RESEARCH ARTICLE
doi: 10.2306/scienceasia1513-1874.2005.31.299

ScienceAsia 31 (2005): 299-306

A Rapid Construction of a Physical Contig across a 4.5 cM Region for Rice Grain Aroma Facilitates Marker Enrichment for Positional Cloning

Samart Wanchana,[a,b] Wintai Kamolsukyunyong,[a] Siriphat Ruengphayak,[a] Theerayut Toojinda,[a] Somvong Tragoonrung,[c] and Apichart Vanavichit[a,d,e] *

[a] Center of Excellence for Rice Molecular Breeding and Product Development, National Center for Agricultural Biotechnology, Kasetsart University, Kamphaeng Saen, Nakhon Pathom, 73140, THAILAND
[b] Interdisciplinary Graduate Program in Genetic Engineering, Kasetsart University, Bangkok, THAILAND
[c] DNA Technology Laboratory, National Center for Genetic Engineering and Biotechnology, at Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, THAILAND
[d] Rice Gene Discovery, National Center for Genetic Engineering and Biotechnology, at Kasetsart University, Kamphaeng Saen, Nakhon Pathom 73140, THAILAND
[e] Agronomy Department, Kasetsart Univerisity Kamphaeng Saen, Nakhon Pathom 73140, THAILAND

* Corresponding author, E-mail: apichart@dna.kps.ku.ac.th

Received 9 Mar 2005
Accepted 8 Jun 2005

ABSTRACT: The whole-genome BAC-end sequences (BES) and Fingerprint Contig (FPC) are valuable resources for map-based cloning. The rice grain aroma quantitative trait loci (QTL)located on a 4.5-cM region of the rice chromosome 8 was targeted for positional cloning. Initially, a fine-scale physical map was developed *in silico* by aligning selected BAC clones using BES and FPC. The total physical maps spanned 1,114 kb across the flanking markers. Forty microsatellite markers were identified from the BES derived from BAC clones existing in the attained FPC. Five polymorphic SSR markers of those, together with an EST and two previously developed SSR markers, were genetically mapped in F_9 and graphically genotyped in F_{11} plants derived from aromatic x non-aromatic sister-line crosses. Aroma and a tightly-linked marker were mapped within 1.1 cM region of approximately 170 kb flanked by newly develped markers, 10L03_FW and CP04133. The fine location of rice grain aroma was narrowed down from the original 1,114 kb to 170 kb by the *in silico* approach. The plants containing such region from the jasmine rice (KDML 105) were all aromatic and able to produce 2-acetyl-1-pyrroline (2AP). The results demonstrated the success of using the *in silico* approaches as an alternative to genetic approach for physical mapping and marker enrichment. Such *in silico* technique is more applicable to crops when genetic tools are limited.

LIST OF ABBREVIATIONS: BES, BAC-end sequence; FPC, Fingerprinting contig; KDML105, Khao Dawk Mali 105; 2AP, 2-acetyl-1-pyrroline; QTL, quantitative trait loci.

KEYWORDS: grain aroma, rice, *Oryza sativa* L., marker enrichment, *in silico* mapping, BAC-end sequence, BAC fingerprint contig, grain aroma, graphical genotype.

INTRODUCTION

Discovering genes underlying quantitative trait loci (QTL) requires well-defined QTL, fine-scale linkage and physical mapping prior to positional cloning steps. In the pre-genomic era, significant progress in positional cloning in plants relied on development of fine-scale mappings and large genomic insert libraries. For example, successful map-based cloning for disease resistance genes were reported for *Pto* in tomato[1], *Xa21* in rice[2], *Mlo* in barley[3], *I2* in tomato[4], *RPM1* in Arabidopsis[5], *RPS2* in Arabidopsis[6] and *RPP13* in Arabidopsis[7].

Fine-scale mapping requires marker enrichment in the target regions where a number of recombination breakpoints must be generated to achieve a desirable mapping resolution. Strategies for marker enrichment in the target region included Bulk Segregant Analysis (BSA)[8], chromosome landing[9] and pooled progeny techniques[10]. Genetically directed representational difference analysis (GDRDA) utilized phenotypic pooling combined with a subtractive method specifically to isolate markers from a locus of interest[11]. In addition, RFLP subtraction has been used to isolate large numbers of randomly located RFLPs spanning the mouse[12], *Volvox carteri*[13] and rice genomes[14]. In many cases, these methods failed to generate enough polymorphic molecular markers in a specific region

for positional cloning. In several organisms for which large inserted libraries are available, bacterial artificial chromosome (BAC) end sequences (BES) and BAC fingerprinting contigs (FPC) were explored as an alternative to a classical physical map. Combined with exhaustive data mining, BES could be ideal for enrichment of informative markers that can be directly isolated at a high density for any low repetitive genomic region. Alternatively, delicate fine-scale mapping and chromosome walking can be performed *in silico*.

