A Rapid Way of Physical Mapping in Coconut and Oil Palm

V. Sniady[1], D. Becker[1], A. Herrán[2] , E. Ritter[2] and W. Rohde[1]

[1]Max-Planck-Institut für Züchtungsforschung, Carl-von-Linné-Weg 10, 50829 Köln, GERMANY, mailto:rohde@mpiz-koeln.mpg.de; [2]NEIKER, Apartado 46, Vitoria, Alava, 01080, SPAIN, mailto:eritter@neiker.net

1 Abstract

Some 120,000 cosmid clones of genomic DNA (corresponding to 2-3 genome equivalents) were individualized for oil palm and coconut, respectively, and multiple filter sets prepared for genomic studies, gene isolation and sequencing. Software was developed for and used on commercial robots for the establishment of multidimensional pools from some 28,800 individualized COS clones for both coconut and oil palm. By AFLP analysis on DNAs from these pools with appropriate primer combinations, individual COS clones were mapped onto the molecular linkage maps of coconut and oil palm by association to mapped polymorphic AFLP markers. This approach allows the rapid association of anonymous AFLP markers to clones of genomic DNA.

2 Introduction

Physical maps provide the likely order of defined genomic DNA segments along the chromosomes. If contiguous regions of overlapping cloned DNAs exist, the physical map provides an estimate of the true distance in base pairs between distinct locations on a specific chromosome. Physical maps have been constructed in human, animals and many other organisms. In plants, they have been established for example in *Arabidopsis* (Mozo *et al.*, 1999), sorghum (Klein *et al.*, 2000) or rice (Tao *et al.*, 2001). These maps serve many purposes such as genome-wide gene discovery, EST mapping (functional genomics), or comparative genomics (synteny studies).

In general high-capacity cloning vectors such as bacterial (BAC) or yeast artificial chromosomes (YAC) are used in the construction of genomic libraries, since due to the large insert size a relatively small number of BAC/YAC clones is required for full genome coverage. Physical map construction then encompasses two steps: (i) anchoring of clones to a genetic map and (ii) closing gaps between neighbouring anchored clones. This requires the identification of adjacent, overlapping clones to obtain contiguous genomic DNA segments (contig building). Clones can be assigned to a position on the molecular linkage map by random sequencing, designing appropriate PCR primers based on the sequence information and by using them to amplify the DNAs of individuals of a mapping population for sequence comparison or SNP (single nucleotide polymorphism) development. Whenever amplification products segregate, they can be mapped onto the genetic linkage map and the cloned DNA segments from which they were derived are anchored automatically to the same location.

A different approach would be to assign DNA clones to DNA markers such as AFLP (amplified fragment length polymorphism) markers which have already been mapped onto a molecular linkage map. This can be done on a clone-by-clone basis by molecular hybridisation using the mapped amplified AFLP DNA fragment as a molecular probe. However, the use of appropriate pools of clones and processing by PCR tremendously reduces the required efforts and resources for physical mapping.

In the frame of an EC-funded INCO-DEV project, high-density (HD) molecular linkage reference maps are being developed for coconut (*Cocos nucifera* L.) and oil palm (*Elaeis guineensis* Jacq.), the two most important perennial oil crops of the tropics. The DNA marker types predominantly used in these studies are amplified fragment length polymorphism (AFLP) and simple sequence repeats (SSR, microsatellites), and currently the two linkage maps for both palm species consist of more than 1000 AFLP markers each (data not shown). The aim of this paper is to present an approach of associating cosmid (COS) clones to mapped AFLP markers via high throughput technology and to present some general considerations for developing appropriate pooling strategies.

