTGGATTTTAACTTGTTTTTTTTAAGTTTTAATGTAATTAAAGTGTTTTTCTTTTTAA
TAAATATAATTTCTATTATTAAAATAATCCGACCCTGGCGTCGTATGCTTCATTTTTTTCTTGACACATG
GAAAAAAAGATATGGCTCGACTTAAATGTTAAGTTATCCAATTGGGACAAAAATGATACGAGTGACATAA
TTAGAACACTTTTATAACCATTTTGTTAACATTTAAAACCTGAAGATATATATATATAGTTTGACATGCT
TATCGTATTGACTTTTTAAATTAACACGTTTTTGCAAACTAACCATCAATCTTTTGGAGATAGCAAGACA
ATAACGTACGTAAGCATTATTGATTTTGGGCAAATTGTGCAGATCTAAATGGAAAAACACCATCTCTAAT
CCAAAATCACTCATTTAAGCATCTGTTTGAGTTCTTTTTTCGTTTCACTAAGAAGCTTGAGTATTCTATA
GTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCTCACA
ATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAACTCA
CATTAATTGCGT

225D9 – B21- B23

Forward: 5138 bp

GTTTAACATAACCATCTGCAATAAAAACACATCATAACTCTCCTATACATTATATAGTTCAAAAAATCAA
ACATATAAAAGGATAAAATACCTCCATAATATATCGCCGCATGACCACTGCTGTAATGGTTAAGTGCACC
AAAAAGGTCCGTATTGTACGCCAAAAGCAAGGTAGAAAGTGTTCCGGGGACTTTAGCCGTTAGGTGCATT
TTCAGGAAGGCTTGGTATAAAGCAGCAATATGGGTTGTCTGACCCACAATCAGATAGTGAGTGAAGAAAT
ACATCGTCTGGAGGATTAAAAGTTGGATATACCATCTGATTGCTAAAGACTTCATAAAGCTGCCTACACG
TCTAGAAAACCATGGTTTGACCCCTAGGACGTTTAGTTGATCCGCCATGCCTAAGAACATTGCGAACCAA
GTCAATGGATTCCAAGCTGGTTTTTCACTTAAGCTGTTGTCCCAGCTCAACATTCTAAGTATCAGTACAA
CCGATAAGCCCATCATAGTTGCAGCCAAAGTTGAAATGTCTACCCTTTCTCTGCATCCACAAAAACCAAC
CATAAGCTACATAATCAAATGGTTACAGACAAGTTGATCGAGCTAGTTAAAGCTTACCCAAAGATCCACA
AAGTGACAGTGACAAGCATGGTTCCCAACATCATCCACTCGTTCCTTTTAACAAGATCCATTTGCTCCAG
TTTTTTCTTAGCCTTAATAGGAGCATCCCTGGTGTACTTTGTTTTAGGCATGGTTGCAAAAGTCGCTAGG
CTCTCCCTAGTCGGTCGACCAGGGAGTTGAGAGTACTCGGTCTATGCGGAGAGTACTCGAGGAGTACTCG
GACGTGTTAAATTATAAAGAAATTATTTTTTGGAGATTAAATATATGTCAAATAACACAAATTTACTAAT
ATTTATAACAAAATATGTGAAAATCATATTCATTCTTTAATATGATAGGCATAGAAATTATCTTATATTG
TTATTTAAGTCAAACTAGGCCCGCGTTGACCTACCAGATCCAATTCTGGCCGAGGTTGACCGCGTTTGAC
CGACTCCGAGTAATTAGGCGAAGTCAAAGAAAGTCGCCTCGGCAGCCTGCCTTGTAGCGACTACTCGGGG
AGTACTCGGCCTTGGAAACCTTGTTTTACAACCATGGTTTTAGGGGAAAATTTTTGTATACAATGAGTGG
AGTGGGCATAAGAGAAGCCAACGCGGGGAAACAGAATTCAACAACTTTTAGGTTAGGTAGGGGTATTAAA
AAAATCCGAATCCGAAACTGAATCCGAAATATCCAAAAATTCGTATCCAAAAATGTGGATGTCCGAAAAC
TGGATAATCCGAATTTCCAGATTCAGATGTGGATTTCTAAAATTTTGATTATCCAATATCCGAAAACTAA
TATTTTTTTAATATATATAGTATATGAGCACTATATTTATTTCGAAATATAGTAAAAAATAGTAAATAGT
CGGATTCGGATGTGGATTTATAAAAGTTAGTTATTAACTTTGAAAATTTGAAAAATCTAACATTGGTTTA
GCTTTAATATATGCACGAAACAAATTTTTTGATTTTTTTTTTTATGAATTCGGATATCCATTTGGGATGGA
GCTGTAGAGCATAGAGCTGTAGCGCAGCTAGATGCCTGTCGATCCAGGGTGTACTTTGTTTAAGGCGGAA
ATTTTATATACGATGAGTGGAGTGAGCATAAGAGAAGCCAACACGGGGAAACAGGAAGCTCTATTCAGTC
GTATTGAAACAATATAGTATTTTAATTTAAACGGTATTGTACGGAAACAGTGTAATATATTCAAACCGTA
CCCGGGTATTTCAAAATAATTAATGTTATAAAAATTTAAAACTGCTATAATATTTTAATTTAAACGCTAT
AGTATGGAAACACGGTGTAATATATTTGAATTGTACCTGAGTAATTTAATAAAACTAACGGTATGAAAAT
CTTAAACGGTATAATATTCGAATTGCACCAAGATTTGGTGTAATAATTTAATTTAAACGGAATAGTATTG
AAATAGTATAATATATTCGAACCGTACCCGAATATTTCAAAATATTTAATGGTTTTAATTTAAACGGTAT
AGTATGGAAACGGTATAATATATTCGAACCGTACTCAAGTATTTAAAAATAATTAACGGTATGATATGGA
AACGTTTATAGCATTTGAACCGCACTAGATGTGGTATAATAATTTAATTTAAACGATAACAGTATAATAT
ATTCGAACCGTATATACGTTTTTAAAAAAATAATTAACGGTATATAATATTAATGGTATAAGATTAACGT

TTATATATTTTAAGCGATTTATGTTGGAACCGGGTGAGAATTTATACCACAAATTCATCAAAATAACCAA
ACAAAATTCTGATGGGTTATATACGTTGTGTATAGCTCTATACGCAGCGTATATAAACTCTGGTTGTCTG
TGCCAATGATTGCTGGGCAGGCTCTGAATTCTATGCAACTTATACGCTGCGTATAGCTCTATACGCGTCG
TATAAGTGTAACCCTATTCGCTACGTATAGAGTTACACTCGGCGTATATAAACACTAAGTGTGCAAATAG
GCAGCCTTGGTGTGTGAATAGTTTCAGCCTTGAATAAAATTGACTCCATTAAAGTGTGTAACGTATGTAT
GATGATTAGTGTGAGTCTCTATCATATATTGATGTTTGTGCTACTAGGTTAAAAGATCAAAGAGGAAACG
GACAACAATAATAATAATAATAATAATGATAATCAAAACATCATGCTTTTAATTGCTTAAAGGTCTT
TCAAGTTTCAAGTAACAAACAAAAGAGGTTTCACAATTACCCTCCAGAGTTATTTAAATAAGGTTCACAA
GTCGTCAATTTTTCCAGGATTCAAACAGAACGCCCAGGTCCGCATTGCCCCCAGAAAGCACAATCCCAAT
GTTTCGAGAACAACTCAAAATAGGATTTTTCTTGAAGTTATCGGATAGAACCGCAGCAAGCCCTATAGCA
GCACTGGGCTCAATTGCAACCTTGAGTATCTGATAACAATATCGAGTTGCTTCAACAATGTCCTTATCTT
CCACGGTTATAACGGCATCAACGAGATCCCGTACAATGGGCCTACATAAAAATAGTAACACTCAAATGTT
AAACATGAGTCTAATTTGAACAATAAACCAGATGCAACAATAAGGCGTAAGTGATTATTTTACTAATGCC
TTTTCAAGTAGTTGATAACCAATTTCTAATGTTTTGAGTATAGTAAGACTGGTTTTGATTATTTGATCAT
GTAGGGAAAAATAACTATGAATATAAGTGAAACAGGTAAACCACATTTAATGTATATTTTAAAGCGGTGA
AGAGATATAAAGATAATATTACCAAGTAAGATCTCCAAGGGAAGCTCGAAGTCCATCAGCAATTGTGTTT
GTTTGACTCAAGGTTATGAGTCTACCGCATGCTTTGGATTGTGCTGCATCGTCTGCTCCTTTTGGCTCAG
CCGCAAAGACCCGTATGGCAGGATTCTTTGCCTTGGCTCCAATCACCGTACCAGATGTTAAACCACCACC
TACAAGTTTCTACTTCAAATGAAAAAGGAAACCAACTAAACAACTTAATAATAAGATTCAAAGATGATAG
AAGAATATACAAACCACTTATTGGAACTATCAGGGTATCAAGCTCTGGGACCTGCTCAAGAAACTCGAGT
GCTAATGTACCCTGACCACTGTAGAAAAAAATAAGAAAACATAAGTACAAGTAATATTTATAACAACTAA
GAAATGTGAGTTCTTGAATATTAACACAAATGCACCTATTAGCATGAACAAGTGATCTCAATAGAAAATG
GAAATGTAATTGCCATAGAACTTTATAATTGACTGGTAGCTTCTTGACAAATCCACTAGTTCAAAGATTT
AATCACTTAAATTCAAATAAATTATACCTTATTATACGAGCATCATTAGAAGAAGGAATGAGAACAGCTC
CCCTCTCGAGTGATACCTTTATTGCTGTCTCTTCTCTAGATTTCATGTTTGACTCACTAAATATAACCTG
ACCCCCGTAACGTTTGACATTCTCTACTTTGCATTTTGGTGCGTTCTTTGGCACCACGACATATGCGGGG
ATTCCACGAATTTGTGCAGCCAACGATAATGCCGCAGCATGGTTGCCACTAAATATAAAAATATTAAAAG
TGAATGAAGTTCCAAAATCATAACATTAAGCTTCTTAGTGAAACGAAAAAAGAACTCAAACAGATGCTTA
AATGAGTGATTTTGGATTAGAGATGGTGTTTTTCCATTTAGATCTGCACAATTTGCCCAAAATCAATAAT
GCTTACGTACGTTATTGTCTTGCTATCTCCAAAAGATTGATGGTTAGTTTGCAAAAACGTGTTAATTTAA
AAAGTCAATACGATAAGCATGTCAAACTATATATATATATCTTCAGGTTTTAAATGTTAACAAAATGGTT
ATAAAAGTGTTCTAATTATGTCACTCGTATCATTTTTGTCCCAATTGGATAACTTAACATTTAAGTCGAG
CCATATCTTTTTTTCCATGTGTCAAGAAAAAAATGAAGCATACGACGCCAGGGTCGGATTATTTTAATAA
TAGAAATTATATTTATTAAAAAGAAAAACACTTTAATTACATTAAAACTTAAAAAAAACAAGTTAAAATC
CACGTTATCACTCGACGTTAGCATCAAATAAAAGGAAAAAAAGAAACAAAAAACTACGTAACCACGGTTA
CCTATGAAATGAATGAAAAAAACTCTTAATTTTAATGAAAATGAATGTTGAGTGACATGATTGAAACATT
TTTAAAATTTGAGTGACCTATTGTAACACTTTTAAAACTTAGTTACTATTATACAAAGTTACCCCTATAC
TATTAACTATAAACCCATGTTGATCCTTTAGTATGTCTGACCTCGATTGGAAAAGTACACCTATCTTACC
TGCTGTGTGTTACAACTCCTTTTGCTGCTTCTTCTTCGTCAAGAGACAAGACAGCGTTTAATGCACCTCT
GAATTTAAAAGCTCCACTGATAACCATCAACAAAATACGTTAAAAATTAACATGTTCATTTTACCACCTC
TAGATCGATAAAGATAACGCAGAAGA