Aroma is an important quality characteristic of rice. Successful breeding for aromatic rice depends mainly on effective identification of desirable traits in combination with strong aroma. The most potent volatile compound that has been determined as the major aroma component in any aromatic rice is 2-acetyl-1-pyrroline, or 2AP[15]. Aroma can be evaluated by simple sensory methods such as heating seeds or leaves in water or incubating leaf tissue in diluted KOH[16]. The 2AP compound can be quantified by gas chromatography (GC)[17]. However, the sensory test is sometimes inconsistent because of human preferences, and the detection of the volatile compound by GC is impractical for rice breeders. Therefore, molecular markers specific for grain aroma are needed for effective breeding for superior rice with aroma.

Aroma QTL has been previously mapped on the rice chromosome 8 between two RFLP markers, RG28 and RG1[18], 4.5 cM away from RG28[19]. In this study, we rapidly constructed a physical map using an *in silico* approach and demonstrated its advantages for dissecting a 4.5 cM region on the rice chromosome 8 for grain aroma.

Materials and Methods

In silico Mapping

In order to retrieve FPC-containing Nipponbare BAC clones aligned in the region proximal to aroma QTL, the SSR marker RM223, which has been located near the QTL[20], was searched against the FPC database (http://www.genome.arizona.edu/fpc/WebAGCoL/rice/WebFPC/). An *in silico* contig was constructed by aligning minimal-tiling path clones selected from the retrieved FPC. End-sequences of selected BAC clones were aligned with the pseudomolecule AP008214 by BLAST2 (http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi) for fine adjustment of the *in silico* contig.

Overgo Hybridization

For BAC library screening, the high-density filters containing BAC clones derived from the rice c.v. KDML105 and Nipponbare were hybridized with [^{32}P]dATP- and [^{32}P]dCTP-labeled probes generated from overlapping oligonucleotides (overgo) according to the method described by Vollrath.[21] The overgo probes were designed using the software Overgo maker (http://genome.wustl.edu/tools/index.php?overgo=1) based on BAC-end sequences of the desired clones that were obtained from the Rice BES database (http://www.genome.arizona.edu/stc/rice/). Hybridization was performed at 60°C and the filters were washed with 4x SSC/0.1% SDS at 58 °C for 30 min and then with 1.5× SSC/0.1% SDS at 58°C for 30 min.

SSR Mining from BAC-End Sequences

BAC-end sequences collected from the Rice BES database were scanned for simple sequence repeats (SSR) using Simple Sequence Repeat Identification Tool (SSRIT; http://www.gramene.org/db/searches/ssrtool). One hundred- to five hundred-bp-long sequences harboring each SSR were collected to design primers using Primer3 (http://www.genome. wi.mit.edu/cgi-bin/primer/primer3_www.cgi) and Genosys Oligomail software (Sigma-Genosys Ltd.) to minimize secondary structures, hairpin loops and primer dimers.

Plant Materials and Phenotypic Methods

Two parental lines, Khao Dawk Mali 105 (KDML105), the aromatic rice strain, and CT9993, the non-aromatic rice strain, and their F_8 families were provided by Rice Gene Discovery at Kasetsart University, Kamphaeng Saen, Nakhon Pathom, Thailand. Segregated individuals were selected in F_9 to F_{11} generation based on the recombination events in the target region flanked by the RM342 and RM223.[22] The resultant 186 F_9 and 174 F_{11} individuals were grown in the field and used for fine-scale genetic mapping and graphical genotyping, respectively. F_{10} and F_{12} seeds were collected from panicles of each F_9- and F_{11}-associated plants for aroma testing either by sensory or by gas chromatography methods. For the sensory method, five individual grains from each panicle were placed into each 2-ml screw-cap tube (AxyGen Inc, USA) containing 200 ml of fresh water and were then incubated at 65°C for 2-3 hours. After storage at 4°C for 20 min, the samples were sniffed by three well-trained panelists. Aroma was noted as either aromatic or non-aromatic. Based on their F_{10} and F_{12} single-grain sensory collections from each plant, F_9 and F_{11} plants were categorized into three types of the aroma phenotype: homozygous aromatic (A), homozygous non-aromatic (B) or segregating (H). For quantitative analysis of 2AP, a small-scale, headspace extraction and detection by GC-MS was performed using five-gram seed samples.[23]

Fine-Scale Mapping and Graphical Genotyping

Genomic DNA was isolated from three-week-old leaves using a standard CTAB method.[24] The SSR markers

in this study were selected from a previously published genetic map[25] and newly developed from BAC-end sequences. EST markers were selected from a YAC-based rice transcript map.[26] Linkage analysis was done using MAPMAKER/EXP version 3.0b.[27] Map distances (cM) are reported in Kosambi units.[28] The grain aroma was mapped in F_9 and genotyped on F_{11}.

