ORIGINAL PAPER

Mohamed Fouad Bouzidi · Jérôme Franchel
Quanzhou Tao · Keith Stormo · A. Mraz
Paul Nicolas · Saïd Mouzeyar

A sunflower BAC library suitable for PCR screening and physical mapping of targeted genomic regions

Abstract A sunflower BAC library consisting of 147,456 clones with an average size of 118 kb has been constructed and characterized. It represents approximately $5\times$ sunflower haploid genome equivalents. The BAC library has been arranged in pools and superpools of DNA allowing screening with various PCR-based markers. Each of the 32 superpools contains 4,608 clones and corresponds to a 36 matrix pools. Thus, the screening of the entire library could be accomplished in less than 80 PCR reactions including positive and negative controls. As a demonstration of the feasibility of the concept, a set of 24 SSR markers covering about 36 cM in the sunflower SSR map (Tang et al. in Theor Appl Genet 105:1124–1136, 2002) have been used to screen the BAC library. About 125 BAC clones have been identified and then organized in 23 contigs by *Hin*dIII digestion. The contigs are anchored on the SSR map and thus constitutes a first-generation physical map of this region. The utility of this BAC library as a genomic resource for physical mapping and map-based cloning in sunflower is discussed.

Introduction

The sunflower, *Helianthus annuus* L. is one of the major crops for oil source and is widely cultivated throughout the globe. It is a diploid plant with an estimated haploid genome size of about 3,000 Mb and $2n=2x=34$ chromosomes (Arumuganathan and Earle 1991). Many genomic resources have been developed for this plant including genetic maps based either on anonymous RFLP markers (Berry et al. 1995; Gentzbittel et al. 1995, 1999) or more recently on microsatellite markers (Tang et al. 2002). These genetic maps are an invaluable resource that can be used not only to map even complex traits such as QTL but also to anchor physical maps (Yim et al. 2002). In addition, a massive effort is currently made for the release of a vast collection of sunflower ESTs and currently more than 66,000 sunflower ESTs are available in the public databases. These sequences constitute a valuable source for the direct access to some genes of interest or for the development of molecular markers for map-based cloning purposes or marker-assisted selection programs. However, the link between these markers and their physical location on the sunflower genome needs to be developed. With this respect, two sunflower BAC libraries have been described recently. A BAC library made up of 158,208 clones and arrayed on nine nylon filters has been described by Gentzbittel et al. (2002) and a second BAC library containing about 100,000 clones and covering $1.9\times$ haploid sunflower genome has been described by Ozdemir et al. (2004). However, both libraries were not designed for systematic and high throughput anchoring of PCR markers on BAC clones. In this paper we describe the design and the construction of a sunflower BAC library suitable not only for southern hybridization but also for PCR screening. Less than 80 PCR reactions are needed to detect one positive BAC clone containing a given PCR-based marker. Furthermore, the pooling and superpooling methods described in this paper could be used for the construction of BAC libraries even from large genomes. We have used this concept to screen the BAC library, with a set of 25 SSR markers covering about 36 cM in the sunflower SSR map (Tang et al. 2002). About 112 BAC clones have been identified and then organized in 23 contigs by *Hin*dIII digestion. The contigs which correspond roughly to 3.1 Mb are

M. F. Bouzidi · J. Franchel · P. Nicolas · S. Mouzeyar (✉)
UMR 1095 INRA-UBP "Amélioration et Santé des Plantes",
Université Blaise Pascal, 24 avenue des Landais,
63177 Aubière Cedex, France
E-mail: Said.MOUZEYAR@univ-bpclermont.fr
Tel.: +33-4-73407911
Fax: +33-4-73407914

Q. Tao · K. Stormo · A. Mraz
Amplicon Express, 2345 NE Hopkins Court,
Pullman, WA 99163, USA

anchored on the SSR map and thus constitutes a first-generation physical map of this region.

Materials and methods

Sunflower genotype

The sunflower line YDQ was used for the construction of the BAC library. This line was bred by INRA (Clermont-Ferrand, France) and is resistant to several races of the biotrophic parasite *Plasmopara halstedii* (Bouzidi et al. 2002).

BAC library construction

This sunflower BAC library was constructed from high molecular weight (HMW) genomic DNA processed at Amplicon Express, Pullman, Washington using the method of Tao et al. (2002). DNA digestion was done with varying amounts of *Hin*dIII to identify an appropriate partial digestion condition. The vector used was derived from the pBeloBAC11 vector and contains two *FRT* and one *oriV* elements, thus resulting in the pBAC(*FRT-oriV*) vector. Ligations were transformed into DH10B *Escherichia coli* cells (Invitrogen) and plated on LB agar with appropriate chloramphenicol, X-gal and IPTG concentrations. Clones were robotically picked with a Genomic Solution G3 into 384 well plates containing LB freezing media. Plates were incubated for 18 h, replicated and then frozen at $-80°C$. The replicated copy was used as a source plate for printing nylon filters and for inoculating plates for pooling.

Insert size estimation

To estimate insert sizes, 10 μl aliqots of BAC miniprep DNA were digested with 5 U of *Not*I enzyme during 3 h at 37°C. The digestion products were separated by pulsed field gel electrophoresis (CHEF-DRIII system, Bio-rad) in a 1% agarose gel in TBE buffer. Insert sizes were compared to those of the Lambda Ladder PFG Marker (New England Biolabs). Electrophoresis was carried out for 18 h at 14°C with an initial switch time of 5 s, a final switch time of 15 s, in a voltage gradient of 6 V/cm.

Filter design and construction

Bioassay dishes containing LB agar plate media and 12.5 μg/ml chloramphenicol were prepared. Hybond N$^+$ (Amersham, Biosciences) nitrocellulose membrane was applied to the media surface. The G3 robot was used to grid 18,432 clones in duplicate on each filter following a 4×4 pattern that allows the well position and plate number of each positive clone to be identified. The filters were incubated at 37°C for 12–14 h and processed by a slightly modified nylon filter lysis method. The filters were fixed according to the Hybond N$^+$ manual by baking at 80°C for 2 h.