Reverse: 3347 bp

ACTGTCAAACGTTACGGGGGTCAGGTTATATTTATGTGAGTCAAACATGAAATCTAGAGAAGAGACAGCA
ATAAAGGTATCACTCGAGAGGGGAGCTGTTCTCATTCCTTCTTCTAATGATGCTCGTATAATAAGGTATA
ATTTATTTGAATTTAAGTGATTAAATCTTTGAACTAGTGGATTTGTCAAGAAGCTACCAGTCAATTATAA
AGTTCTATGGCAATTACATTTCCATTTTCTATTGAGATCACTTGTTCATGCTAATAGGTGCATTTGTGTT
AATATTCAAGAACTCACATTTCTTAGTTGTTATAAATATTACTTGTACTTATGTTTTCTTATTTTTTTCT
ACAGTGGTCAGGGTACATTAGCACTCGAGTTTCTTGAGCAGGTCCCAGAGCTTGATACCCTGATAGTTCC
AATAAGTGGTTTGTATATTCTTCTATCATCTTTGAATCTTATTATTAAGTTGTTTAGTTGGTTTCCTTTT
TCATTTGAAGTAGAAACTTGTAGGTGGTGGTTTAACATCTGGTACGGTGATTGGAGCCAAGGCAAAGAAT
CCTGCCATACGGGTCTTTGCGGCTGAGCCAAAAGGAGCAGACGATGCAGCACAATCCAAAGCATGCGGTA
GACTCATAACCTTGAGTCAAACAAACACAATTGCTGATGGACTTCGAGCTTCCCTTGGAGATCTTACTTG
GTAATATTATCTTTATATCTCTTCACCGCTTTAAAATATACATTAAATGTGGTTTACCTGTTTCACTTAT
ATTCATAGTTATTTTTCCCTACATGATCAAATAATCAAAACCAGTCTTACTATACTCAAAACATTAGAAA
TTGGTTATCAACTACTTGAAAAGGCATTAGTAAAATAATCACTTACGCCTTATTGTTGCATCTGGTTTAT
TGTTCAAATTAGACTCATGTTTAACATTTGAGTGTTACTATTTTTATGTAGGCCCATTGTACGGGATCTC
GTTGATGCCGTTATAACCGTGGAAGATAAGGACATTGTTGAAAGCAACTCGATATTGTTATCAGATACTC
AAGGTTGCAATTGAGCCCAGTGCTGCTATAGGGCTTGCTGCGGTTCTATCCGATAACTTCAAGAAAAATC
CTATTTTGAGTTGTTCTCGAAAACATTGGGATTGTGCTTTCTGGGGGGCAATGCGGACCCTGGGCGTTTC
TGTTTGAATCCTGGAAAAATTGACGACTTGTGAACCTTATTTAAATAACTCTGGAGGGTAATTGTGAAAC

CTCTTTTGTTTGTTACTTGAAACTTGAAAGACCTTTAAGCAATTAAAAGCATGATGTTTTGATTATCATT
ATTATTATTATTATTATTATTGTTGTCCGTTTCCTCTTTGATCTTTTAACCTAGTAGCACAAACATCAAA
TATATGATAGAGACTCACACTAATCATCATACATACGTTACACACTTTAATGGAGTCAATTTTATTCAAG
GCTGAAACTATTCACACACCAAGGCTGCCTATTTGCACACTTAGTGTTTATATACGCCGAGTGTAACTCT
ATACGTAGCGAATAGGGTTACACTTATACGACGCGTATAGAGCTATACGCAGCGTATAAGTTGCATAGAA
TTCAGAGCCTGCCCAGCAATCATGGGCACAGACAACCAGAGTTTATATACGCTGCGTATAGAGCTATACA
CAACGTATATAACCCATCTAGAATTTTGTTTGGTTATTTTGATGAATTTGTGGTATAAATTCTCACCCGG
TTCCAACATAAATCGCTTAAAATATATAAACGTTAATCTTATACCATTAATATTATATACCGTTAATTAT
TTTTTTAAAAACGTATATACGGTTCGAATATATTATACTGTTATCGTTTAAATTAAATTATTATACCACA
TCTAGTGCGGTTCAAATGCTATAAACGTTTCCATATCATACCGTTAATTATTTTTAAATACTTGAGTACG
GTTCGAATATATTATACCGTTTCCATACTATACCGTTTAAATTAAAACCATTAAATATTTTGAAATATTC
GGGTACGGTTCGAATATATTATACTATTTCAATACTATTCCGTTTAAATTAAATTATTACACCAAATCTT
GGTGCAATTCGAATATTATACCGTTTAAGATTTTCATACCGTTAGTTTTATTAAATTACTCAGGTACAAT
TCAAATATATTACACCGTGTTTCCATACTATAGCGTTTAAATTAAAATATTATAGCAGTTTTAAATTTTT
ATAACATTAATTATTTTGAAATACCCGGGTACGGTTTGAATATATTACACTGTTTCCGTACAATACCGTT
TAAATTAAAATACTATATTGTTTCAATACGACTGAATAGAGCTTCCTGTTTCCCCGTGTTGGCTTCTCTT
ATGCTCACTCCACTCATCGTATATAAAATTTCCGCCTTAAACAAAGTACACCCTGGATCGACAGGCATCT
AGCTGCGCTACAGCTCTATGCTCTACAGCTCCATCCCAAATGGATATCCGAATTCATAAAAAAAAAATCA
AAAAATTTGTTTCGTGCATATATTAAAGCTAAACCAATGTTAGATTTTTCAAATTTTCAAAGTTAATAAC
TAACTTTTATAAATCCACATCCGAATCCGACTATTTACTATTTTTTACTATATTTCGAAATAAATATAGT
GCTCATATACTATATATATTAAAAAAATATTAGTTTTCGGATATTGGATAATCAAAATTTTAGAAATCCA
CATCTGAATCTGGAAATTCGGATTATCCAGTTTTCGGACATCCAAATTTTTGGATACGAATTTTTGGATA
TTTCGGATTCAGTTTCGGATTCGGATTTTTTTAATACCCCTACCTAACCTAAAAGTTGTTGAATTCTGTT
TCCCCGCGTTGGCTTCTCTTATGCCCACTCCACTCATTGTATACAAAAATTTTCCCCTAAAACCATGGTT
GTAAAACAAGGTTTCCAAGGCCGAGTACTCCCCGAGTAGTCGCTACAAGGCAGGCTGCCGAGGCGACTTT
CTTTGACTTCGCCTAATTACTCGGAGTCGGTCAAACGCGGTCAACCTCGGCCAGAATTGGATCTGGTAGG
TCAACGCGGGCCTAGTTTGACTTAAATAACAATATAAGATAATTTCTATGCCTATCATATTAAAGAATGA
ATATGATTTTCACATATTTTGTTATAAATATTAGTAAATTTGTGTTATTTGACATATATTTAATCTCCAA
AAAATAATTTCTTTATAATTTAACACGTCCGAGTACTCCTCGAGTACTCTCCGCATAGACCGAGTACTCT
CAACTCCCTGGTCGACCGACTAGGGAGAGCCTAGCGACTTTTGCAACCATGCCTAAA

401E15 – B5 – B6

Forward: 1522 bp

TTGCTTCGGAAATGACCCATCTCGCCACAGTTGAAACAAGAACCTGGTAGGAAACGACCTTGAGCGGGAT
TGTTGCCAGCTTGGTTTGGTGCAGCTGCAGCTGGGCAGATTCTTGTTGTATGGCCTATGAGTCCACAAGT
TTGGCATTTGCGGCACTGTACATTGGCGTGATGATGGAGGCTACATTGGTTGCACAAGGGTGCAGTTCCA
TTGTATGGTTTTCTAGCAGGGGGTTGAGCTGGTTGGTTGGGGACGGCTTGACCATTGTGAGCGACCACGG
CGAAGTTCTGAGAAGCCTTTCGCTTCCTTGACTTCTTAGGAGATTCAGTCTGTTCATCCATGGTCTTTGA
TTCCTCGATTGGGTTCGATTCCCCGGCCGATTTCTTGTCACCTTTTCGAAAAAGTTTATGCTTTCTGATC
TGCGATTCAGTCAAGGTAGCAGATAGCTCAATGGCCTGACGGAGTGTGGTAGGGTTGCTACCAGTGACAA
TGTCTTGTACCGAGTCAGGCAGGCCGTCGATGTACCTTTCGATAGCCTTGTCGAGTGGGGCGACCATAGT
CGGGCAAAGTAGACTCAACTCCTCAAACCTATCAGTATATGCCCTGTGTTCGCCACTATCTTGTTTCAAA
TCATCAAACTTCTTCTCCAACGCTCGTTGTTCATGACGAGGACAAAACTCTCTCATCATAAGTGCTCTCA
GTTCAGCCCATGTTTGTGCTAGAGCAACATCTGCACCACGGTCTCTCATTACCCCGTTCCACCATGTAAG
AGCCCTCTTCTGAAACACACTCGAAGAAAACTCGACCTTGCGATTTTCCGGACACTGCACGTGGCGAAAG
GTATTCTCGATGCTCTCGAACCATTGGAGAAGCCCAGTTGCTCCCTCAGAACCACTAAACTTGAGTGGTT
TAGCCGAGTTAAAGTTCTTGAAATTGCATGTACCATTGTTGTTGTTGTTGGCTTGGTTCCATTGAGCAAA
GAGATTTGGGAATTGAGCAGCCATCTGCTGCGCAATAATTTCTGCCAGCTCGGCAGTTGCTATCTGGTGT
TCGCGTCGAGGAGGCATTCTAAAAGAGGAGAACATGAAAGGAAACAAATAAAATGGTATTGGGTAAGAAT
AGGATGTTACAATCACTGACCATAATCACGGTTGATGGTCACTTGTTTGAATCATGAAGCAAACAAACAA
CACCAAGGCTGAATCAAAGCAAATAGATCCTCATAGTGTATCGCGAAGACATGCTCGCCTATAAGTGGAC
ACTCACCCCAAGAGTTCCCAGGTAAGAGTGACTGGTCCGATTATGTGGATTTGTACGAACACTCTAACCT
TAGACAGAAAACCCAGGGTACAGGCATTCACTCTTCCAGTTCGCACGTGTTCACACTAGTTAGGACCCAA
AACTTTGACGGGGGTTTTGAAAATTCAAAGGGGTTCAAAACCATATAACAGAGGGTTCAAAACCTTACAA
TAGAGGGTTCAAAACCTAGTAATCAATTATCCTAGAACAGATGATTAGTTTT