Polymerase Chain Reaction

PCR was performed in 5 μl reaction mixtures containing 25 ng of DNA template, 0.1 mM of dNTPs, 0.25 mM of each forward and reverse primer, 0.25 unit of *Taq* DNA polymerase, 2.0 mM $MgCl_2$ and 1x Thermophilic DNA Polymerase buffer (Promega). After pre-heated at 94°C for 2 min, the PCR reaction was carried out for 35 cycles under the following conditions: 94°C denaturation for 30 sec, 55 °C annealing for 30 sec and a 72°C extension for 1 min, and allowed a final extension at 72°C for 5 min. The PCR products were resolved on 4.5% denaturing silver-stained polyacrylamide gel electrophoresis (PAGE).

Result and Discussion

In silico Physical Map across a 4.5 cM Region for Rice Grain Aroma

A genome-wide physical map of Nipponbare containing 72,703 BACs has been previously developed by the restriction fingerprinting method (FPC) (http://www.genome.clemson.edu/ projects/rice/fpc/). This physical map consists of 180 BAC contigs, 3,276 singletons and 2,427 anchored markers. End-sequences of those BAC clones, collectively called Sequence Tag Connector or STC, represent 48-Mb genomic sequence with the average of 707 bp per STC. It was estimated that STC should be found on every 9-kb across the 430-Mb rice genome. Therefore, STC is desirable for high-density marker enrichment, as well as rapid construction of physical maps, at any regions of interest.

In this study, we targeted a 4.5 cM genomic region on chromosome 8 for demonstrating the advantages of the *in silico* approach. As a result of searching RM223 against the FPC database, FPC #58 comprising 2,163 BAC clones with a total genomic distance of 8,082 kb was obtained. To demonstrate a rapid *in silico* construction of a physical map encompassing the genomic region responsible for rice aroma, ten minimum-overlapping clones were selected from the FPC #58 across the marker RM223 and RM342 (Figure 1). To verify the *in silico* contig, eight overgo probes derived from the BAC clone a0010H14, b0032E15, a0014C22 and a0006F09 (Figure 1A, 1B) were used to perform high-density filter hybridizations. As a result, all four seed-BACs, as well as their adjacent clones existing in the same FPC, were positive to the Overgo probes (Figure 1B). This result ensures that the existences and locations of clones in the FPC #58 are reliable. To finely adjust the *in silico* BAC contig, end sequences of those ten BAC clones were aligned onto the pseudomolecule accession AP008214 representing a non-overlapping assembled sequences of the rice chromosome 8 (Figure 1C). It was calculated that the total *in silico* contig spans 1,114 kb encompassing the aroma QTL. When compared to the classical mapping approach in which the physical mapping requires

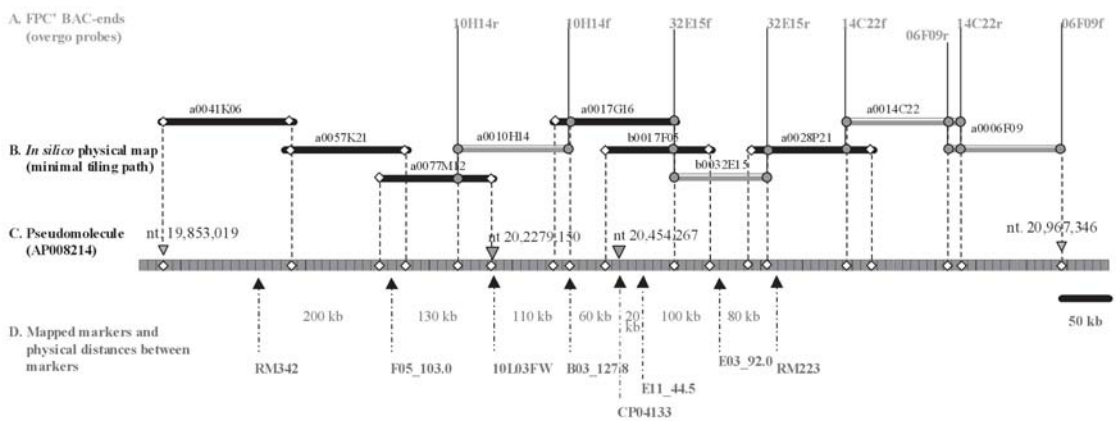

Fig 1. *In silico* physical map spanning the 1,100 kb which harbored the aroma gene region.