3 Materials and Methods

Software tools Two relative simple software programs were developed to simulate and analyse the influence of different parameters on the efficiency of a particular pooling strategy and to provide assistance for the practical pooling and clone assignment work:
(i) The BACMIX programme visualises pooling patterns of numbered clones individualised in 384well microtiter plates for specified clone and level numbers and writes the corresponding pooling instructions for robots. By inverting the clone-pool assignment process, this tool allows also to identify the specific clone which corresponds to a particular AFLP band based on its appearance in the specific pools of each dimension. An unambiguous assignment is only possible if a band is present in exactly one level (represented by a particular pool) of each dimension.
(ii) The BACSIM programme was developed to simulate and analyse the genomic distribution of the clones generated by a particular pooling experiment. The user specifies as variables the genome size, the average insert size of the clones, the number of clones in each set (set size) and the number of sets to be generated. BACSIM performs the following analyses for each specific design: Genome equivalents (GE) for a set of clones from a specific design are calculated using the formula GE=*v2 v3*/*v1* (v1= haploid genome size in bp, v2= set size (i.e. number of clones per set), v3= average insert size in bp). The programme then proceeds to determine the fragment distribution over the genome. For this purpose each chromosomal unit representing a potential "AFLP" fragment is screened for presence in the collection of all DNA pieces of each set. In this way the percentages of fragments which are in no clone (OF), fragments which are in exactly one clone (1F) and fragments which are in more than one clone (MF) are determined in one step, as well as the portion of the genome which is covered by any fragment (=1F+MF). The number of PCR reactions required is calculated using as model pools of four levels in each possible dimension resulting from the given set of *v3* clones. Notice that otherwise the simulation is independent of the particular levels and dimensions of the experiment.
For contig analysis start and endpoint of each clone from each set are mapped to the genome and the region covered by the clone is marked. The programme then scans the whole genome counting and determining the lengths of each contiguous region (= contig). In the simplest form a contig may be represented by only one single clone. In this way the software determines contig numbers, minimum and maximum length of contig lengths and average contig length. In addition the average length differences between neighboured contig lengths are computed in order to evaluate the length dispersion which exists between adjacent contigs.

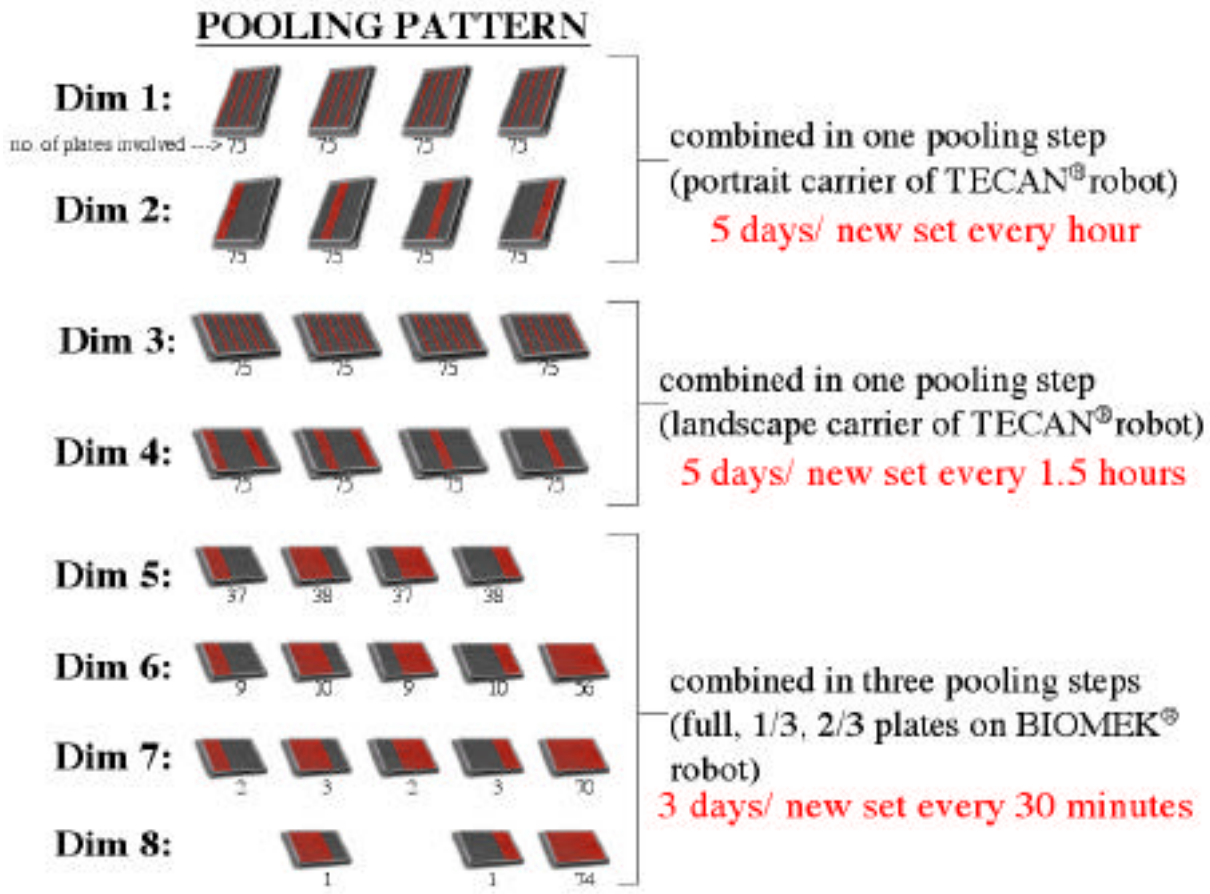
DNA extraction For the preparation of plant DNA, total nucleic acids were extracted from 10 g of leaf material (frozen in liquid nitrogen and converted to a powder in a coffee mill) by the CTAB method of Doyle and Doyle (1990) and treated with RNAse for the removal of RNA (Rohde, 1996). The integrity and concentration of the isolated DNA was determined by electrophoresis in 0.7% agarose gels with the 1 kb ladder (GIBCO-BRL, Groningen, Netherlands) as size and concentration marker.
DNA from single cosmid clones (grown in 80 ml cultures) or COS pools were isolated via reversible adsorption to QIAGEN-Tip100 Midi Prep columns (QIAGEN, Hilden, Germany) according to the protocol of the supplier.