Reverse: 1161 bp

CACGTGCAGTGTCCGGAAAATCGCAAGGTCGAGTTTTCTTCGAGTGTGTTTCAGAAGAGGGCTCTTACAT
GGTGGAACGGGGTAATGAGAGACCGTGGTGCAGATGTTGCTCTAGCACAAACATGGGCTGAACTGAGAGC
ACTTATGATGAGAGAGTTTTGTCCTCGTCATGAACAACGAGCGTTGGAGAAGAAGTTTGATGATTTGAAA

CAAGATAGTGGCGAACACAGGGCATATACTGATAGGTTTGAGGAGTTGAGTCTACTTTGCCCGACTATGG
TCGCCCCACTCGACAAGGCTATCGAAAGGTACATCGACGGCCTGCCTGACTCGGTACAAGACATTGTCAC
TGGTAGCAACCCTACCACACTCCGTCAGGCCATTGAGCTATCTGCTACCTTGACTGAATCGCAGATCAGA
AAGCATAAACTTTTTCGAAAAGGTGACAAGAAATCGGCCGGGGAATCGAACCCAATCGAGGAATCAAAGA
CCATGGATGAACAGACTGAATCTCCTAAGAAGTCAAGGAAGCGAAAGGCTTCTCAGAACTTCGCCGTGGT
CGCTCACAATGGTCAAGCCGTCCCCAACCAACCAGCTCAACCCCCTGCTAGAAAACCATACAATGGAACT
GCACCCTTGTGCAACCAATGTAGCCTCCATCATCACGCCAATGTACAGTGCCGCAAATGCCAAACTTGTG
GACTCATAGGCCATACAACAAGAATCTGCCCAGCTGCAGCTGCACCAAACCAAGCTGGCAACAATCCCGC
TCAAGGTCGTTTCCTACCAGGTTCTTGTTTCAACTGTGGCGAGATGGGTCATTTCCGAAGCAATTGCCCA
AAGCTTGAGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGA
AATTGTTATCCGCTCACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCT
AATGAGTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCCAGTCGGGAAACCTGTCGTG
CCAGCTGCATTAATGAATCGGCCAACGCGAACCCCTTGCGGCCGCCCGGGCCGTCGACCAATTCTCATGT
TTGACAGCTTATCATCGAATCTCTGCCATTCATCCGCTTAT

115P9– B4 – B5

Forward: 542 bp

GGTGTTTTTCCATTTAGATCTGCACAATTTGCCCAAAATCAATAATGCTTACGTACGTTATTGTCTTGCT
ATCTCCAAAAGATTGATGGTTAGTTTGCAAAAACGTGTTAATTTAAAAAGTCAATACGATAAGCATGTCA
AACTATATATATATATCTTCAGGTTTTAAATGTTAACAAAATGGTTATAAAAGTGTTCTAATTATGTCAC
TCGTATCATTTTTGTCCCAATTGGATAACTTAACATTTAAGTCGAGCCATATCTTTTTTTCCATGTGTCA
AGAAAAAAATGAAGCATACGACGCCAGGGTCGGATTATTTTAATAATAGAAATTATATTTATTAAAAAGA
AAAACACTTTAATTACATTAAAACTTAAAAAAAACAAGTTAAAATCCACGTTATCACTCGACGTTAGCAT
CAAATAAAAGGAAAAAAAGAAACAAAAAACTACGTAACCACGGTTACCTATGAAATGAATGAAAAAAACT
CTTAATTTTAATGAAAATGAATGTTGAGTGACATGATTGAAACATTTTTAAA

175L13 – B1 – B2

Forward: 596 bp

TTTGTTGCCCTTCACCTTTGAAAGCCGTTCTAGCAACAGCAGCATAGAACCTGCTCGATTGTTTTGCCTG
ATCAGATTCGTATATAACTATATATTCAGTTGATGCTAGACTATTGAAACATTTGAACTTACATTCCCAA
CGTCATCACTGTTAATACTCCAGATATATCAGCACCTTCTTGAGACTATCGCAACGGAGACACAAATTAG
CTTATTATACAAAACGTATATAGGGAACAAATGTAGTTTTGCGAGTTTAAAAGGTAACATACCACGAAGT
AAGCAAGGTAGCTCACAGCAAGAGTCAATGTAATTTCTATCACAGTATCGTTGAATATAAACCCGAGCCA
CAAGTAAGACACCAGACCAAACGCGATACCCATTCCTACACTGGACCAAGGAAACTATTTTATCAATCAA
ATATAACCGTCATATTACAAATCAAAGTCATATTACAAATCAAAGGTAAATACTTACGCTCCAAGAGATA
CTTGCGACAAGAACTTGATGATGGTACCCCAGCTGAACGTCGATCCCGTTACCATTCGGAAAAAAAGTGT
ATACACAACGATTGCCGTCCTGAACGTAAGTACTAA

231M03 – B4 – B5

Forward: 544 bp

TTCTCAAGCTCGAGCTTGGCTCGTTTTTTATTTTATAGTTAGTTTATATATATATATATATATATATATA
TATATAGACATATTTCATATATTCTTATAAGTCTTAAAAGATAAAGTCGGTGCTTTCTCTCACCTAGCAC
CCTTCAAATATGTAAAATCTTTTAATTTTAACCCTTATAATATTGTTTTGTATTTACACCTTTCACTTTT
TGAAAGTTACAATTTTAACCTTTGTGCTATTGCTTTGTTTTACAATTTTTACAACTACTTTTAAAATTTA
CAGTTTTAACCCATGCGTTATTACTTTATTTTACACCCCCTCAATTTTTAAACTTTGTGTTTTTTTTCCT
TTCACTATTACTCCATTTTGACCACCCCACTCAATTTTTAAACTTTGCGTTTTTTATCCTTTCACTATTA
CTCCATTTTTAGTGCCACCACACCGTCAAGTTTTAAACTTTACGTTTTTGCTTCTTGTACTATTACTCCT
TTTTTACACTTGCTCAACTTTGAAAATTTCATATGGTTACTGTATCACATAATC

194A21 – B1 – B27

Forward: 655 bp

CGTTAAACGGATAGAAAGAAGAATTTATGTTAAAAGCGAAAAACCCAATTACTTGGGTAGTCGAATGTGT
ATGCTTAACTCTCATTGAGAGTGTTTATGCCAAACCCAATTACTTGGGTAGTCGAATGTGTATGCTTAAC
TCTCATTGAGAGTGTTTATGCCAAACCCAATTACTTGGGTAGTCGAATGTGTATGCTTAACTCTCATTGA
GAGTGTTTATGCCAAACCCAATTACTTGGGTAGTCGAATGTGTATGCTTAACTCTCATTGAGAGTGTTTA
TGCCGAACCCAATTACTTGGGTAGTCGAATGTGTAAGAAAGAAGAAAGCTCATATATGGATGGAACTAAA
ATGAAGTAATTATGATTCGACAACTGAGATGACTAACTTTAAAATCTTGTTGCTGAAGGTAGGTAAAACT

TTAAGTTTTTGTTAAACACATGTAAGGAAGAAAAGGTACGAATGATATGTTTGAAACCGCTAGATGGATT
CAAACATAGAAGGAATGAAGGGTATGGATGTTACATTTAAATCCGTGGGACGGATTTAAAACATAAAAAG
GATGTAATGAATGTCTTTATAAGTAAGCATGAATGGAAGGAAGTACGAATGTGTACCCCAAAGCATCTAT
GATAAGTAAGAAAAGTCAGTTTACC

94F15 – B105 – B111

Forward: 1905 bp

TTTATTAGTCGCCATGTGTACCGATTCAAGAAGCAATAGATGACCCTTTGCAGGGTTTATACTACTTAAT
GCAATAATTCCATAACAAGCATATCATTATCTTTTAATCCCATTTCTTTTCTGACTGATTTTCTTAACAA
AAGTTTCTTTTCAAACATTTTTTCGGTGGTGAATGCAGGGGTATTTAGGGAACAAGAAATTCCAGCTACG
AAAGCTAGTTCCTCATTGACAAAAAGTGGAACTACAGAGGGCATGTGTTTAAATTTAATATTTTCTTCTT
CACACCAATCTAGCTATTGCTTAGATTGTGATTCGGAGATAAATGCTAATGTTTTAACGCGATTCAGAAC
GAGCTTAGGTCGGTCAAAATATTCCCGTCGGCTTTCCATGATCCACCAAACAAGTTGTCGTGTTTAAGGC
AATATGAAAAAGAGGTCTTACCAATCCTTTAACAATTATTAATACTTAAAACTGCATTATTTAAGGCAAT
ATGAAAAAGAGGTCTAACCAATCCATGATGCACAAACAGCTGAACCTGCTATAACAAGATCTGCTTGCGA
GGCCGCTTTGAAACTGAGTTTATCTTTATCTTCAAGAACTTTAATTCCTTTTCTAGAAAGTTCCGGAAGC
AATCCGCCTCTTCTGCCAAGAGCAACTACATAAACTGTGGCTCCACAACTCAAAAGCTCTAAAGCCAACT
CCATCATCGAAAGTGGAGCTCCGGTCATTGAAAGCTCATGATATATCAAAACAAATTTTCTTGACCAAAC
AAGCCGAGCAAATTGACTTCCTCTATCGCATGTTTCGGATCTCTTCGCAGGACTCCATTCCAAAACCTGG
TCTTCAATCGAACCAAACGGACCGACAAGCATCCCGTAAGTCGAATTTGTTTGAGGAATTAATGTTTCCT
ATTAAACTTTAGTTAAATTATACTACACATTCTATAAACCTAACAAAAATCTATATACAATCACGGATAT
AACTTTGATGAAAACTCACCACACACATTCTTTTACCTTCAACATAATAGGAAACTCACCTCACTGGGAT
ACAAATCTACAAACTACATAAGAAATAACACTGTCAAATCAATTAAACAAACACACATCAATTAAACTGA
CCTGGATCTATATCAAATGTTATTAAGAAATTTATTCATTGAACAATAAGTACCATCCTCGAAGGGAAAT
CACATCGAAATAGAAGGCAAACAGACAACAAGCGGCACCGCTAACCGTTTTTGGATAGCAAAACTGAGTC
TTTTATAATAATCTTGATTTTTAATAAAATAATAAGGGCTTTTGTATGAAAAATTACACTCTTTTGCTTA
ATATGTTGAAAGAAGCATGTCTAAGGTAATAGAATCAAACATTCAACACCCCAAATCCAAGGTACGGTGA
TATTTTTGGATTCAAATTACAAATTATTCATCAAACCAAAGCCCAATTTCCATGAAAAACAAAAGGTCTA
AAACTTTAATCTAAAGCCAACAAGATTCAAGCAGCAAGAAGCCCAATGAACCTAACAAATTAAAAAAAAA
CTCCAAACATTAAAGAAAAATAAAAAATCAAGAAACCAAACCTCCAAATAGCGAGTTTCTTGTTGAAGAA
TCAATCGTATTCATTCACCCCTGGAATTCTATGCCATCGATCTTATTCGGTTTATGTTGATTGGACTCTT
TGAGTACCTATAAACCCCATAATTTTGCCCTAATTCAATTTGCGTATTTAATTATATTCAATATGAAAAC
TGAATCGTGGGTTGTTCAATCGATCCATTGATTATATCATCCAGAAAACCGAAGAAATAATTGAAAACTC
ATTAAATGGGGGCAGATTTGGGAGAGGGGATGAAAAAGCAGGGAAACAAGACAACGTTTGCCGAATTGAA
GTCCAATTTGCCTGA