A. BAC-end markers used in hybridization of high-density filters.
B. Horizontal bars representing minimal-tiling path Nipponbare BAC clones selected from the FPC contig #58. Clones that was hitted by BAC-end overgo probes are shown as the symbol (■).
C. Pseudomolecule AP008214 representing non-overlapping of genome sequence of rice chromosome 8. Position of each BAC-end anchored on this pseudomolecule are shown as the symbol (◇).
D. Markers used for fine-scale mapping and graphical genotyping. Locations and distances on pseudomolecule AP008214 are indicated.

exhaustive chromosome walking by repeated filter hybridizations, the *in silico* physical mapping demonstrated here is simple, less expensive, and time-saving. This strategy is possible for application to other crops, such as maize, tomato, cotton and barley, of which BES and FPC are available for rapid construction of physical contigs at any regions of interest, or would be applied comparatively to construct physical maps in related species or subspecies to elucidate genome evolution and functions.

SSR Marker Enrichment

In order to generate a fine-scale map for positional cloning of the gene controlling rice grain aroma, markers in high density are needed to place between the flanking markers, RM342 and RM223. In this study we proposed a recent advance in molecular marker development by mining simple sequence repeats (SSRs) from BAC-end sequence data. Initially, several BAC clones located between the flanking markers were selected from the FPC#58. Forward and reverse end-sequences of those BAC clones were collected and subsequently screened for simple sequence repeats (SSRs). As a result, forty SSR loci containing di, tri, or tetra nucleotide repeats were identified from sixteen BAC-ends in the minimum-tiling path retrieved from the contig #58. Most abundant SSRs mined from the BAC-end sequences in the region of interest contained GC-rich (~27%) and (AG)n dinucleotide repeats (~27%) (data not shown). PCR primers were consequently designed to flank those SSRs. The primers were screened for polymorphisms between parental lines, KDML105 and CT9993. Successful PCR amplification was achieved with twenty-two primer pairs, but only five loci were polymorphic between the aromatic and non-aromatic parents (Table 1). The missing SSRs were either loci whose primers were located in the highly repetitive area or loci whose annealing temperatures of primer pairs were difficult for optimization. The physical locations and distances of these five markers on the pseudomolecule of rice chromosome 8 are shown (Figure 1D).

SSRs are important molecular marker tools for applied rice research. Traditionally, SSRs have been discovered through the laborious task of constructing and sequencing SSR-enrich genomic libraries. Mining SSRs from BESs is an alternative to rapidly produce informative markers for either breeding or a gene cloning project. In the case of rice, the utilization of the completed whole genomic DNA sequence may have simplified positional cloning as well as marker generation. However, BES and FPC resources are useful for marker enrichment in many other crops in which whole genomic sequences may not be available.

Fine-Scale Genetic Mapping

The recent work on locating the grain aroma QTL by data mining the rice genomic sequence revealed a microsatellite marker (SSR)[29] and a single-nucleotide polymorphism (SNP) marker[30] closely-linked to the grain aroma QTL by the genetic distances of 4 and 2 cM, respectively. In this study, informative SSR loci generated based on BAC-end sequences retrieved from the BES database were mapped using 186 F_9 individuals derived from KDML105 and CT9993, the aromatic and non-aromatic parents. Five newly developed SSR markers as well as two early SSR markers, RM342 and RM223, and an EST, CP04133, were surveyed for segregation distortions in 177-186 F_9 mapping population. All markers except for RM223 fit the 1:2:1 segregation ratios that corresponded to the single Mendelian ratio (Table 2). The segregation ratio of the grain aroma phenotype in F_{10} seeds also fit the 1:2:1 ratio of single Mendelian segregation ($\chi^2 = 0.194$). F_9 plants were classified as fixed aromatic (A) for 48 lines, segregating (H) for 90 lines and fixed non-aromatic (B) for 48 lines (Table 2). This phenotypic data strongly confirms that aroma is controlled by a single recessive gene.

The mapping individuals were segregated only in the target region on chromosome 8, so the fine-scale map is illustrated only in this region in Figure 2. The total genetic distances of the fine-scale map encompassing the two flanking markers, RM342 and E03_92.0 was 3.8 cM. As expected, the newly developed SSR markers derived from sequence analysis were all

Table 1. Five SSR markers derived from Nipponbare's BAC-end sequences and a re-designed EST marker. For each SSR the repeat motif is shown. All markers was designed to fit annealing temperature at 55 °C.