Sunflower pooling and superpooling method

The pooling strategy chosen (Fig. 1) was the standard Matrix Pool and Superpool Strategy by Amplicon Express. The strategy consists of a two-step approach in which the entire BAC library of 147,456 clones on 384 microtiter plates is arranged into 32 superpools, each composed of 12 sequential 384 well plates. The clones in these 12 plates are systematically combined into plate, row, column and diagonal (PRCD) pools. These PRCD pools are further combined following the Matrix Technology Strategy, which reduces the number of pools to be screened in the second step from 76 (PRCD) to 36 (Matrix) for each superpool. Each superpool contains 4,608 unique BAC clones with an average insert size of 118 kb, ($\sim$0.18 $\times$ coverage of the 3,000 Mb haploid sunflower genome). Therefore, it is expected that zero or one template copy of any single copy sequence will be found in a specific Superpool.

Screening the pools and superpools involves two steps: step one is to identify which of the 32 superpools contain a signal for the sequence of interest. Step two is to screen the 36 matrix pools for as many superpools identified in step one as the researcher wants to pursue. For quality control purposes, an internal standard plasmid or total genomic DNA is placed in one of the empty wells of the BAC library. The plate pool that contains this positive control clone is also placed in one well of each PRCD plate to facilitate quality control verification and PCR positive control for each superpool.

Each BAC clone was grown independently to prevent growth competition between clones in microtiter plates containing 2xYT broth with 12.5 μg/ml chloramphenicol. The plates were incubated for 16 h at 37°C. Following growth, the culture for each set of 12 plates was pooled using the method described. Following pooling, sequencing grade BAC DNA preparation was performed on all pooled clones in 96 well format deep blocks using an optimized alkaline lysis method. BAC plasmid DNA was resuspended in 200 μl TE ($\sim$20 ng/μl), then pooled to create the Matrix Master pools.

Pooling of BAC clone cultures was performed on a Biomek 2000 liquid handling robot with customized software and hardware. The growth cultures were combined into 12 plate pools, 16 row pools, 24 column pools and 24 diagonal pools for each superpool. The coordinates of any positive BAC clone can be identified by interpreting the resulting PCR data from experiments using superpool, plate pool, row pool, column pool and diagonal pool DNA as template. The diagonal pools help reduce the number of potential false positive clones when multiple positive clones are contained in a single superpool.

Extraction of BAC DNA occurs on the intermediate stage of the PRCD pools. The BAC DNA from these PRCD pools is further pooled so that each PRCD pool is contained in two matrix pools. Each individual clone has eight occurrences in the matrix pool set (for each superpool), allowing a smaller number of PCR experiments

to correctly identify the unique well position in each super-pool, and providing redundant results to minimize the impact of false negative and false positive PCR results. See Fig. 1 for the composition of the 36 matrix pool.

Superpooled and matrix pooled BAC DNA is prepared as a high concentration master stock and as a diluted PCR working stock, allowing a backup master stock to be kept in case of accidental contamination of the working stock. The highly concentrated resource can also be used in a variety of analytical techniques. All 32 superpools have an individual matrix pool plate. The superpool collection plate contains DNA from the entire library with 4,608 clones in each superpool well. The matrix pools and superpools are contained in 1 ml half height plates sealed with a silicone mat.

Screening of the BAC library

Filter hybridization: filters were hybridized using standard protocols. The probes were labeled with [^{32}P]-dCTP

by random priming according to the manufacturer's instructions (Amersham-Pharmacia Biotech). Filters were prehybridized for 4 h at 65°C in a solution containing 1× HSB, 5× Denhardt's buffer and salmon sperm DNA. They were then hybridized overnight in the same solution at 65°C but containing the labeled probes. Filters were then washed at 65°C in a 2× SSC and 0.1% SDS solution for 15 min, then in a 1× SSC and 0.1% SDS solution for 15 min, and finally in a 0.1× SSC and 0.1% SDS solution for 15 min. The filters were then exposed to X-ray films with two intensifying screens.