Reverse: 1555 bp

TAGCCCTTATTATTTTATTAAAAATCAAGATTATTATAAAAGACTCAGTTTTGCTATCCAAAAACGGT*TA*
*GCGGTGCCGCTTGTTGTCTGTTTGCCTTCTATTTCGATGTGATTTCCCTTCGAGGATGGTACTTAT*TGTT
CAATGAATAAATTTCTTAATAACATTTGATATAGATCCAGGTCAGTTTAATTGATGTGTGTTTGTTTAAT
TGATTTGACAGTGTTATTTCTTATGTAGTTTGTAGATTTGTATCCCAGTGAGGTGAGTTTCCTATTATGT
TGAAGGTAAAAGAATGTGTGTGGTGAGTTTTCATCAAAGTTATATCCGTGATTGTATATAGATTTTTGTT
AGGTTTATAGAATGTGTAGTATAATTTAACTAAAGTTTAATAGGAAACATTAATTCCTCAAACAAATTCG
ACTTACGGGATGCTTGTCGGTCCGTTTGGTTCGATTGAAGACCAGGTTTTGGAATGGAGTCCTGCGAAGA
GATCCGAAACATGCGATAGAGGAAGTCAATTTGCTCGGCTTGTTTGGTCAAGAAAATTTGTTTTGATATA
TCATGAGCTTTCAATGACCGGAGCTCCACTTTCGATGATGGAGTTGGCTTTAGAGCTTTTGAGTTGTGGA
GCCACAGTTTATGTAGTTGCTCTTGGCAGAAGAGGCGGATTGCTTCCGGAACTTTCTAGAAAAGGAATTA
AAGTTCTTGAAGATAAAGATAAACTCAGTTTCAAAGCGGCCTCGCAAGCAGATCTTGTTATAGCAGGTTC
AGCTGTTTGTGCATCATGGATTGGTTAGACCTCTTTTTCATATTGCCTTAAATAATGCAGTTTTAAGTAT
TAATAATTGTTAAAGGATTGGTAAGACCTCTTTTTCATATTGCCTTAAACACGACAACTTGTTTGGTGGA
TCATGGAAAGCCGACGGGAATATTTTGACCGACCTAAGCTCGTTCTGAATCGCGTTAAAACATTAGCATT
TATCTCCGAATCACAATCTAAGCAATAGCTAGATTGGTGTGAAGAAGAAAATATTAAATTTAAACACATG
CCCTCTGTAGTTCCACTTTTTGTCAATGAGGAACTAGCTTTCGTAGCTGGAATTTCTTGTTCCCTAAATA
CCCCTGCATTCACCACCGAAAAAATGTTTGAAAAGAAACTTTTGTTAAGAAAATCAGTCAGAAAAGAAAT
GGGATTAAAAGATAATGATATGCTTGTTATGGAATTATTGCATTAAGTAGTATAAACCCTGCAAAGGGTC
ATCTATTGCTTCTTGAATCGGTACACATGGCGACTAATAAAATGCCAGGTGGCACGGTTGATAACGTTGA
AGGTTTTAAAAAAAAAAAAGTTGAGGGGAAGTGAAGAAAAGAAGGTGAAATATGTTAAGCTTGAGTATTC
TATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCTGTGTGAAATTGTTATCCGCT
CACAATTCCACACAACATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGAGTGAGCTAA
CTCACATTAATTGCG

100L22 – B3 – B10

Forward: 592 bp

TCAATGAACTTGCTAAATGTTAACGAGTTAAACCCTTTTTTATTCCTTAACATCATAAGATACGTGCCCC
AAGTTTCATGTGGCAACGCATCAGCCAACTTTTCAATCAACTCCTCATTGTCTTTTGTGACGCTCAACCT
TTTCATGTTCATTACTAGATTACAATATCTTTCAATAATCTGCTTAGTACTTTCAGTTCTTAAACCTCGA
AATAGATCGAACTCTTTTTTCAGAAGCGATTTTTTGTTCTTAACCATCTCTTGGCTTCCAACAAACTTTG
ATTTTAATGCTTTCCAAATTGAGTATGCACTTCCATTGTGCTGCAACAAAACCAAGATATCTTCTTTAAC
AGCCTGTTGTAGTAAACTGATCATCATTTTCTCGTTTTTGTATTTCTTTTTCTCATCCACACTAAGATCT
TTAATAGCAATCTCATCTTCGTCATCATTCAATGGTCTAACATATTTCGTTTCAACACATTCCCAAGCAT
CCAAGTAATTTGCTTGTACCCAATTCTCAAATCGTTCTTCCCATCCTTTATACTCCTCAATGTTCATGAG
CTTAGGCGGTTTTTGCATGGTGCCCGTTTCAT

Reverse: 515 bp

CAAATTACTTGGATGCTTGGGAATGTGTTGAAACGAAATATGTTAGACCATTGAATGATGACGAAGATGA
GATTGCTATTAAAGATCTTAGTGTGGATGAGAAAAAGAAATACAAAAACGAGAAAATGATGATCAGTTTA
CTACAACAGGCTGTTAAAGAAGATATCTTGGTTTTGTTGCAGCACAATGGAAGTGCATACTCAATTTGGA
AAGCATTAAAATCAAAGTTTGTTGGAAGCCAAGAGATGGTTAAGAACAAAAAATCGCTTCTGAAAAAAGA
GTTCGATCTATTTCGAGGTTTAAGAACTGAAAGTACTAAGCAGATTATTGAAAGATATTGTAATCTAGTA
ATGAACATGAAAAGGTTGAGCGTCACAAAAGACAATGAGGAGTTGATTGAAAAGTTGGCTGATGCGTTGC
CACATGAAACTTGGGGCACGTATCTTATGATGTTAAGGAATAAAAAAGGGTTTAACTCGTTAACATTTAG
CAAGTTCATTGAAAAGCTTGAGTAT

147A3 – B20 – B23

Forward: 1090 bp

GTCCCAAATAGGAGTTATGCCCGATATCGTAATTAGAAGCTTTTAACTAAGTAAATTACGATATCTTGCA
TAAAGCATGTTTAAACTTGGTTTTTGACCCAAAACTCATTACCTACTGTTAAGATATTATTTTTGAGGGA
TTGTTGGAGTTATTGATCAGATTATAAACTGATCATATCATAGAGTTCTTGTATTAATTCGATAAATGAC
GATAATACCCTTTTCACGCATAAAATGAGATTTACAAATGATTTGAATGCCAAACCTTTTCCTACTGATT
TTATACATTAAATAAATTATTTTGAGCATTCAAAACTAATCAAAATCTCAGATTTCCATAAAACCCCTTA
ATGATCGTCAATTAGCGATTTTTACGCTATTTAGGTGCATAGTATGGTTTTAAACCATAATTAACACCTA
AGACTTGTCACCTACCGAATTCTTAAACAAATTTTAATATTATTATAGTAGGTATAAGTCGAGAACTCAG
ACTTCCAAAATGCCCGTAAAAGCATATGTAAAATGACCAAAATGCCCTTTCGGGACATAGTTTGGCCATA
AATGGTAAACCTCACATATGTATGATATCTTACTGTTGTAATGAGATAAAATAAATATTTTTACTGATTA
GAAAGGACCAGAACTTCAGTTTGTTATAAAATCTCTTTTATAAATATTAAAATGACCAAAATGCCCTTAC
GGGACTTAAAATGGTTTTAAATACCTTTTGGGCATATATGTTGACATCCTACTGATGTATTGACATATTC
GAAGCATAATAACATATGGAACTTGTATATGATTCATTTGGTTACCCGTTACGCATTTTACGCGTTCGGA
TCGGTTTATGTAACTAGTTTACGTAAAAATTGCCGAAACGGGTTTAACCTTATCGTTTTTGTCTCAAAAT
CCGGAATGTGTTTAGTTTACCCATATTATACAAGTATCCAAACTTGTTGGGTCTAAATAACGTTCTATTC
CGGTCATCGCTTAATCTAGCATGTCGTACCGCTTTAGATACTTCAAGCTAGCCGGTCTAAGTCTATGACT
TAAATAAGACCATTAGCATTCTGAATTTGGTTATATATTA