No.	Marker name	Marker type	L primer (5'-3')	R primer (5'-3')	SSR motif
1	F05_103.0	SSR	CTCCGTGCAAAGGAGGTACT	CCATACCCCAGCTTCTCTCA	CCG
2	10L03_FW	SSR	AACTAAGCACAACGCAAGGC	TCACTCTAATTGGCCTGGTTTT	GA
3	BO3_127.8	SSR	CGTGGCTCGACCTTTTTAAT	TCAAACCCTGGTTACAGCAA	ATG
4	CP04133	EST	CAGAGGCAATGCAAGAACAG	CGATCATCCCGAATTCATTT	-
5	E11_44.5	SSR	TCATGATATGGGCACTGCTG	TCGATCCAGGAGGAGAGCTA	TC
6	E03_92.0	SSR	GCCATGGCTAAGCTAGGATTC	ATCCGCGTACTCTCTCCTCA	CGG

Table 2. List of molecular markers used for genetic mapping around the heterozygous region in 177-186 F_9 population.

Marker name	*Thai Hom Mali Rice (A)*	*Hetero-zygous (H)*	*CT 9993 (B)*	Total	χ^2	Probability
RM342	49	89	45	183	0.311	0.85578363
F05_103.0	48	90	47	185	0.146	0.92962595
10L03_FW	46	84	47	177	0.469	0.79099528
B03_127.8	47	91	48	186	0.097	0.9527649
CP04133	48	83	47	178	0.820	0.66357569
E11_44.5	52	86	48	186	1.226	0.54177569
E03_92.0	52	87	47	186	1.043	0.59362624
RM223	69	78	38	185	14.935	<0.001
Aroma	48	90	48	186	0.194	0.90776096

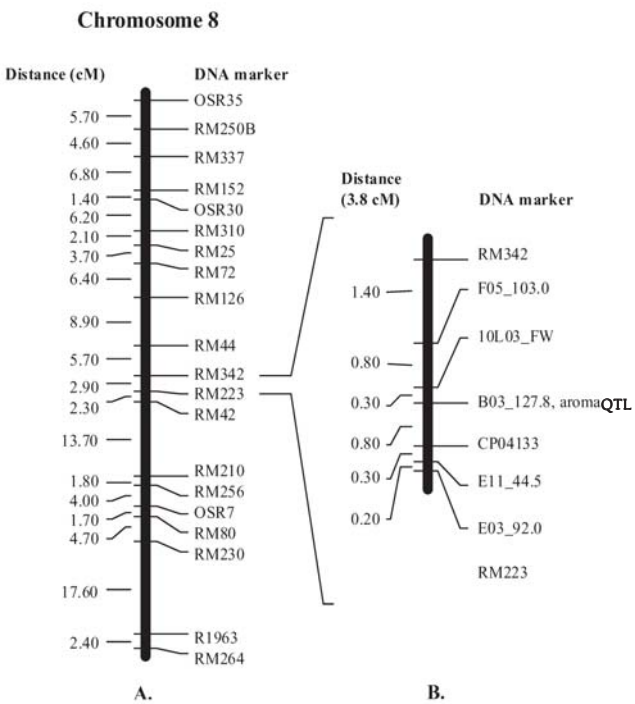

Fig 2. Genetic mapping at the target region controlling grain aroma.

A. Integrated map from three populations as described by Toojinda *et al*. 2003. RM223 and RM342 were separately mapped by the distance of 2.9 cM.

B. The fine-scale map containing five newly developed SSR markers, as well as the early SSR marker RM223 and the EST marker CP04133. Total distance of the fine-scale map is 3.8 cM. The map does not include the marker RM223, since its genotypic segregating data among F_9 population distorted from the 1:2:1 ratio.

located within the target region (Figure 1D). Combined with the two existing SSRs and one EST, the marker density in this target region became 0.6 cM per marker interval. Using 186 recombinant F_9 plants, the genetic location of the grain aroma QTL was mapped within the 1.1-cM interval between the new SSR marker 10L03_FW and the EST marker CP04133 (Figure 2). The aroma QTL completely co-segregated with the SSR marker B03_127.8 at this mapping resolution.

Recombination Events in the Region Proximal to the Rice Grain Aroma QTL

BLAST2 alignments of the flanking markers, RM342 and E03_92.0, and the pseudomolecule AP008214 revealed that the physical distance between the flanking markers is approximately 700 kb (Figure 1D). The physical order of the molecular markers and their genetic positions corresponded well in this mapping population (Figure 1D and 2). Graphical genotyping was performed by examining the marker genotypes and grain aroma phenotypes among 174 F_{11} plants. As a result, eight recombination patterns were found in the genomic region between the markers RM342 and RM223. To demonstrate recombination events in the target genomic region harboring the grain aroma QTL, only eight recombinant plants segregating for the rice grain aroma QTL were selected (Figure 3). Evaluated their respective F_{12} seeds by the single-seed sensory method, F_{11} plants no. 4, 10 and 55 were classified as non-aromatic (B) because no aromatic seed was detected from these plants. Plants no. 125, 16, 19 and 83 were classified as segregating (H) owing to the presence of both aromatic and non-aromatic seeds. Plant no. 117, in which all seeds were aromatic, was classified to aromatic group (A). By the same token, the results of 2AP quantification revealed that 2AP was not detected in class B, whereas 2AP amount was detected at a high level (0.46 ppm.) in class A and a trace amount (0.06-0.11 ppm.) in class H (Figure 3A). The graphical genotypes of the eight F_{11} plants, revealed that the aroma QTL was located in a 170-kb interval between two flanking markers, 10L03_FW and CP04133, and closely associated with the SSR marker B03_127.8 (Figure 3B). Only the plant that contains the B03_127.8 region associated to KDML105 is aromatic, while plants containing either heterozygous or homozygous to CT9993 are non-aromatic. According to graphical genotypes of these eight F_{11} plant, the genotype at RM223 was absolutely fixed for KDML105 while genotypes of other markers were segregated. This confirms the distortion of the marker RM223 in this population. These aroma tightly-linked markers could be useful not only for aromatic rice breeding but also for positional cloning of the aroma gene.