PCR screening

The PCR amplifications were carried out with 2 μl of each template in the presence of 0.2 mM of each dNTP, 0.2 μl of Taq DNA polymerase (Advantage 2, Clontech), 1× Taq polymerase buffer and 1 μM of each primer, in a final volume of 20 μl. Following initial denaturation at 95°C for 5 min, 35 PCR cycles of 95°C for 30 s, 60°C for
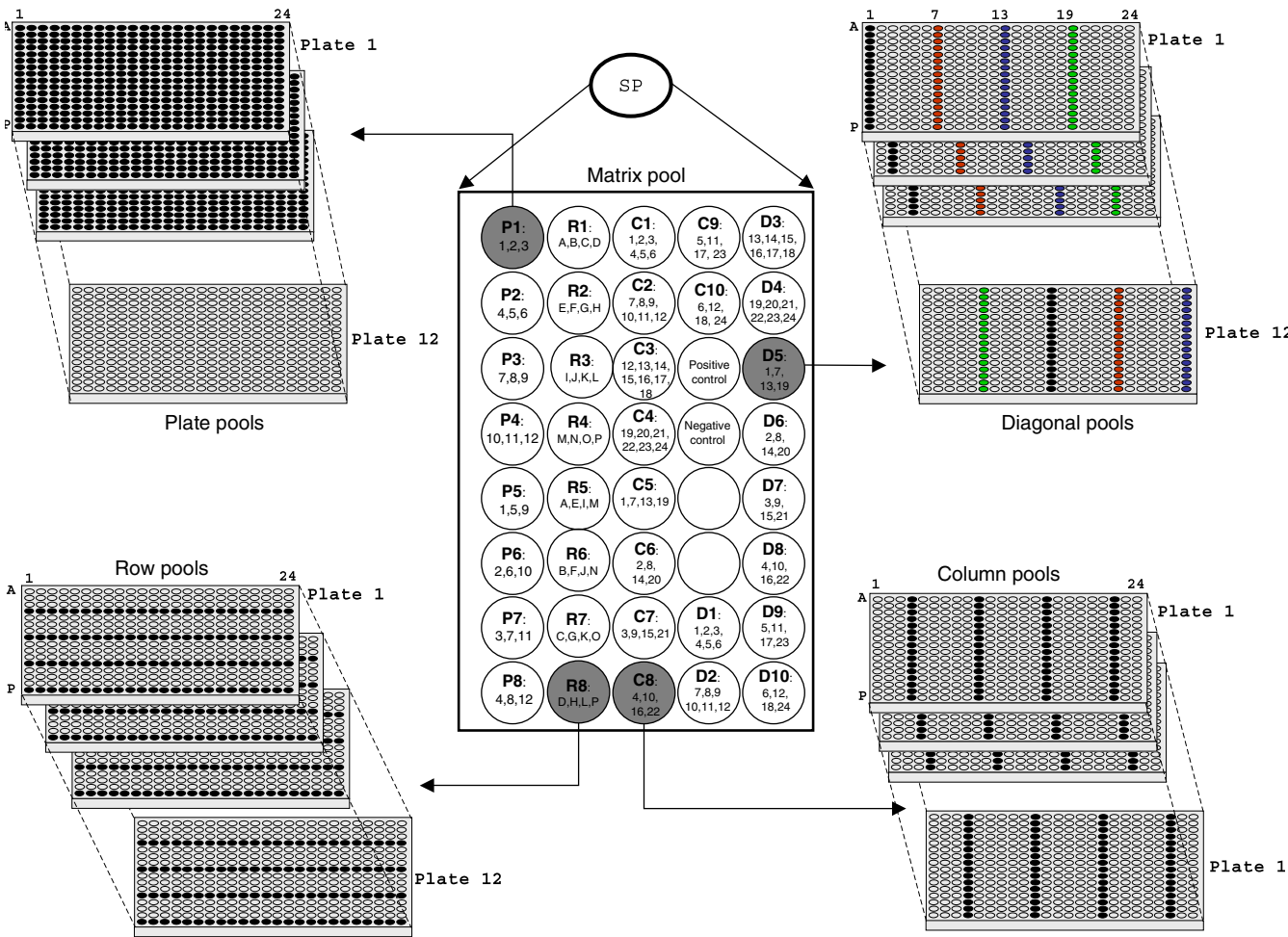

Fig. 1 Schematic representation of the pooling strategy. Twelve 384-well plates are pooled to form a superpool (*SP*). The entire library consists of 32 superpools. Each of the 32 superpools corresponds to a matrix pool containing pools of Plates, Rows, Columns and Diagonals as depicted in the figure. Examples are given (in *grey*) for the Plate pool "P1:1,2,3", the Row pool "R8:D,H,L,P", the Column pool "C8:4,10,16,22" and the Diagonal pool "D5: 1,7,13,19". The *black circles* indicate the clones that have been mixed to make up each pool. Positive control contains DNA from the sunflower line YDQ. Negative control contains no DNA

30 s and 72°C for 1 min were performed. PCR products were separated using 2% standard TAE agarose gel electrophoresis.

BAC fingerprinting and contig assembly

BAC DNA were isolated and digested with *Hin*dIII, end-labeled with [^{32}P]dATP and then subjected to 3.5% polyacrylamide DNA sequencing gel electrophoresis. The gels were dried and autoradiographed. The fingerprints on the autoradiographs were scanned into images and edited using the Image 4.0 software (Soderlund et al. 1997). The BAC contigs were assembled using the version 4.7 of the FPC program (Soderlund et al. 2000).

Results

BAC library construction and characterization

The present library consists of 147,456 BAC clones and is represented by a total of 384 384-well microtiter plates. In order to evaluate the average BAC insert sizes, 297 clones were randomly chosen and the corresponding DNA was extracted, digested with the rare-cutter *Not*I enzyme and analyzed by PFGE (Fig. 2a). All fragments generated by *Not*I digestion contained 8.7 kb vector band and various insert fragments, except

six presumed to be non-insert clones. The estimated insert sizes ranged from 40 to 210 kb with an average of 118 kb (Fig. 2b). The haploid genome size of the cultivated sunflower *H. annuus* L. is estimated to be 2,800–3,400 Mb (Arumuganathan and Earle 1991), hence the coverage of this library is predicted to be 5–6 haploid genome equivalents, and thereby the probability of finding any specific sequence in this library is estimated to be greater than 99%.

PCR screening of the library

To test the feasibility of the pooling scheme used for this library, we used one primer pair for PCR screening. The primers amplify a partial fragment of about 150 bp of Ha-PR5 gene in sunflower (Radwan et al. 2005). The first PCR round detected two superpools (16 and 28) with positive amplification compared to the signal detected using as template the DNA from the parent from which the library was constructed (Fig. 3a). The corresponding PRCD pools were subsequently used in a second PCR round (Fig. 3b). The example presented corresponds to the matrix pools of the superpool number 28. Thus, the PRCD coordinates of this particular clone are respectively: plate number 336 in the library, Row: N, Column 16 and Diagonal 5 which contains the 16th column of the last plate in the superpool (see Fig. 1, Materials and methods section). To ensure that the amplified fragments correspond to the initial target, these fragments were