Reverse: 1119 bp

CTAAAGCGGTACGACATGCTAGATTAAGCGATGACCGGAATAGAACGTTATTTAGACCCAACAAGTTTGG
ATACTTGTATAATATGGGTAAACTAAACACATTCCGGATTTTGAGACAAAAACGATAAGGTTAAACCCGT
TTCGGCAATTTTTACGTAAACTAGTTACATAAACCGATCCGAACGCGTAAAATGCGTAACGGGTAACCAA
ATGAATCATATACAAGTTCCATATGTTATTATGCTTCGAATATGTCAATACATCAGTAGGATGTCAACAT
ATATGCCCAAAAGGTATTTAAAACCATTTTAAGTCCCGTAAGGGCATTTTGGTCATTTTAATATTTATAA
AAGAGATTTTATAACAAACTGAAGTTCTGGTCCTTTCTAATCAGTAAAAATATTTATTTTATCTCATTAC
AACAGTAAGATATCATACATATGTGAGGTTTACCATTTATGGCCAAACTATGTCCCGAAAGGGCATTTTG
GTCATTTTACATATGCTTTTACGGGCATTTTGGAAGTCTGAGTTCTCGACTTATACCTACTATAATAATA
TTAAAATTTGTTTAAGAATTCGGTAGGTAACAAGTCTTAGGTGTTAATTATGGTTTAAAACCATACTATG
CACCTAAATAGCGTAAAAATCGCTAATTGACGATCATTAAGGGGTTTTATGGAAATCTGAGATTTTGATT
AGTTTTGAATGCTCAAAATAATTTATTTAATGTATAAAATCAGTAGGAAAAGGTTTGGCATTCAAATCAT
TTGTAAATCTCATTTTATGCGTGAAAAGGGTATTATCGTCATTTATCGAATTAATACAAGAACTCTATGA
TATGATCAGTTTATAATCTGATCAATAACTCCAACAATCCCTAAAAATAATATCTTAACAGTAGGTAATG
AGTTTTGGGTCAAAAACCAAGTTTAAACATGCTTTATGCAAGATATCGTAATTTACTTAGTTAAAAGCTT
CTAATTACGATATCGGGCATAACTCCTATTTGGGACCAAGAACTGATGTCAAATTTTCGGGACAAGCTTG
AGTATTCTATAGTGTCACCTAAATAGCTTGGCGTAATCATGGTCATAGCTGTTTCCGGTGTGAAATTGT

447N6 – B2- B23

Forward: 1042 bp

CCCACATAAACTGCTACTGGCATATGATCCGTTAGCCAAATGTTCTAAATCATCATTTCTGAATTTGGAA
TGAACTAATTTCATGCTTTTAGCTGCCTGTTCCACACTTTGCACTCCTTTCAGTTGCAACTTCTGTTTCT
TTATCTCTGCAACCAACATACTTATCGAAGTTTGTCTTTCATGCCTCGTTGGTACCATCATCATCGACAC
AAAATTTCCCACAGTAGTTTGTGGCAGCTTGTCCACAAAATGTTTACGTACATCGGCCATCATAAACAAA
TAAGATGGTTTAAAGCAACCAGATTTTGTGGTTGCAGCTGCCACCGCTGTTCTGTAAAGTAGAGAGGTTA
ATACTTCGACCCGAGTAGGATTGTTTGTTTCATCGGCAACAACCTTCTTAAGATCACTTAACTTTGAGTT
AGGAAACACGAATTTCTTCATCACACGAGTCGTACTGCTGGCAAGATGTTGGTTCATAACGGGAGCTTCG
GACAGAGGAGAGTTGGTCCTTGGGACATGAATGAAATGAGGGTTCAGAGGAAGTACTTCTTTATGGTGGG
GCGATCCATAGCGTGCTACGGATGCCCAGTGTAACATGAATGATCCTAGAGTGCAACCGTCACCAACGAG
GTGAGACACAGAGATTGCAAGCCCAACTCCACCACATGCGAAGTGGTTGAATTGAACCCCAACAAGATTC
CTATTATGAGGAGAGTTGTAGCACACCATATCATCCGCAAATAGATAGTCAAGATTTCCATATTGTGCAC
TAGTAAGCTGGAATGCATCAAGCCTGCTATCGTTTCGGGCTTCAAGAAACACAACTCCCTCATCGTTACA
GTCAACGTAAGGTGTTGTGCGTGTGGGTAATCTGCCTGCAAAAGTGATGGTATTTTGTTAGGCTTTGTGA
TAAAGATTTCTTCAGCACGCTGGCCTTGTCTTGGGCAGTTAAACTGCAGCTGGCATTTTTTGGGTAAAAG
AGAATAAGAGGCATATATGCTTTCTCCGCGAGTAGATCGATCTCAGAAAGATTATACGTTTG

Reverse: 669 bp

ATCCTTGGCGCATCCGGAGCACGCTATGGATCGCCCCACCATAAAGAAGTACTTCCTCTGAACCCTCATT
TCATTCATGTCCCAAGGACCAACTCTCCTCTGTCCGAAGCTCCCGTTATGAACCAACATCTTGCCAGCAG
TACGACTCGTGTGATGAAGAAATTCGTGTTTCCTAACTCAAAGTTAAGTGATCTTAAGAAGGTTGTTGCC
GATGAAACAAACAATCCTACTCGGGTCGAAGTATTAACCTCTCTACTTTACAGAACAGCGGTGGCAGCTG
CAACCACAAAATCTGGTTGCTTTAAACCATCTTATTTGTTTATGATGGCCGATGTACGTAAACATTTTGT
GGACAAGCTGCCACAAACTACTGTGGGAAATTTTGTGTCGATGATGATGGTACCAACGAGGCATGAAAGA
CAAACTTCGATAAGTATGTTGGTTGCAGAGATAAAGAAACAGAAGTTGCAACTGAAAGGAGTGCAAAGTG
TGGAACAGGCAGCTAAAAGCATGAAATTAGTTCATTCCAAATTCAGAAATGATGATTTAGAACATTTGGC
TAACGGATCATATGCCAGTAGCAGTTTATGTGGGATGGCGTTTGGCAAAGTCGATTTCGGGTGGGGAAAG
CCCATGGCTAAAGCTTGAGTATTCTATAGTGTCACCTAA

450B6 – B71 – B73

Forward: 576

TAGCAGATTGTCAATCGGACAGTTTATTGAAAAGTTGGAGGCACAGGATCTTGAACAGCAGAAAATTGCA
AGAATGAATAGTTCAAGCCATCAACAAGACATCAAACTGTACTATAAAGGAAATGTTCAAACTGCTGAAG
CAAGTCCAAAGATTCAAACCGCATTTAGTGCAGGAAATCAATCTGGAACAGGTAGTCAAAGTTCAGTCAA
CACTAGTGGGTTTTCAAATGTTAATCCTCCAAGTGTTCAAAGTGCGAATGTTGGTAATGGGCATATGATT
CAGTGCAATGTGGCATTGCATCTACAAAATGGACAGAATTTTTCTGAAGAAGTTGTAAAAAGTCATATGG
GATTACTGGTTACTGTTTTGGAATCTTATGAAGGGTTGATTGCCGGAAAAATTGGTAATCCGATGCTAAC
GAAGGAAGATTATGATCAAATAGATGCAGAGGAATTGGAGCTCATGGACATCAAATGGTGTCTAGCTAGT
GTCCTGAGAAAAGCTGAAAAGTTTAAAATGATAACAGGAAGAAATGACTTTTTGGATGCTCATCTTTCAG
CTTTAGGTTTTGATAA

480G4 – B2 – B4

Forward: 369 bp

CAAAACCGGTTCACATGGTGTGGATAATATTTGCAGTTTCAAAAATCAAAAGGATATTATATGCAAGCTA
ATCGATAAGGATTTGGTTGCTTTGTGGGCTTCGTTTAATCGGAACATCTCCTGTTGGTGTCTTGCTTGTG
TGTCCGTACATTTTTTTTTTATTTTTGTCATTAAATATATTCCATCTTGACGTTAGAAAAAATGATTAAT
TGTGTTACTAACTTCAGCATAGAAATAACTCTGGCTGGATAATATAAATCTAAAAATGTAAGGAATAAAT
AAAACCACTTAAAATACAGTATTATAGATATTTCTTTCAATTTTAGGTTTTCATTTATTTATTTTTCTCT
ATAGATGTCTGATTGAAGA