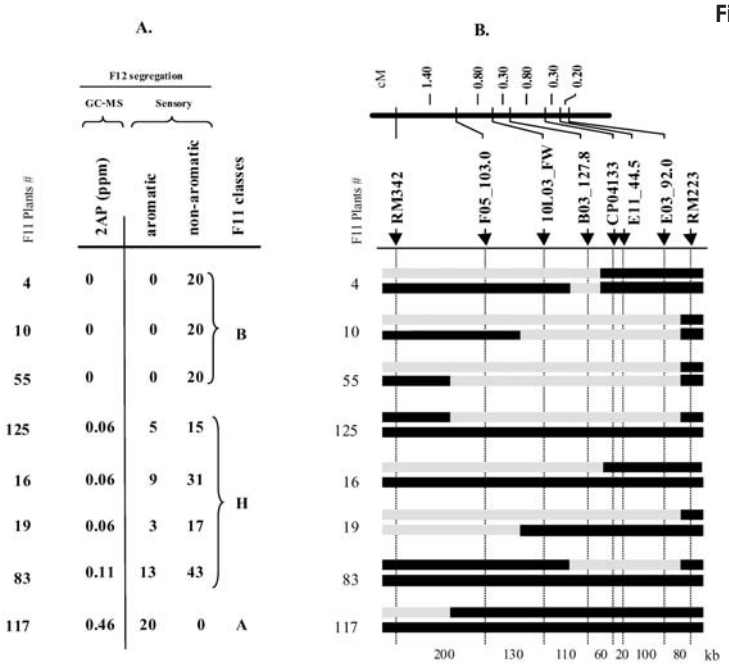

Fig 3. Demonstration of graphical genotyping among the eight F_{11} plants containing different recombination patterns in the region between the markers RM342 and RM223

A. Aroma phenotypes of F_{12} families analyzed by GC-MS and sensory method. The amount of 2AP in pooled seeds from each F_{11} plant is shown in the first column. Evaluation of collections of aromatic and non-aromatic F_{12} seeds evaluated by the sensory method are shown next to the 2AP column. A, B, and H refer to aromatic homozygote, non-aromatic homozygote and heterozygote, respectively.

B. The solid and hatched bars represent aromatic (KDML105) and non-aromatic (CT9993) DNA, respectively. Physical distances in kilobase (kb) between marker intervals are indicated under the graphics. The fine-linkage map is presented above the graphical genotypes.

Table 3. Candidate genes co-segregating with aroma in the region between markers 10L03FW and CP04133