Construction of COS libraries in coconut and oil palm High quality genomic DNA was isolated for both coconut palm (East African Tall, palm number 0707; EAT_{0707}) and oil palm (LM2T) and used according to standard protocols after restriction with *Nde*II for ligation into the cosmid vector SuperCos 1 (Stratagene). Packaged recombinant DNAs (initial colony forming units: 340,000 for oil palm, 420,000 for coconut palm) were not amplified, but stored over weeks for successive infection / transformation and picking of single colonies into 384well plates by robots at the ADIS service unit of MPIZ, Köln. A total of roughly 120,000 cosmid (COS) clones were individualized for oil palm and coconut palm corresponding to 2.5 and 2.3 genome equivalents (oil palm 1.95×10^9; coconut 2.15×10^9 bp genome size), respectively. Both an original library as well as 3 replicates were picked and stored at -80^0 for further usage.

Pool preparation A total of 28,800 individualized COS clones (approximately 0.8 GE) for oil palm and coconut were regrown from the respective libraries in 384well plates. For assembly into 8dimensional pools the robots TECAN®Genesis RSP150 and BIOMEK®Fx were used. The corresponding pooling pattern and the time consumption for each pooling step is depicted in **Figure 1**. According to the

BACMIX program and its practical application, a 7x4+2 design was chosen, indicating the distribution of the selected COS clones into 30 pools (7 dimensions of 4 pools each and the 8th dimension with two pools). Due to the unique presence of each clone in each dimension, each clone has to be pooled 8 times in total. Aliquots of 20 µl were withdrawn from replicated and freshly grown COS clone cultures in a NUNC 384well plate to ensure a sufficient amount of each COS clone in the pool for each dimension. The pooling of the dimensions 1 and 2 (rows) as well as 3 and 4 (columns) was combined in two processing steps (**Fig. 1**) on a TECAN robot (Genesis RSP150) by using permanent 8-channel tips and a portrait carrier for rows pooling and landscape carrier for columns pooling. The pooling of the dimensions 5 to 8 (full, 1/3rd, 2/3rd plates) was processed in three steps (**Fig. 1**) on a BIOMEK robot (FX) by using one-way 96-channel tips. The pool portions were frozen in 50 ml Falcon tubes immediately after each process. For cosmid DNA preparation from the pools, the QIAGEN-Tip100 Midi Prep-protocol for low-copy plasmids was used. COS pool DNAs were redissolved in 200 µl of 0.2xTE-buffer and stored at -20°C.

Figure 1: Display of a pooling experiment with 28,800 COS clones organized into 30 8dimensional pools.



AFLP analysis For AFLP, DNA was restricted by digestion with *Eco*RI and *Mse*I followed by the ligation of the adapters (*Eco*RI: CTCGTAGACTGCGTACC, CATCTGACGCATGGTTAA; *Mse*I: GACGATGAGTCCTGAG, TACTCAGGACTCAT), pre-amplification in the presence of universal *Eco*RI and *Mse*I primers and final amplification with specific primers as described previously (Castiglioni *et al.*, 1998). PCR amplification of genomic DNA followed established protocols (Vos *et al.*, 1995). PCR amplifications were carried out in a final volume of 30 µl containing 10-50 ng of DNA template, 0.2 mM dNTPs, 2.5 mM $MgCl_2$, 1x PCR buffer (GIBCO-BRL, Groningen, Netherlands), 1 pmol each of ^{33}P-labelled AFLP primers and 1 unit of Taq DNA polymerase (GIBCO-BRL). The amplification programs followed the protocols established for AFLP (Vos *et al.*, 1995). Aliquots of the reaction mixture were separated on a 4% polyacrylamide sequence gel. After electrophoresis, the gel was fixed, dried and exposed to X-ray film.