8.3 Automated contigs assembly

Sequences of all BAC-ends were subjected to an automated contig assembly using VectorNTI software (Invitrogene) for sequence analysis and data management. BAC-end sequences of 115P09 (from RHA325) and 216F17, 384H04 (from HA383) were determined to be overlapping.

```
                1          10          20          30          40                  56
» 216F17    (1) ATGCTTAAATGAGTGATTTTGGATTAGAGATGGTGTTTTTCCATTTAGATCTGCACA
» 384H4     (1)                                   GGTGTTTTTCCATTTAGATCTGCACA
» 115P9     (1)                                   GGTGTTTTTCCATTTAGATCTGCACA
Consensus   (1) ATGCTTAAATGAGTGATTTTGGATTAGAGATGGTGTTTTTCCATTTAGATCTGCACA

               50          60          70          80          90                 105
» 216F17   (50) TCTGCACAATTTGCCCAAAATCAATAATGCTTACGTACGTTATTGTCTTGCTATCTC
» 384H4    (19) TCTGCACAATTTGCCCAAAATCAATAATGCTTACGTACGTTATTGTCTTGCTATCTC
» 115P9    (19) TCTGCACAATTTGCCCAAAATCAATAATGCTTACGTACGTTATTGTCTTGCTATCTC
Consensus  (50) TCTGCACAATTTGCCCAAAATCAATAATGCTTACGTACGTTATTGTCTTGCTATCTC

              107            120         130         140         150                162
» 216F17  (107) CAAAAGATTGATGGTTAGTTTGCAAAAACGTGTTAATTTAAAAAGTCAATACGATAA
» 384H4    (76) CAAAAGATTGATGGTTAGTTTGCAAAAACGTGTTAATTTAAAAAGTCAATACGATAA
» 115P9    (76) CAAAAGATTGATGGTTAGTTTGCAAAAACGTGTTAATTTAAAAAGTCAATACGATAA
Consensus (107) CAAAAGATTGATGGTTAGTTTGCAAAAACGTGTTAATTTAAAAAGTCAATACGATAA

              163      170         180         190         200                218
» 216F17  (163) AGCATGTCAAACTATATATATATATCTTCAGGTTTTAAATGTTAACAAAATGGTTAT
» 384H4   (132) AGCATGTCAAACTATATATATATATCTTCAGGTTTTAAATGTTAACAAAATGGTTAT
» 115P9   (132) AGCATGTCAAACTATATATATATATCTTCAGGTTTTAAATGTTAACAAAATGGTTAT
Consensus (163) AGCATGTCAAACTATATATATATATCTTCAGGTTTTAAATGTTAACAAAATGGTTAT

              217            230         240         250         260                272
» 216F17  (217) TATAAAAGTGTTCTAATTATGTCACTCGTATCATTTTTGTCCCAATTGGATAACTTA
» 384H4   (186) TATAAAAGTGTTCTAATTATGTCACTCGTATCATTTTTGTCCCAATTGGATAACTTA
» 115P9   (186) TATAAAAGTGTTCTAATTATGTCACTCGTATCATTTTTGTCCCAATTGGATAACTTA
Consensus (217) TATAAAAGTGTTCTAATTATGTCACTCGTATCATTTTTGTCCCAATTGGATAACTTA

              273      280         290         300         310                328
» 216F17  (273) AACATTTAAGTCGAGCCATATCTTTTTTTCCATGTGTCAAGAAAAAAAATGAAGCATA
» 384H4   (242) AACATTTAAGTCGAGCCATATCTTTTTTTCCATGTGTCAAGAAAAAAAATGAAGCATA
» 115P9   (242) AACATTTAAGTCGAGCCATATCTTTTTTTCCATGTGTCAAGAAAAAAAATGAAGCATA
Consensus (273) AACATTTAAGTCGAGCCATATCTTTTTTTCCATGTGTCAAGAAAAAAAATGAAGCATA

              328            340         350         360         370                383
» 216F17  (328) TACGACGCCAGGGTCGGATTATTTTAATAATAGAAATTATATTTATTAAAAAGAAAA
» 384H4   (297) TACGACGCCAGGGTCGGATTATTTTAATAATAGAAATTATATTTATTAAAAAGAAAA
» 115P9   (297) TACGACGCCAGGGTCGGATTATTTTAATAATAGAAATTATATTTATTAAAAAGAAAA
Consensus (328) TACGACGCCAGGGTCGGATTATTTTAATAATAGAAATTATATTTATTAAAAAGAAAA

              385    390         400         410         420         430        440
» 216F17  (385) ACACTTTAATTACATTAAAACTTAAAAAAAAACAAGTTAAAATCCACGTTATCACTCG
» 384H4   (354) ACACTTTAATTACATTAAAACTTAAAAAAAAACAAGTTAAAATCCACGTTATCACTCG
» 115P9   (354) ACACTTTAATTACATTAAAACTTAAAAAAAAACAAGTTAAAATCCACGTTATCACTCG
Consensus (385) ACACTTTAATTACATTAAAACTTAAAAAAAAACAAGTTAAAATCCACGTTATCACTCG

              441         450         460         470         480                496
» 216F17  (441) GACGTTAGCATCAAATAAAAGGAAAAAAAGAAACAAAAAACTACGTAACCACGGTTA
» 384H4   (410) GACGTTAGCATCAAATAAAAGGAAAAAAAGAAACAAAAAACTACGTAACCACGGTTA
» 115P9   (410) GACGTTAGCATCAAATAAAAGGAAAAAAAGAAACAAAAAACTACGTAACCACGGTTA
Consensus (441) GACGTTAGCATCAAATAAAAGGAAAAAAAGAAACAAAAAACTACGTAACCACGGTTA
```

```
                496               510          520          530          540          551
» 216F17 (496) TACCTATGAAATGAATGAAAAAAACTCTTAATTTTAATGAAAATGAATGTTGAGTGA
  » 384H4 (465) TACCTATGAAATGAATGAAAAAAACTCTTAATTTTAATGAAAATGAATGTTGAGTGA
  » 115P9 (465) TACCTATGAAATGAATGAAAAAAACTCTTAATTTTAATGAAAATGAATGTTGAGTGA
Consensus (496) TACCTATGAAATGAATGAAAAAAACTCTTAATTTTAATGAAAATGAATGTTGAGTGA

                550          560          570          580          590          605
» 216F17 (550) TGACATGATTGAAACATTTTTAAAATTTGAGTGACCTATTGTAACACTTTTAAAACT
  » 384H4 (519) TGACATGATTGAAACATTTTTAAA
  » 115P9 (519) TGACATGATTGAAACATTTTTAAA
Consensus (550) TGACATGATTGAAACATTTTTAAAATTTGAGTGACCTATTGTAACACTTTTAAAACT
```

8.4 Supplementary tables

Tab. 9 Primers developed for sequencing the BAC-ends. Forward (F) and reverse (R) primer were used to sequence the BAC-ends in both directions.

67N4-B F1	ACACAAGTTTAACATTGCTGC	67N4-B R1n	GGTAATGCTGCATGATTCAATG
67N4-B F0	CCATTCATTTGTCTTGATCTC	67N4-B R2n	AAGAAAGAGATAAAGAGGAACG
67N4-B F2	CTTGTTCTTACCCATCTAGCTG	67N4-B R3	TTTATATAGTGCTTCACCTGT
216F17-B10-F1	GAAATTATATTTATTAAAAAGAAA	216F17-B10-R1	TAGTTTGACATGCTTATCGTA
216F17-B10-F2	TAAAAGCTCCACTGATAACC	216F17-B10-R2	GTTACAATAGGTCACTCAAAT
216F17-B10-F3	ATTAGTGGATTACCTTAGGTTG	216F17-B10-R3	CTTAGAATGCCTGTATATTTG
94F15 B F1	TCCAGCTACGAAAGCTAGTTC	94F15 B R1	ACTTGTTTGGTGGATCATG
94F15 B F2	TAAACTGTGGCTCCACAA	94F15 B R2	AATGGAGTCCTGCGAAGA
94F15-B-F3	CAAACAGACAACAAGCGG	94F15-B-R3	GATTCTATTACCTTAGAC
139A17-B1 F1	GTATAGAGTAGCAGACTCTCC	139A17-B1 R1	GTTTGCCGGAGAGAATTTCC
139A17-B1 F2	GAAACCTAAAATTTCACGTCCG	139A17 B1-R2	TCACATGTTGTAATATACATACC
401E15-B5-F1	TTCCTCGATTGGGTTCGATTCC	401E15-B5-R1	ACCAATGTAGCCTCCATCATC
401E15-B5-F2	TCTCATCATAAGTGCTCTCAG	401E15-B5-R2	GGCGAACACAGGGCATATACTG
401E15-B5 F3	CCAGTTGCTCCCTCAGAACCAC	401E15-B5 R3	TAGTGGTTCTGAGGGAGCAACTG
115P9 B4-F1	ACGACGCCAGGGTCGGATTATTTTA	115P9 B4-R1	GTTGAGTGACATGATTGAAACAT
126N19-B8-F1	GGACAAATATACAGAATC	126N19-B8-R1	GTGTCCTTGGGTCTTCCTCG
126N19-B8-F2	CTAGGGAAGTTATTCTTA	126N19-B8-R2	CCAATATGTAAGTGTAGT
126N19-B8-F3	TAGGGTATGATTCAAACA	126N19-B8-R3	TAATTAAGTGGGTGTTTG
126N19-B8-F4	CCCGAAGAGTTTACCGCA	126N19-B8-R4	TATGTGATGGACAATATG
126N19-B8-F5	AGTTATTCCATAACCATC	126N19-B8-R5	ATGCATGCCGTGCATCGC
126N19-B8-F6	TGGTATACTATGTTTACG	126N19-B8-R6	ATTTTTCATGCTTTCCCT
126N19-B8-F7	CATTTCAGATCGAGTAGA	126N19-B8-R7	GCTAAAGACAAGGGGGTG
126N19-B8-F8	TCTTCAAAGCAGACAATG	126N19-B8-R8	GTATATGATTCCTTTTAGC
401E15-B5 F3	CCAGTTGCTCCCTCAGAACCAC	401E15-B5 R3	TAGTGGTTCTGAGGGAGCAACTG
115P9 B4-F1	ACGACGCCAGGGTCGGATTATTTTA	115P9 B4-R1	GTTGAGTGACATGATTGAAACAT
126N19-B8-F1	GGACAAATATACAGAATC	126N19-B8-R1	GTGTCCTTGGGTCTTCCTCG
126N19-B8-F2	CTAGGGAAGTTATTCTTA	126N19-B8-R2	CCAATATGTAAGTGTAGT
126N19-B8-F3	TAGGGTATGATTCAAACA	126N19-B8-R3	TAATTAAGTGGGTGTTTG
126N19-B8-F4	CCCGAAGAGTTTACCGCA	126N19-B8-R4	TATGTGATGGACAATATG
126N19-B8-F5	AGTTATTCCATAACCATC	126N19-B8-R5	ATGCATGCCGTGCATCGC
126N19-B8-F6	TGGTATACTATGTTTACG	126N19-B8-R6	ATTTTTCATGCTTTCCCT
126N19-B8-F7	CATTTCAGATCGAGTAGA	126N19-B8-R7	GCTAAAGACAAGGGGGTG
126N19-B8-F8	TCTTCAAAGCAGACAATG	126N19-B8-R8	GTATATGATTCCTTTTAGC