ORF	Location on AP008214	Exon	Coding region size (kb)	RiceFl-cDNA	Coding region %GC content	Protein ID	a.a	kD	Candidate
1	20300020 - 20304269	4	0.369		60.81	BAC99798.1	122	14.092	Hypothetical protein
2	20305341 - 20307475	4	0.831		57.81	BAC99799.1	267	30.082	Putative dioscorin
3	20321930 - 20322357	2	0.378		74.14	BAC99800.1	125	13.399	Hypothetical protein
4	20335508 - 20335317	1	0.612		71.94	BAC99801.1	203	19.999	Hypothetical protein
5	20341923 - 20343731	4	0.421		55.81	BAC99802.1	139	15.05	Hypothetical protein
6	20346386 - 20343864	4	0.843		60.35	BAC99803.1	277	30.299	Putative dioscorin
7	20352613 - 20359121	10	1.719	AK060716, AK058814	50.34	BAC99804.1	572	60.816	Putative methylcrotonyl-CoA carboxylase beta chain, mitochondrial precursor beta chain, mitochondrial precursor
8	20364848 - 20365230	2	0.351		50.00	BAC99805.1	116	13.114	Hypothetical protein
9	20372728 - 20378587	15	1.512	AK071221, AK060461	52.14	BAC98555.1	503	54.682	Putative betaine-aldyhyde dehydrogenase
10	20382323 - 20385046	1	2.724		48.18	BAC55740.1	907	106.671	Putative disease resistance gene homolog
11	20388246 - 20391263	2	2.772		43.09	BAC98556.1	923	104.965	Putative disease resistance gene homolog
12	20392892 - 20393272	1	0.381		64.65	BAC98557.1	126	13.344	Hypothetical protein
13	20394898 - 20398471	7	1.644	AK105302	55.86	BAC98558.1	547	60.911	Unknown protein
14	20401217 - 20410580	22	2.628	AK105302, AK069134	46.36	BAC98559.1	875	95.552	Putative dynamin homolog
15	20416573 - 20420219	6	0.561	AK099569	57.47	BAC98560.1	186	21.585	Unknown protein
16	20422219 - 20423808	1	1.59		70.01	BAC55745.1	529	57.924	Putative endo-beta-1,4-glucanase
17	20432396 - 20432595	1	0.225		52.65	BAC98561.1	74	8.188	Hypothetical protein
18	20439317 - 20441345	4	0.276		44.04	BAC98562.1	91	10.954	Hypothetical protein
19	20442988 - 20444781	7	1.245	AK106998	68.05	BAC55746.1	414	45.65	Inositol polyphosphate-5-phosphatase-like
20	20450912 - 20451115	1	0.204		54.63	BAC98563.1	67	7.57	Hypothetical protein
21	20451629 - 20452222	2	0.402	AK049408, AK104119	66.50	BAC55747.1	133	13.363	Hypothetical protein
22	20452492 - 20454107	3	0.966		70.42	BAC55748.1	321	35.475	Putative calcium-binding protein annexin

Candidate Gene Analysis

The flanking markers, F05_103.0 and E03_92.0, were aligned on the pseudomolecule, AP008214, using BLAST2 to determine the candidate region for the rice aroma gene. As a result, the 170-kb region was aligned from nucleotide 20,279,150 to nucleotide 20,454,267 of the pseudomolecule. The aroma co-segregating marker, B03_127.8, was located in the region from nucleotide 20,394,413 to 20,394,554. Because the pseudomolecule, AP008214, retrieved from the GenBank did not include gene annotation data, we collected gene annotation summary from its original seed clones, AP005537 and AP004463, instead. According to the GenBank CDS features of the clones, AP005537 and AP004463, the 170-kb region was found to contain 22 open reading frames (ORFs). Seven of these ORFs are considered potential candidate genes with hits to full-length cDNA in KOME (http://cdna01.dna.affrc.go.jp/cDNA/) (Table 3), including a putative methylcrotonyl-CoA carboxylase beta chain, a putative betaine-aldyhyde dehydrogenase, two unknown proteins, a putative dynamin homolog, an inositol polyphosphate-5-phosphatase-like protein, and a putative calcium-binding protein annexin. The marker B03_127.8 is placed between the annotated ORF 12 (hypothetical protein) and 13 (putative dynamin homolog).

At present, little is known regarding the molecular mechanism of aroma biosynthesis in rice. Therefore, it seems difficult to specify the key gene responsible for rice aroma based solely on this gene annotation data. Further narrowing down of the candidate region by using additional molecular markers or advanced populations, such as Near Isogenic Lines (NILs) or recombinant sister lines that contain identical genomic backgrounds, except for the target region, is necessary. Alternatively, monitoring differentially-expressed candidate genes between aromatic and non-aromatic lines would also be useful in determining the gene(s) that might be the key gene for rice aroma.

Acknowledgements

This project was jointly funded by the Royal Golden Jubilee Ph.D. Program, DNA Technology Laboratory, Rice Gene Discovery and Kasetsart University. These experiments complied with the current Biosafety guidelines of Thailand.