4 Results and Discussion

Mapping of associating COS clones

For coconut, selected AFLP marker - COS clone associations and their linkage group affiliations on the EAT_{0707} are shown in **Table 1**. These 10 AFLP primer combinations have previously produced 90 mapped polymorphic fragments (data not shown) for the coconut mapping population EAT_{0707} x PRD (Herrán *et al.*, 2000). For 7 primer combinations, we were able to identify a total 9 COS clones with association to correspondingly mapped AFLP markers.

Table 1: Association of COS clones to AFLP markers and linkage groups

DNA marker - COS clone associations
of the mapping population EAT0707 x PRDmari

PC no.	PC	MT	NPF	NAF	P/F no.	P/F (M)	LG	Loc	COS
#115	E42xM38	AFLP	11	1	#374	P115/1	3	L-1/3	#49/8/6
#118	E42xM41	AFLP	7	1	#388	P118/4	2	G2	#70/9/12
#143	E44xM41	AFLP	8	0					
#148	E44xM45	AFLP	9	0					
#155	E45xM32	AFLP	11	2	#620	P155/5	2	a-184/5	#59/23/4
					#623	P155/8	4	a-159/5	#56/18/10
#158	E45xM37	AFLP	5	1	#643	P158/5	14	G14	#64/2/3
#159	E45xM38	AFLP	16	2	#647	P159/4	4	G4	#51/21/3
					#652	P159/9	16	G16	#65/3/7
#162	E45xM41	AFLP	5	0					
#163	E45xM44	AFLP	10	1	#688	P163/9	11	G11	#28/20/5
#166	E46xM35	AFLP	8	1	#708	P166/6	2	L13/1	#7/24/3
Sum	**10**		**90**	**9**					

Legend to Table 1:

PC: primer combination*
MT: marker type
NPF: number of polymorphic fragments
NAF: number of associating fragments
P/F: polymorphic fragment*
M: marker
LG: linkage group*
Loc: location*
COS: COS clone position (plate/column/row)
*www.neiker.net/link2palm/coconut-online.org

One example for these analyses is shown in **Figure 2** for the AFLP primer combination E45/M38 which generates 16 polymorphic markers (**Table 1**) including the markers 159/4 and 159/9 (derived from EAT): these have been mapped on linkage group 4 and 16 of the EAT_{0707} molecular linkage map as markers 159/4 and 159/9, respectively (**Fig. 3**). The analysis of an autoradiogram originating from the AFLP amplification of the 30 COS pools (**Fig. 2**) shows that the 159/4 and 159/9 AFLP markers are present once in every dimension as expected (red circles in **Fig. 2**). Applying the BACMIX program for clone identification, the two individual COS clones #51/21/3 and #65/3/7 were identified as producing the markers 159/4 and 159/9, respectively. This was verified by isolating DNA from the two indivdual COS clones and subjecting them to AFLP analysis together with the two parents EAT_{0707} and PRD of the mapping population. Thus the physical position of the two identified clones #51/21/3 and #65/3/7 is on linkage groups LG:4 and LG:16 as shown in **Figure 3**.

Current efforts are devoted to further substantiate this assiciation by i) independent mapping of the two COS clones via SNP to the molecular linkage map, and by ii) sequencing the two polymorphic fragments from the COS clones and comparing them to the corresponding EAT_{0707} sequence.

Figure 2: Association of coconut COS clones to mapped AFLP markers.