Tab. 9 Continuation

401E15-B5 F3	CCAGTTGCTCCCTCAGAACCAC	401E15-B5 R3	TAGTGGTTCTGAGGGAGCAACTG
115P9 B4-F1	ACGACGCCAGGGTCGGATTATTTTA	115P9 B4-R1	GTTGAGTGACATGATTGAAACAT
126N19-B8-F1	GGACAAATATACAGAATC	126N19-B8-R1	GTGTCCTTGGGTCTTCCTCG
126N19-B8-F2	CTAGGGAAGTTATTCTTA	126N19-B8-R2	CCAATATGTAAGTGTAGT
126N19-B8-F3	TAGGGTATGATTCAAACA	126N19-B8-R3	TAATTAAGTGGGTGTTTG
126N19-B8-F4	CCCGAAGAGTTTACCGCA	126N19-B8-R4	TATGTGATGGACAATATG
126N19-B8-F5	AGTTATTCCATAACCATC	126N19-B8-R5	ATGCATGCCGTGCATCGC
126N19-B8-F6	TGGTATACTATGTTTACG	126N19-B8-R6	ATTTTTCATGCTTTCCCT
126N19-B8-F7	CATTTCAGATCGAGTAGA	126N19-B8-R7	GCTAAAGACAAGGGGGTG
126N19-B8-F8	TCTTCAAAGCAGACAATG	126N19-B8-R8	GTATATGATTCCTTTTAGC
225D9-B7-F1	TAAGAACATTGCGAACCAAGTC	225D9-B7-R1	TTCTACCTTGCTTTTGGCGTAC
225D9-B7-F2	TTTCTTAGCCTTAATAGGAGC	225D9-B7-R2	TGATTATGTAGCTTATGGTTGG
225D9-B7-F3	ATTTAAGTCAAACTAGGCCCGC	225D9-B7-R3new	TTTAGAAATCCACATCTG
225D9-B7-F4	TCCAATATCCGAAAACTAATA	225D9-B7-R4	CAAAAAATTTGTTTCGTGC
225D9-B7-F5	ATATATTCAAACCGTACCCGGGT	225D9-B7-R5	TTGGTGCAATTCGAATATTATAC
225D9-B-F6	AATGGTATAAGATTAACG	225D9-B7-R6	GGCACAGACAACCAGAGT
225D9-B7-F7	CAATTACCCTCCAGAGTT	225D9-B7-R7	GGTTCTATCCGATAACTT
225D9-B7-F8	TTTTAAAGCGGTGAAGAG	225D9-B7-R8	ATCCTGCCATACGGGTCT
225D9-B7-F9	ATTCAAATAAATTATACC	225D9-B7-R9	GTGCCAAAGAACGCACCA
225D9-B7-F10	ATGTCACTCGTATCATTT	225D9-B7-R10	GTAATTAAAGTGTTTTTC
261F19-B5-F1	AGTACCATTGTCACTTCT	261F19-B5-R1	ATGGGTCAGAACTGAAAC
261F19-B5-F2	TTAAACGACGTAATTAGC	261F19-B5-R2	AAACCGACAACTACCGAA
261F19-B5-F3	TACCCTACTGTTATAACT	261F19-B5-R3	GATATCTTCTATTATGCC
261F19-B5-F4	TTCGCTTAATCATGCGTTTG	261F19-B5-R4	AGTATTTACCTTTGCAAGTAG
261F19-B5-F5	GTAATTCTTTTACAAGAT	261F19-B5-R5	ACAGTTTACACTGGTAAA
261F19-B5-F6	GATTTCTCGTTTTGTGTG	261F19-B5-R6	TTGAAGGATGCTTTTCAA
261F19-B5-F7	CATCAGTTTTAGAAATCC	261F19-B5-R7	CAAGTTCATTAACTTGAAAG
261D17-B2-F1n	AACATACGAGCCGGAAGC	261D17-B2-R1n	CGCGTTGGCCGATTCATT
261D17-B12-F1n	GTGCAGCTTGGATGCTTG	261D17-B12-R1n	GCAAGAGGCTTTGGTTCT
261D17-B12-F2n	CACTTTCCAACGTAAGTA	261D17-B12-R2n	TTAGGTTCCGCTCTTTGG
261D17-B12-F3n	TGCGCTGAAGACTCTTGT	261D17-B12-R3n	GGGCTGTGTGCTAATGCA
447N6 B23 F1	AGATCACTTAACTTTGAG	447N6 B23 R1	GCACTCTAGGATCATTCA
100L22-B3-F1	AATAGCAATCTCATCTTC	100L22-B3-R1	ATGAAACGGGCACCATGC
147A3-B20-F1	CCATAATTAACACCTAAG	147A3-B20-R1	ACCATTTATGGCCACACT
147A3-B20-F2	TACAAGTATCCAAACTTG	147A3-B20-R2	TAACCAAATTCAGAATGC

Tab. 10 Manufacturer and sources of supplies of equipments and chemicals

Equipments and supplies	Manufacturer and sources of supply
Centrifugation for Falcon tubes (3K30)	Sigma
for Eppendorf tubes (5415R)	Eppendorf
Electrophoresis chamber (Mini sub cell GT, Powerpac300, 1000, 3000)	Bio-Rad
Falcon tubes, Tips, Tubes	Sartedt AG & Co
Film: Amersham Hyperfilm™ MP	GE Healthcare
BioMax MS Film	Kodak
Fuji Medical XRay Super HR-E30	Fuji
500-ml flask	Schott Duran
Hybridization Glass (300ml)	OCHS
Imaging software (Gel Documentation)	Quantity One, Protean®II xi
Incubator shaker (IH 50)	Incutec GmbH
Incubator (37°C)	Heraeus Instruments
PCR thermocycler	Applied Biosystems

Petri dish	Roth
pH meter	WTW GmbH
Photo-spectrometer (Ultrospec 3100 Pro)	Amersham Biosciences
Photo chemical (developer – fixer solutions)	AGFA
Pulse field gel electrophoresis CHEF III	Bio-Rad
Rotisserie oven radioactive (OV3)	Biometra
non-radioactive (GFL7601)	Fritz Gössner
Sequencer (CEQ™ 2000 XL DNA Analysis System)	Beckman Coulter GmbH
Sequences alignment	BioEdit Software
Speed vac (Con-Jet)	Fröbel Labortechnik GmbH
UV cross-linking treatment	VILBER Lourmat
Water bath radioactive (GFL1083)	Fritz Gössner
Non-radioactive (M3)	Lauda
Acetic acid	Roth
Agar	Roth
Agarose gel	Roth
Ammonium acetate	Roth
Antibiotics	Roth
Boric acid	Roth
Bovine albumin serum	Roth
Bromophenol blue	Millipore
Calcium Chloride	Roth
Calf intestinal alcaline phosphatase (CIAP+Buffer)	Fermentas
Chloridric Acid	Roth
dGTP, dTTP	peqLab Biotechnologie
Dimethyl-Formamide	Roth
ECL™ kit	Amersham
Escherichia coli (DHα5)	Invitrogen
Enzyme endonucleases + buffers	Fermentas
Ethanol	Roth
Ethidium bromide	Roth
Ethylenediaminetetra-Acetate (EDTA)	Roth
Phenol-Chloroform-Isoamyl	Sigma
GFX™ Micro Plasmid Prep Kit	Amersham Biosciences
Illustra™ Plasmid Prep Mini Spin Kit	GE Healthcare
Glucose	Roth
Glycerol	Roth
Hybond N+ membrane	Amersham Bioscience
Amersham Hybond™-N$^+$	GE Healthcare
HEPES-NaOH	Roth
Isopropanol	Roth
Isopropyl-ß-D-thiogalactopyranoside (IPTG)	Roth
Klenow DNA polymerase	Fermentas
Lamda-Phage marker	Promega + New England Biolab
6-Load Dye Blue/Orange	Promega
Magnesium Chloride	Roth
Marker MidRange II PFG	New England Biolab
2-Mercaptoethanol	Roth
*Not*I + buffer	New England Biolab
Overgo primer pairs	Invitrogen + MWG- Biotech AG
Phenol chloroform	Roth
Potassium acetate	Roth
Potassium chloride	Roth
Primers for sequencing	Invitrogen + MWG- Biotech AG
RNase	Roth
Sodium acetate	Roth
Sodium chloride	Roth
Sodium dodecyl sulphate (SDS)	Roth
Sodium hydroxide	Roth
Sodium phosphate	Roth

SP6 primer promoter	Invitrogen
Standard saline citrate (SSC)	Roth
T4 ligase + Buffer	Fermentas
Tris-hydroxymethyl-amino-methane	Roth
Trypton	Roth
Urea	Roth
Wathman 3 MM paper	Whatman International Ltd
Xylene cyanol	Roth
Yeast extracts	Roth
[α-^{32}P] dATP, [α-^{32}P] dCTP	Amersham Bioscience+ Hartman
1kb Ladder marker	Promega + New England Biolab
5-bromo-4-chloro-3indol-ß-D-galactopyranoside (XGal)	Roth

List of figures

List of tables

Abbreviations

A	Adenine
AFLP	Amplified Fragment Length Polymorphism
Amp	Ampicillin
BAC	Bacterial artificial chromosome
*Bam*HI	*Bacillus amyloliquefaciens*
BC	Back crosses
bp	Base pair
BSA	Bovine serum albumin
BSA	Bulked Segregant Analysis
C	Cytosine
CIAP	Calf intestinal alkaline phosphatase
cM	centimorgan
CMS	Cytoplasmic male sterility
Contig	Contiguous segment of DNA
dATP	Desoxyadenosin-5'-triphosphat
dCTP	Desoxycytidin-5'-triphosphat
ddH_2O	Double distilled water
dGTP	Desoxyguanine-5'-triphosphat
DH	Double haploid
DNA	Desoxyribonucleic acid
dNTP	Desoxynucleotide
dTTP	Desoxythymine-5'-triphosphat
E. coli	*Escherichia coli*
ECL	Enhanced chemiluminescence
*Eco*RI	*Escherichia coli* RY13
EDTA	Ethylendiamintetra-acetat
et al.	e*t alii*, and others
EtBr	Ethidium bromide
EtOH	Ethanol
F_2	second filial (generation)
G	Guanine
GTE	Glucose-Tris-EDTA
h	hour
HCL	Hydrochloric acid
*Hind*III	*Haemophilus influenzae* Rd
HOAc	Acetic acid
IPTG	Isopropyl-ß-D-thiogalactopyranoside
K	Potassium
kb	kilobase pair
KCl	Potassium chloride
kDa	Kilo Dalton
KOAc	Potassium acetate
*Kpn*I	*Klebsiella pneumoniae*
L	liter
LB	Luria Bertani
LOD	Logarithm of odds
mMeter	
M	Molar
mA	Milliampere (10^{-3}A)
mg	Milligram (10^{-3}A)
Mg	Magnesium
$MgCl_2$	Magnesium Chloride
min	Minute

ml	Milliliter (10^{-3}l)
mm	Millimeter (10^{-3}m)
mM	Millimolar (10^{-3}M)
mmol	Millimol (10^{-3}mol)
mtDNA	Mitochondrial DNA
NaCl	Sodium Chloride
NaOAc	Sodium acetate
NaOH	Sodium hydroxide
ng	Nanogramm (10^{-9}g)
NH_4Oac	Ammonium acetate
NILs	Near isogenic lines
nm	Nanometer (10^{-9}m)
NMS	Nuclear male sterility
nt	Nucleotide
OD	Optical density
OLB	Oligo labeling buffer
ORF	Open reading frame
PCR	Polymerase chain reaction
Pmol	Pico mol (10^{-12}mol)
*Pst*I	*Providencia stuartii*
pUC	Plasmid of the University of California
RAPD	Random amplified polymorphic DNA
Rf	Restorer of fertility
RFLP	Restriction fragment length polymorphism
RILs	Recombinant inbreed lines
RNase	Ribonuclease
Rpm	Rounds per minute
RT	Room temperature
s	Seconds
SCAR	Sequence characterized amplified region
SDS	Sodium dodecyl sulfate
SNP	Single nucleotide polymorphism
SOB	Super optimal broth
SOC	Super optimal broth with catabolite repression
SSC	Standard saline citrate
SSR	Single sequence repeat
STS	Sequence tagged site
T	Thymine
TAE	Tris acetate EDTA
Taq	*Thermophilus aquaticus*
TE	Tris-EDTA-buffer
Tris	Tris-(hydroxymethyl)-amino-methane
UV	Ultraviolet light
V	Volt
w	weight
XGal	5-bromo-4-chloro-3-indolyl-ß-D-galactoside
YAC	Yeast artificial chromosome
λ	*Lambda-phage*
%(v/v)	Volume percentage
%(w/v)	Weight percentage
°C	Grade Celsius
µg	Microgram (10^{-6}A)
µl	Microliter (10^{-6}l)

Acknowledgement, Danksagung, Remerciement

Many thanks to my doctoral advisor Frau Prof. Dr. Renate Horn for her invaluable advices, suggestions and support during this work. Without her tireless assistance and constructive critics, this work would not have been possible.