Fig. 2 Characterization of the sunflower BAC library. **a** Pulsed field gel electrophoresis of randomly selected BAC clones. DNA samples from 24 BAC clones were digested with *Not*I, and the average insert size and rate of non-recombination were determined. **b** Insert size distribution in the library. Insert sizes were determined on 297 randomly selected BAC clones. Six clones are non-recombinant. The *horizontal axis* refers to the size in kb and *vertical axis* indicates the number of clones in each size range. *M* refers to the high range PFG markers (New England Biolabs). The 8.7 kb band corresponds to the plasmid vector
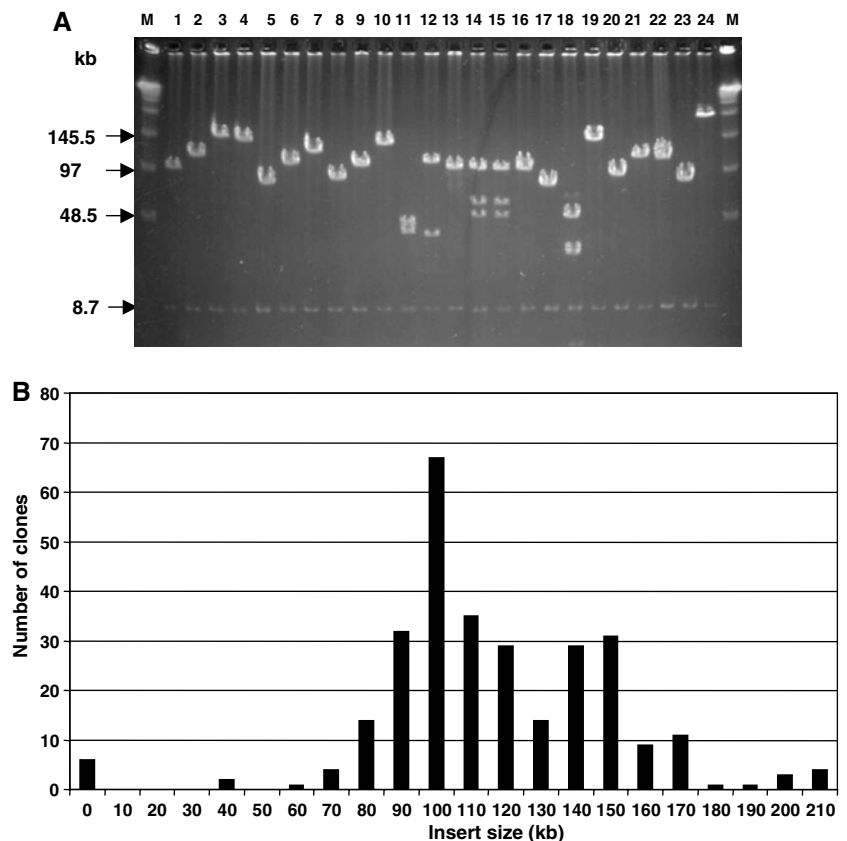
Fig. 3 PCR screening of the pools and superpools of the sunflower library. **a** First round PCR using the 32 superpools which represents the entire library or 384 384-well microtiter plates. Each superpool contains DNA of 12 plates or 4,608 individual BAC DNAs. Two superpools (16 and 28) contain a signal for Ha-PR5 gene coding for the pathogenesis-related protein 5. **b** Second round PCR using the 36 matrix pools to screen the superpool number 28 identified in step one. The clones in the 36 matrix pools are systematically combined into Plate (X 8), Row (X 8), Column (X 10) and Diagonal (X 10) (PRCD) pools. Sunflower genomic DNA from the line YDQ represents the positive control (*YDQ*). *W* indicates the negative controls (water replaced BAC DNA)
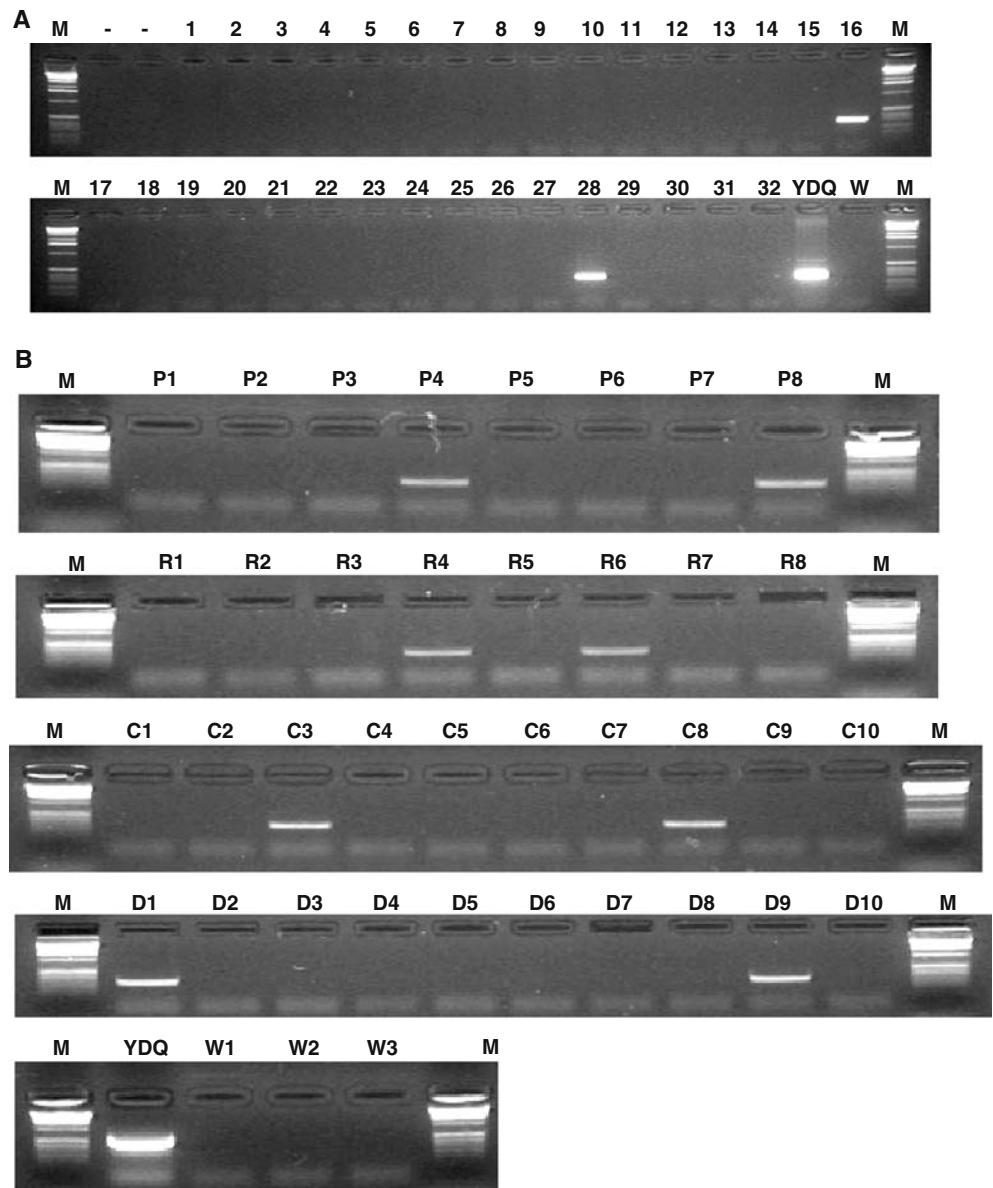
cloned and sequenced. The sequences obtained correspond exactly to the one from which the primers were derived, indicating that the different clones are containing the same target sequence. Subsequent fingerprinting of the two BAC clones by *Hin*dIII digestion proved that they are overlapping.