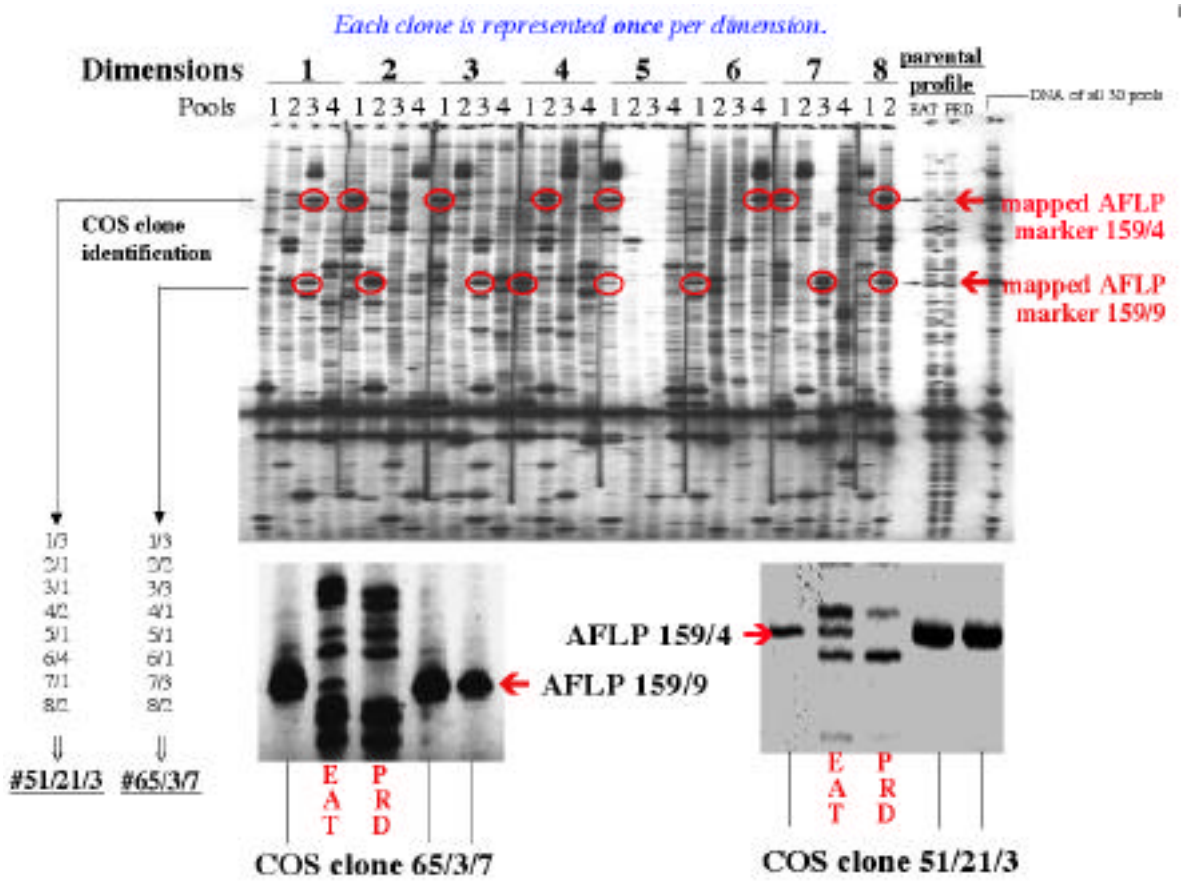
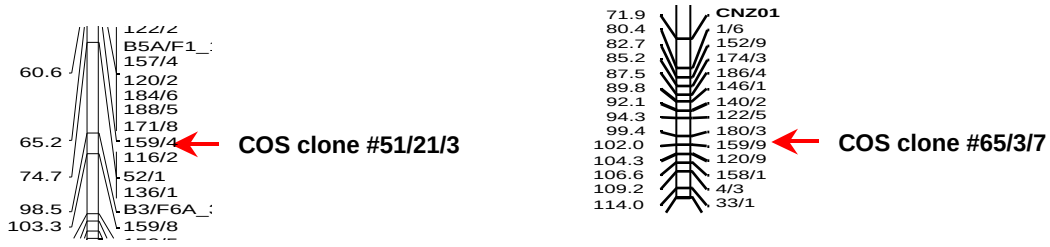


Figure 3: Position of positive clones on EAT_{0707} linkage groups LG:4 (left) and LG:16 (right)