Sincere thanks go to Herr Prof. Dr. Wolfgang Friedt for reviewing my work.
I would like to thank sincerely Frau Dr. Sabine Fulda for having always an open ear for my daily problems, for her scientific assistance as well as her helpfulness and friendship. I thank you for the unforgettable nice and pleasant moments spent in your house.
Special thanks go to Gesine Krentz for the time we spent together in our common office, for her tolerance and helpfulness. Thanks a lot for your encouraging words and for being there with your support during the time I needed the most.
I would like to thank Ossama Sajer for the nice time spent together as well as Antje Raddemann I am very fortunate for having the chance to meet friends at our department (the list is too long…).

Ich möchte mich bei dir, Dörte Warber, bedanken für deine Freundschaft, die vielen Hilfestellungen, die die tägliche Laborarbeit so angenehm machten und nicht zuletzt für die Geduld und dein Verständnis, auch in Situationen, wenn es im Labor mal angespannt war. Unvergesslich bleiben natürlich auch die unterhaltsamen Mittagspausen.
Rico Fürstenberg, möchte ich herzlich danken für die Hilfsbereitschaft im Labor und die vielen kleinen Tipps und Tricks beim Umgang mit dem Computer.
Allen Mitgliedern unserer Gruppe, auch denen, die hier nicht gennant wurden, möchte ich für die angenehme Arbeitsatmosphäre danken.

I highly appreciate the critical discussions with Prof. Dr. Martin Hagemann and his valuable recommendations, without forgetting the whole group of Plant physiology for creating a pleasant environment, especially Dirk Hasse and Klaudia Michl.
Especial thanks go to Dr. Barbara Kusterer for her orientations, and valuable hints concerning her previous work. Particularly, I want to thank Dr. Michael-Frank Gumpert for his continual help, for his support and for his great patience during this time. And, of course, last but not least, I would like to thank Dr. Ralf Bastrop and his staff especially Lukas Krebes and Jens Frankowski for their efforts in making in time all the sequences needed during this work.

This research was financially supported by grants from the Deutsche Forschungs-gemeinschaft (DFG).

Je tiens tout particulièrement à remercier et à témoigner toute ma reconnaissance à mes parents, Zakaria et F. Zohra, à mon frère Samy, à mes sœurs Lamia, D. Wassila et Soumia et sans oublier ma tante Rebiha pour leur support, leur encouragement, et leur confiance qu'ils m'ont accordée et qui n'ont jamais cessez de m'accompagner tout le long de mon travail. Je voudrais juste leur dire merci beaucoup car sans leur support et leur confiance je n'aurai eu le courage de persévérer. Je tiens aussi à dire à mes neveux, Seif, Anis, Akram et à ma nièce Aya qu'il ne faut jamais abandonner quand il s'agit de réaliser un rêve ou d'atteindre un but…

Enfin, j'adresse mes plus sincères remerciements à tous mes proches et à tous mes amis, sans exception aucune, qui m'ont toujours soutenue et encouragée durant cette période.

Erklärung

Ich versichere hiermit an Eides statt, dass ich die vorliegende Arbeit selbständig angefertigt und ohne fremde Hilfe verfasst habe, keine außer den von mir angegebenen Hilfsmitteln und Quellen dazu verwendet habe und die den benutzten Werken inhaltlich und wörtlich entnommenen Stellen als solche kenntlich gemacht habe.

Sonia Hamrit Rostock, den 3. Juni 2009

Publications

Horn R**, Hamrit S**, (**2009**) Gene cloning and characterization. In: Genetics, Genomics and Breeding in Sunflower, (eds) Hu J, Seiler G (in press)

Hamrit S, Kusterer B, Friedt W, Horn R (**2008**) Verification of positive BAC clones near the *Rf1*gene restoring pollen fertility in the presence of the PET1-cytoplasm in sunflower (*Helianthus annuus* L.) and direct isolation of BAC ends. 17th Int. Sunflower Conf., Cordoba, Spain pp 623-628

Hamrit S, Kusterer B, Friedt W, Horn R (**2008**) Verifizierung positiver BAC-Klone am Restorerlocus *Rf1* der Sonnenblume (*Helianthus annuus* L.) über BAC-Fingerprinting und BAC-End-Klonierung. Gesellschaft für Pflanzenzüchtung (GPZ) e.V. 70, 243-244

Stich B, Melchinger AE, Piepho HP, **Hamrit S**, Schipprack W, Maurer HP, Reif JC (**2007**) Potential causes of Linkage Disequilibrium in European maize Breeding Program Investigated with computer simulation. Theor Appl Genet 115: 529-536

Reif, JC, **Hamrit S**, Melchinger AE (**2006**) Genetic diversity trends in central European heterotic groups. Acta Agronomica Hungarica. 54: 315-320

Krentz G, **Hamrit S**, Horn R (**2006**) Charakterisierung der Genstruktur von *SLY1* in der Sonnenblume (*Helianthus annuus* L.). Gesellschaft für Pflanzenzüchtung (GPZ) e.V. 70, 247-248

Reif JC, **Hamrit S**, Heckenberger M, Schipprack W, Maurer HP, Bohn M, Melchinger AE (**2005**) Genetic Structure and Diversity of European Flint Maize populations Determined with SSR Analyses of Individuals and Bulks. Theor Appl Genet 111:906-913

Reif JC, **Hamrit S**, Heckenberger M, Schipprack W, Maurer HP, Bohn M, Melchinger AE (**2005**) Trends in genetic diversity among European maize cultivars and their parental components during the past 50 years. Theor Appl Genet 111:838-845

International conferences

Hamrit S, Kusterer B, Friedt W, Horn R (**2008**) Verification of positive BAC clones near the *Rf1*gene restoring pollen fertility in the presence of the PET1-cytoplasm in sunflower (*Helianthus annuus* L.) and direct isolation of BAC ends. 17th Int. Sunflower Conf., Cordoba, Spain pp 623-628. **Poster and oral presentation**

Hamrit S, Schnabel U, Warber D, Kusterer B, Lazarescu E, Özdemir N, Friedt W, Horn R (**2006**) Identification and DNA fingerprinting of positives BAC-clones at the restorer locus *Rf1* in sunflower. Seventh European Conference on Sunflower biotechnology. September 3-6, 2006 Gengenbach, (Germany). **Poster and oral presentation**

Hamrit S, Engelman U, Schnabel U, Warber D, Kurutz S, Kusterer B, Lazarescu E, Özdemir N, Friedt W, Abratti G, León A, Horn R (**2006**) Comparative mapping of restorer genes restoring pollen fertility in the presence of different CMS cytoplasms in the genus *Helianthus*. Seventh European Conference on Sunflower biotechnology. September 3-6, 2006 Gengenbach, (Germany). **Poster**

Horn R, Kusterer B, Lazarescu E, Özdemir N, Friedt W, Warber D, **Hamrit S**, (**2006**). Comparative Analyses of Cytoplasmic Male Sterility and Fertility Restoration In Sunflower. Plant and animal Genomes XIII Conference. January 14-18, 2006 San Diego (USA). **Poster**

Kusterer B, Warber D, **Hamrit S**, Özdemir N, Lazarescu E, Friedt W, Horn R. (**2005**). Molecular Analyses of the restorer genes involved in fertility restoration of PET1, PEF1 and PET2 cytoplasm in sunflower. Plant and animal Genomes XIII Conference. January 15-19, 2005 San Diego (USA). **Poster**

National conferences

Hamrit S, Kusterer B, Friedt W, Horn R (**2008**) Verifizierung positiver BAC-Klone am Restorerlocus *Rf1* der Sonnenblume (*Helianthus annuus* L.) über BAC-Fingerprinting und BAC-End-Klonierung. Gesellschaft für Pflanzenzüchtung (GPZ) e.V. 70, 243-244. **Poster**

Krentz G, **Hamrit S**, Horn R (**2006**) Struktur und Funktion von SLY1 bei der Sonnenblume (*Helianthus annuus* L.). Bericht über die 57 Tagung der Vereinigung der Pflanzenzüchter und Saatgutkkaufleute Österreichs. Raumberg-Gumpenstein, 21-23 November. **Poster**

Engelman U, Schnable U, **Hamrit S,** Warber D, Kurutz S, Kusterer B, Lazarescu E, Özdemir N, Friedt W, Horn R (**2006**) Vergleichende Analyse der Restorergene beim PET1-, PET2- und PEF1-Cytoplasma der Sonnenblume. Bericht über 57 Tagung der Vereinigung der Pflanzenzüchter und Saatgutkkaufleute Österreichs. Raumberg-Gumpenstein, 21-23 November. **Poster**

Hamrit S, Kusterer B, Schnabel U, Warber D, Özdemir N, Lazarescu E, Friedt W, Horn R. (**2005**). Vergleichende Analyse von CMS-Quellen und deren Fertilitätsrestauration bei der Sonnenblume. Gesellschaft für Pflanzenzüchtung (GPZ) e.V. „Genomforschung an Pflanzen: Vom Modell zur Sorte“. 67, 151-165. **Poster**

Hamrit S, Kusterer B, Özdemir N, Lazarescu E, Friedt W, Horn R (**2005**). Development of a contig around the *Rf1* gene restoring fertility of the PET1-cytoplasm in sunflower. Gesellschaft für Pflanzenzüchtung (GPZ) e.V. „Genomforschung an Pflanzen: Vom Modell zur Sorte“. 22-23. Februar 2005. Bonn, Nußallee 17. **Poster**

Curriculum Vitea

Name		Sonia Hamrit
Date of Birth		23.01.1965 in Constantine – Algeria
School background	*71 - 81*	Elementary and Primary school
	81 - 84	Secondary school, Lycée Massinissa
	06/84	Graduation: Baccalaureate certificate high school (Biological science)
Academic background	*09/84 – 09/89*	Bachelor program, University of Sétif, Department of plant production. "Valeur nutritive de six associations fourrageres: Vicia-Avena, Pea-Avena Vicia-Barley, Pea-Barley, Vicia-Triticale, and Pea-Triticale.
	09/89	Graduation: Ingénieur d´Etat Agronome
	09/01 – 07/04	Master program, in Plant breeding and Applied Genetics, University of Hohenheim, Germany. "Genetic diversity within and Among maize populations determined with SSR markers" (Prof. Dr. AE Melchinger).
	07/04	Graduation: MSc.
	03/05 – 05/09	Doctorate candidate in Plant Genetics (Prof. Dr. R. Horn), University of Rostock, Germany
Professional and practical experiences	*91 - 01*	Plant breeder at the Institut Technique des Grandes Cultures, Algeria
	02 – 04	Short consultancies GTZ (Deutsche Gesellschaft für Technische Zusammenarbeit) DNA Marker technologies, automated analysis of SSR on ALF express sequencer, detection and quantification of GMOs using real time PCR.