Anchoring the BAC clones on the SSR map of sunflower

To test whether this strategy could be widely used for different PCR markers, we used a set of SSR markers to screen the library. One region highly covered by SSR markers was selected from the sunflower SSR map described by Tang et al. (2002). Twenty-five SSR markers detecting unique loci within the linkage group 10 of this map were used to screen the BAC library. All the SSR markers detected at least one positive BAC clone in the library. A total of 113 BAC clones were identified

and the number of BAC clones detected by each SSR marker ranged from 1 to 8 giving on average 4.8 BAC clones per SSR marker. This is in the range of the estimated 5–6 redundancy. Conversely, no BAC clone was found to harbor two different SSR markers. These results are summarized in Fig. 4.

Fingerprinting of the positive BAC clones

Since several BAC clones were detected by the same SSR marker, it was questionable whether these BAC clones are identical or overlapping. All the BAC clones (i.e., 112 BAC clones) were thus *Hin*dIII digested and fingerprinted using various FPC iterations and parameters (Soderlund et al. 2000). The results are given in Fig. 4. As it was expected, the majorities of the BAC clones that share the same SSR marker are overlapping and thus belong to the same BAC contig. This is consistent with
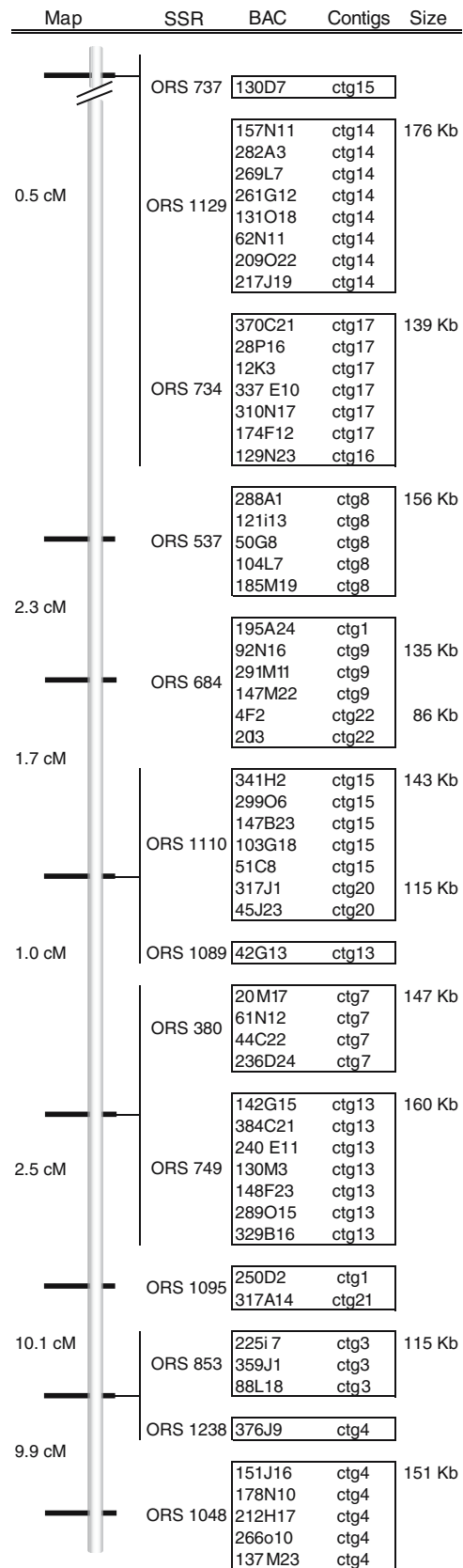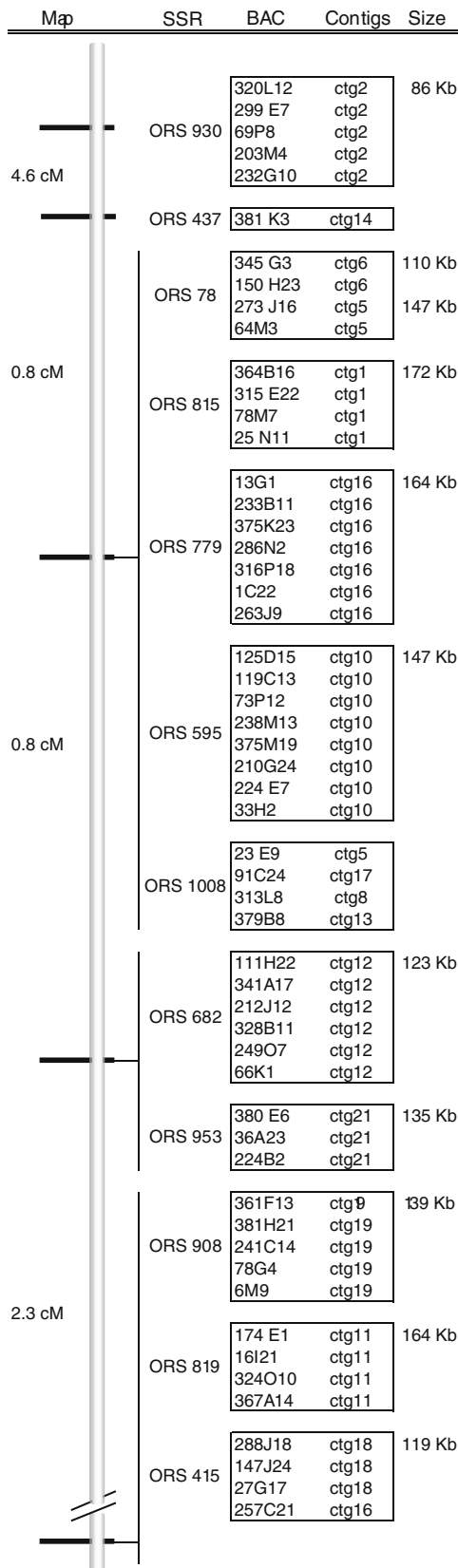
Map	SSR	BAC	Contigs	Size
		320L12	ctg2	86 Kb
		299 E7	ctg2	
	ORS 930	69P8	ctg2	
		203M4	ctg2	
4.6 cM		232G10	ctg2	
	ORS 437	381 K3	ctg14	
		345 G3	ctg6	110 Kb
	ORS 78	150 H23	ctg6	
		273 J16	ctg5	147 Kb
0.8 cM		64M3	ctg5	
		364B16	ctg1	172 Kb
	ORS 815	315 E22	ctg1	
		78M7	ctg1	
		25 N11	ctg1	
		13G1	ctg16	164 Kb
		233B11	ctg16	
		375K23	ctg16	
	ORS 779	286N2	ctg16	
		316P18	ctg16	
		1C22	ctg16	
		263J9	ctg16	
		125D15	ctg10	147 Kb
		119C13	ctg10	
		73P12	ctg10	
0.8 cM	ORS 595	238M13	ctg10	
		375M19	ctg10	
		210G24	ctg10	
		224 E7	ctg10	
		33H2	ctg10	
		23 E9	ctg5	
	ORS 1008	91C24	ctg17	
		313L8	ctg8	
		379B8	ctg13	
		111H22	ctg12	123 Kb
		341A17	ctg12	