Model validation in a practical approach

The influence of the different variables mentioned above on fragment distribution and for contig building can be evaluated using the BACSIM programme and varying the different variable values e.g. the influence of increasing set sizes for a single set on fragment distribution and contig building, or the influence of varying set numbers with different set sizes (GE<1, GE=1, GE>1) on the pooling efficiency. For coconut, the COS library was prepared for the EAT parent of the mapping population EAT_{0707}xPRD (Herrán *et al.*, 2000). Of the 120,000 individualized clones, 28,800 were assembled into 30 8dimensional pools as shown in **Figure 2**, and DNA was extracted from these pools and processed for AFLP analysis. Currently, 10 AFLP primer combinations have been applied to the analysis of pooled DNAs and the genomic DNAs of the two parents. **Figure 2** shows the AFLP pattern after amplification with the AFLP primer combination E45/ M38 (E45: GACTGCGTACCAATTCATG; M38: GATCAGTCCTGAGTAAACT; primer combination 159 in Table 2.3.1d; http://www.neiker.net/link2palm/Coco/for1-4a.htm). This combination gives rise the two AFLP markers P159/4 and P159/9 which map on the EAT_{0707} linkage groups LG:4 (P159/4) and LG:16 (P159/9), respectively, of the EAT_{0707} linkage map (http://www.neiker.net/link2palm/).

These polymorphic markers appear once in every dimension and also in the amplification of the parental DNAs (see close-up in **Figure 2**). From the particular pools of every dimension in which they occur, the COS clones #51/21/3 (AFLP marker 159/4) and #65/3/7 (AFLP marker 159/9) were identified. When the DNAs of these individual COS clones were analyzed by AFLP, they in fact produce the AFLP marker of the identical size as compared to the EAT parent. Since the COS library was prepared from the DNA of the EAT parent, the identified polymorphic AFLP DNAs from COS clone and EAT parent are currently sequenced to verify the validity of the approach. Also oil palm COS pools are currently subjected to the identical strategy.

In this paper we present an empirical approach to develop appropriate pooling strategies for physical map construction rather than determine explicitely mathematical formulae explaining the underlying biological processes. As shown above, physical map construction can be greatly facilitated by choosing an appropriate pooling strategy. One crucial point is to maximise the number of anchor points for the clones of a library, represented by unique and mapped AFLP fragments. For this purpose an initial set size of one genome equivalent is optimal, which detects around 37% of all possible fragments and is independent of the particular genome or average insert size. Furthermore, several of these sets have to be processed in parallel to increase rapidly the efficiency of detecting anchor points. However, with increasing set numbers small efficiency increments require increasing PCR efforts.

Acknowledgements

This work was supported by the EC INCO-DEV project # ICA4-CT-2001-10066 (LINK2PALM; http://www.mpiz-koeln.mpg.de/~rohde/link2palm.html).

References

CASTIGLIONI P; POZZI C; HEUN M; TERZI V; MÜLLER KJ; ROHDE W and SALAMINI F (1998). An AFLP-based procedure for the efficient mapping of mutations and DNA probes in barley. *Genetics* **149**: 2039-56.

DOYLE JJ and DOYLE JL (1990). Isolation of plant DNA from fresh tissue. *Focus* **12**: 13-5.

HERRÁN A; ESTIOKO L; BECKER D; RODRIGUEZ MJB; ROHDE W and RITTER E (2000). Linkage mapping and QTL analysis in coconut (*Cocos nucifera* L.). *Theoretical and Applied Genetics* **101**: 292-300.

KLEIN PE; KLEIN RR; CARTINHOUR SW; ULANCH PE; DONG J; OBERT JA; MORISHIGE DT; SCHLUETER SD; CHILDS KL; ALE M. and MULLET JE (2000). A high-throughput AFLP-based method for constructing integrated genetic and physical maps: progress toward a sorghum genome map. *Genome Research* **10**:789-807.

MOZO T; DEWAR K; DUNN P; ECKER JR; FISCHER S; KLOSKA S; LEHRACH H; MARRA M; MARTIENSSEN R; MEIER-EWERT S and ALTMANN T (1999). A complete BAC-based physical map of the *Arabidopsis thaliana* genome *Nature Genetics* **22**:219-20 .

ROHDE W (1996). Inverse sequence-tagged repeat (ISTR) analysis: a novel and universal PCR-based technique for genome analysis in the plant and animal kingdom *Journal of Genetics and Breeding* **50**: 249-61.

TAO Q; CHANG YL; WANG J; CHEN H; ISLAM-FARIDI MN; SCHEURING C; WANG B; STELLY DM and ZHANG HB (2001). Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. *Genetics* **158**:1711-1724.

VOS P; HOGERS R; BLEEKER M; REIJANS M; VAN DE LEE T; HORNES M; FRIJTERS A; POT J; PELEMAN J; KUIPER M and ZABEAU M (1995). AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**: 4407-14.