	ORS 682	212J12	ctg12	
		328B11	ctg12	
		249O7	ctg12	
		66K1	ctg12	
		380 E6	ctg21	135 Kb
	ORS 953	36A23	ctg21	
		224B2	ctg21	
		361F13	ctg19	139 Kb
		381H21	ctg19	
	ORS 908	241C14	ctg19	
		78G4	ctg19	
2.3 cM		6M9	ctg19	
		174 E1	ctg11	164 Kb
	ORS 819	16I21	ctg11	
		324O10	ctg11	
		367A14	ctg11	
		288J18	ctg18	119 Kb
	ORS 415	147J24	ctg18	
		27G17	ctg18	
		257C21	ctg16	

Map	SSR	BAC	Contigs	Size
	ORS 737	130D7	ctg15	
		157N11	ctg14	176 Kb
		282A3	ctg14	
		269L7	ctg14	
0.5 cM	ORS 1129	261G12	ctg14	
		131O18	ctg14	
		62N11	ctg14	
		209O22	ctg14	
		217J19	ctg14	
		370C21	ctg17	139 Kb
		28P16	ctg17	
		12K3	ctg17	
	ORS 734	337 E10	ctg17	
		310N17	ctg17	
		174F12	ctg17	
		129N23	ctg16	
		288A1	ctg8	156 Kb
		121i13	ctg8	
	ORS 537	50G8	ctg8	
		104L7	ctg8	
2.3 cM		185M19	ctg8	
		195A24	ctg1	135 Kb
		92N16	ctg9	
	ORS 684	291M11	ctg9	
		147M22	ctg9	
		4F2	ctg22	86 Kb
1.7 cM		2O3	ctg22	
		341H2	ctg15	143 Kb
		299O6	ctg15	
		147B23	ctg15	
	ORS 1110	103G18	ctg15	
		51C8	ctg15	
		317J1	ctg20	115 Kb
		45J23	ctg20	
1.0 cM	ORS 1089	42G13	ctg13	
		20M17	ctg7	147 Kb
	ORS 380	61N12	ctg7	
		44C22	ctg7	
		236D24	ctg7	
		142G15	ctg13	160 Kb
		384C21	ctg13	
		240 E11	ctg13	
2.5 cM	ORS 749	130M3	ctg13	
		148F23	ctg13	
		289O15	ctg13	
		329B16	ctg13	
	ORS 1095	250D2	ctg1	
		317A14	ctg21	
		225i 7	ctg3	115 Kb
10.1 cM	ORS 853	359J1	ctg3	
		88L18	ctg3	
	ORS 1238	376J9	ctg4	
		151J16	ctg4	151 Kb
		178N10	ctg4	
9.9 cM	ORS 1048	212H17	ctg4	
		266o10	ctg4	
		137 M23	ctg4	

◄

the fact that the SSR markers used in this study correspond to single mapped loci (Tang et al. 2002). Seven SSR markers detected three or more BAC clones with predicted overlaps that have been confirmed by fingerprinting analysis. For example, the SSR marker ORS 595 identified eight BAC clones that formed the unique contig 10. In contrast, some of the BAC clones that share the same SSR marker were found within different contigs even when the cutoff value of 10 was used (low fingerprinting stringency, Soderlund et al. 2000). For example, the ORS 78 SSR marker detected four positive BAC clones which were placed by fingerprinting in two distinct contigs.

Discussion

BAC library construction and characterization

We have constructed a sunflower BAC library consisting of 147,456 clones with an average insert size of 118 kb, thus representing approximately 5–6 sunflower haploid genome equivalents. The insert size of around 100–200 kb makes it an ideal source for physical mapping, FISH analysis to localize the position on chromosomes, isolation of closely linked polymorphic markers, exon trapping, and identification of regulatory elements in gene expression studies (Buckler et al. 1999; Ota and Amemiya 1996).

The clones were produced using *Hin*dIII partially digested DNA from nuclei of the sunflower line YDQ which is resistant to several races of the sunflower downy mildew pathogen *P. halstedii* (Bouzidi et al. 2002). The BAC clones were arrayed on high-density filters for hybridization with radiolabeled probes and several screening by different probes have been successfully carried out (data not shown). A sunflower BAC library with comparatively similar characteristics (i.e., 5- to 6-fold and contains about 150,000 BAC clones) has been previously described by Gentzbittel et al. (2002). A second BAC library with 60 kb as average insert length and covering about 1.9× sunflower haploid genome has been reported by Ozdemir et al. (2004). However, these libraries are made up from different sunflower lines with different agronomic traits which render them complementary genomic resources for the scientific community working on this crop. Along with the production of high-density filters for colony hybridization with radioactively labeled probes, we designed a pooling and superpooling strategy

suitable for large and systematic screening of this library using a wide range of PCR-based markers.

Pooling and superpooling strategy

This pooling design was based on a BAC library with 147,456 clones (average insert 118 kb) in 384 384-well plates with approximately 5.8× haploid genome coverage. The Matrix pooling strategy was developed to address the following important factors. First was to minimize the number of PCR experiments required to screen the library. In this strategy, only 72 PCR reactions (32 superpools + 36 matrix pools + 4 positive and negative controls, see Fig. 3) are needed to detect a positive BAC clone. The second consideration in the Matrix pooling design was oversampling the pooled resource to minimize the impact of false positive or false negative PCR results. The final important design factor was to allow the researcher to perform large numbers of PCR experiments on the Pool and Superpool resource (+ 20,000 independent screenings).

Ozdemir et al. (2004) described a pooling strategy consisting of 273 plate pools, 144 column pools, and 96 row pools. Thus, 513 PCR reactions are needed to identify a particular BAC clone which is significantly higher than in the strategy proposed in this study (72 PCR reactions).

The Asakawa et al. (1997) pooling strategy used to study the human genome that consisted of 96,000 clones (∼ 3× haploid coverage) arranged for screening as 10 superpools with 9,600 clones in each superpool. Their pooling strategy was a two-step method where the superpool was first identified. Then the 40 pools (from four-dimensionally assigned pools of the specific superpool) were screened to identify the specific clone in the second step. Their pooling method did not include any oversampling which will always lead to significant difficulties if multiple clones, false negative or false positives occur during PCR screening.

Klein et al. (2000) used a six dimensional pooling strategy to screen 24,576 sorghum BAC clones that were about 4× haploid genome coverage. Their strategy allowed a single screen of 184 pools that would identify between two and six positive clones, 72% of the time and 90% of the time being able to reliably assign the coordinates. Their oversampling insures good identification of up to six clones and allowed screening of the complete library in one step. However, this strategy appears to have difficulty in reliably assigning coordinates if more than six positive clones are found in the library. For a large library with 147,456 clones, this strategy would

have to be repeated at least once or performed on a large block design. For example, using a large block size of $64 \times 48 \times 48$ would have 3,072 in five of the six dimensions and 2,304 in the sixth dimension pool. This $64 \times 48 \times 48$ design would require a total of 336 pools to maintain the six dimension design. The design would have a decreased ability to identify correct clones when compared to a small block design, especially with a $5.8 \times$ coverage BAC library. With the Matrix pooling design, the same 336 PCR reactions could follow eight superpool hits in eight matrix plates to find eight positive clones with very high reliability. For example, if there were six hits in six superpools, following them all would only take a total of 248 PCRs.

Screening the library by PCR-based markers

Several different PCR markers have been successfully used to screen this BAC library (data not shown). These markers were derived either from sunflower EST representing unique genes and small multigene family, or from SSR markers which correspond to repetitive sequences within the sunflower genome (Tang et al. 2002). Nevertheless, we were able to detect specific BAC clones harboring these repetitive sequences and corresponding to unique loci in the sunflower map. The repetitive nature of these markers almost precludes their use as radiolabeled probes since they will detect several hundreds of BAC clones. For example, in a survey of the distribution of LINEs and Gypsy-like retrotranspons elements in *Hordeum* species, Vershinin et al. (2002) showed that in hybridization conditions where sequences with 70–75% identity will cross-hybridize, the Gypsy-like probe HvGyp13 detected up to 4,792 BAC clones. Thus, we expect that any PCR marker can be used to identify the corresponding BAC clone with an accuracy that is limited only by the specificity of the primers used. Similarly, we prepared DNA from the 32 superpools enough to achieve up to 20,000 first-round PCR screenings. Considering that the sunflower genome size is about 3,000 Mb and sufficient PCR-based marker evenly distributed throughout the genome are provided, this amount of DNA is roughly sufficient to anchor up to one PCR-based marker every 150 kb. This is beyond the average size of the BAC clone inserts of this library. Alternatively, PCR reactions can be multiplexed which should significantly increase the number of total possible PCR reactions. Therefore, by multiplexing three PCR reactions, it is possible to anchor up to one marker per 50 kb on average. However, these theoretical expectations may be out ruled by practical difficulties such as the development of such a huge number of PCR markers (Klein et al. 2000).

Rapid identification of BAC contigs within a specific region in the sunflower genome

In the present work, we presented the rapid anchoring of a set of 25 SSR markers that were previously mapped to a unique region in the sunflower genome (Tang et al.

2002). Several SSR markers were found tightly linked (i.e., 0 cM) during the map construction; however, none of these SSR detected the same BAC clone in our screening which indicates that at least in this region, the SSR markers were separated by more than 100 kb which is the average size of the BAC inserts. In addition, this distance is probably underestimated since the map was generated using a set of 94 F7 recombinant inbred lines (Tang et al. 2002) which is too small to detect all the cross-over that could happen between SSRs. Alternatively, the selected markers are likely within the centromeric region where the recombination may be severely inhibited.

The BAC clones were fingerprinted and placed in 23 contigs. The fingerprinting results were predictably consistent with the mapping data in that the majority of the BAC clones sharing the same SSR marker were placed in the same contig. However, there are cases that need further analysis such as the cases where BAC clones containing SSR markers at distant loci were placed in the same contig. Since the fingerprinting is based only on shared bands of similar size between clones, this could be explained either by duplication events or the presence of repeated sequences such as transposable elements (Klein et al. 2000). BAC end sequencing and reiterative screening of the library could help to eliminate false positive contigs. Alternatively, Klein et al. (2000) used a method based on AFLP technology to build up a whole integrated genetic and physical map of the sorghum genome. This method identified 2,400 BAC clones organized in 700 contigs and 891 unique new markers. To anchor the physical map to the genetic map of maize, three BAC libraries have been constructed and screened with 90 RFLP core markers providing a framework for the physical mapping and sequencing of the maize genome (Yim et al. 2002). In an other work, Yuan et al. (2000) used an *in silico* strategy to anchor 586 BAC clones to 418 genetic markers of the rice map providing a set of nucleation points throughout the 430 Mb rice genome. However, the sunflower genome is larger (3,000 Mb, Arumuganathan and Earle 1991) and sequencing the ends of all the 150,000 BAC clones should necessitate a huge effort that is beyond the capacity solely of one laboratory. Therefore, the public resource described and the results presented here provide an alternative and simple strategy to provide a framework by detecting seed contigs throughout the sunflower map. The SSR markers described by Tang et al. (2002) are likely to be used in the future in several mapping of agronomic traits in sunflower. Therefore, when an interesting region is identified by SSR mapping, the SSR markers could be used to screen the BAC library and to develop additional markers. The anchoring of the whole set of the 419 sunflower SSR markers is being initiated and the identification of these minimal BAC 'tiling paths' should in turn set the stage for the exploration of targeted genomic regions of interest including the development of markers and positional cloning.

References

Arumuganathan K, Earle ED (1991) Nuclear DNA content of some important plant species. Plant Mol Biol Rep 9:208–219

Asakawa S, Abe I, Kudoh Y, Kishi N, Wang Y, Kubota R, Kudoh J, Kawasaki K, Minoshima S, Shimizu N (1997) Human BAC library: construction and rapid screening. Gene 191:69–79

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

Bouzidi MF, Badaoui S, Cambon F, Vear F, De Labrouhe DT, Nicolas P, Mouzeyar S (2002) Molecular analysis of a major locus for resistance to downy mildew in sunflower with specific PCR-based markers. Theor Appl Genet 104:592–600

Buckler AJ, Chang DD, Graw SL, Brook JD, Haber DA, Sharp PA, Housman DE (1999) Exon amplification: a strategy to isolate mammalian genes based on RNA splicing. Proc Natl Acad Sci USA 88:4005–4009

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

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 9:218–234

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 Genomics 266:979–987

Klein PE, Klein RR, Cartinhour SW, Ulanch PE, Dong J, Obert JA, Morishige DT, Schlueter SD, Childs KL, Ale M, Mullet JE (2000) A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. Genome Res 10:789–807

Ota T, Amemiya CT (1996) A nonradioactive method for improved restriction analysis and fingerprinting of large P1 artificial chromosome clones. Genet Anal 12:173–178

Ozdemir N, Horn R, Friedt W (2004) Construction and characterization of a BAC library for sunflower (Helianthus annuus L.). Euphyt 138:177–183

Radwan O, Mouzeyar S, Nicolas P, Bouzidi MF (2005) Induction of a sunflower CC-NBS-LRR resistance gene analogue during incompatible interaction with *Plasmopara halstedii*. J Exp Bot 56:567–575

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. CABIOS 13:523–535

Soderlund C, Humphray S, Dunham I, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 11:934–941

Tang S, Yu JK, Slabaugh B, Shintani K, Knapp JS (2002) Simple sequence repeat map of the sunflower genome. Theor Appl Genet 105:1124–1136

Tao Q, Wang A, Zhang HB (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

Vershinin AV, Druka A, Alkhimova AG, Kleinhofs A, Heslop-Harrison JS (2002) LINEs and gypsy-like retrotransposons in Hordeum species. Plant Mol Biol 49:1–14

Yim YS, Davis GL, Duru NA, Musket TA, Linton EW, Messing JW, McMullen MD, Soderlund CA, Polacco ML, Gardiner JM, Coe EH Jr (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

Yuan Q, Liang F, Hsiao J, Zismann V, Benito MI, Quackenbush J, Wing R, Buell R (2000) Anchoring of rice BAC clones to the rice genetic map in silico.Nucleic Acids Res 28:3636–3641