References

1. Martin GB, Brommonschenkel SH, Chunwongse J, Frary A and Ganal MW (1993) Map-based cloning of a protein kinase gene conferring disease resistance in tomato. *Science* **262**, 1432-6.
2. Song WY, Wang GL, Chen LL, Kim HS and Pi LY (1995) A receptor kinase-like protein encoded by the rice disease resistance gene, *Xa-21*. *Science* **270**, 1804-6.
3. Buschges R, Hollricher K, Panstruga R, Simons G and Wolter M (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance *Cell* **88**, 695-705.
4. Simons G, Groenendijk J, Wijbrandi J, Reijans M and Groenen J (1998) Dissection of the *Fusarium I2* gene cluster in tomato reveals six homologs and one active gene copy. *Plant Cell* **10**, 1055-68.
5. Grant MR, Godiard L, Straube E, Ashfield T and Lewald J (1995) Structure of the Arabidopsis *RPM1* gene enabling dual specificity disease resistance. *Science* **296**, 843-6.
6. Bent AF, Kunke BN, Dahlbeck D, Brown KL and Schinidt R (1994) *RPS2* of *Arabidopsis thaliana* a leucine-rich repeat class of plant disease resistance gene. *Science* **265**, 1856-60.
7. Bittner-Eddy P, Can C, Gunn N, Pinel M and Tor M (1999) Genetic and physical mapping of the *RPP13* locus in Arabidopsis, responsible for specific recognition of several *Peronospora parasitica* (downy mildew) isolates. *Mol Plant Microb Interac* **12**, 792-802.
8. *Michelmore RW, Paran I and Kessli 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****, 9829-32.*
9. *Tanksley SD, Ganal MW and Martin GB (1995) Chromosome landing: a paradigm for map-based gene cloning in plants with large genomes.* Trends Genet ***11****, 63-8.*
10. Thomas CM, Vos P, Zabeau M, Jones DA and Norcott KA (1995) Identification of amplified restriction length polymorphism (AFLP) markers tightly linked to the tomato *Cf-9* gene for resistance to *Cladosporium fulvum*. *Plant J* **8**, 785-94.
11. Lisitsyn N (1995) Representational difference analysis: finding the differences between genomes. *Trends Genet* **11**, 303-7.
12. Rosemberg M, Przybylska M and Straus D (1994) "RFLP Subtraction": A method for making libraries of polymorphic markers. *Proc Natl Acad Sci USA* **91**, 6113-7.
13. Corrette-Bennett J, Rosenberg M, Przybylska M, Ananiev E and Straus D (1998) Positional cloning without a genome map using 'Targeted RFLP Subtraction' to isolate dense markers tightly linked to the *regA* locus of *Volvox carteri*. *Nucleic Acids Res* **26**, 1812-8.
14. Xuelin T, Vanavichit A, Amornsilpa S and Tragoonrung S (1998) Mapping of rice *Rf* gene by bulked line segregation. *DNA Research* **5**, 15-8.
15. Buttery RG, Ling LC, Juliano BO and Turnbaugh (1983) Cooked rice aroma and 2-acetyl-1-pyrroline. *J Agric Food Chem* **31**, 821-6.
16. Sood BC and Siddiq EA (1978) A rapid technique for scent determination in rice. *Indian J Genet Plant Breed* **38**, 268-71.
17. Mahatheeranont S, Promdang S and Chaimpiriyakul A (1995) Volatile aroma compounds of Khao Dawk Mali 105. *Kasetsart J (Nat Sci)* **29**, 508-14.
18. Lorieux M., Petrov M, Huand N, Guiderdoni E and Ghesquire A (1996) Aroma in rice: genetic analysis of a quantitative trait. *Theor Appl Genet* **93**, 1145-51.
19. Ahn SN, Bollich CN and Tanksley SD (1992) RFLP tagging of a gene for aroma in rice. *Theor Appl Genet* **84**, 825-8.
20. Garland S., Lewin L, Blakeney A, Reinke R and Henry R (2000) PCR-based molecular markers for the fragrance gene in rice (*Oryza sativa. L.*). *Theor Appl Genet* **101**, 364-71.

21. Vollrath D (1999) DNA markers for physical mapping. In: *Genome analysis A Laboratory Manual Vol 4 mapping genomes,* pp. 187-215. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.
22. Jain N, Saini N, Rana P, Jain S and Jain RK (2002) Microsatellite diversity for chromosome number 8 in Basmati rice. *Rice Genet Newsl* **19**, 103-6.
23. Mahatheeranont S, Keawsa-ard S and Dumri K (2001) Quantification of the rice aroma compound, 2-acetyl-1-pyrroline, in un-cooked Khao Dawk Mali 105 brown rice. *J Agric Food Chem* **49**, 773-9.
24. Rogers SO and Bendich AJ (1994) Extraction of total cellular DNA from plants, algae and fungi. *Plant Mol Biol Manual* **D2**, 1-8.
25. Toojinda T, Siangliw M, Tragoonrung S and Vanavichit A (2003) Molecular Genetics of Submergence Tolerance in Rice: QTL Analysis of Key Traits. *Ann Bot* **91**, 243-53.
26. Wu J, Maehara T, Shimokawa T, Yamamoto S, Harada C, Takazaki Y, Ono N, *et al*. (2002) A comprehensive rice transcript map containing 6591 expressed sequence tag sites, *Plant Cell* **14** (3), 525-35.
27. Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE and Newburg L (1987) MAPMAKER: an interactive computer package for constructing genetic linkage maps of experimental and natural populations. *Genomics* **1**, 174-81.
28. Kosambi DD (1944) The estimation of map distance from recombination values. *Ann Eugen* **12**, 172-5.
29. Corderio GM, Christopher MJ, Henry RJ and Reinke RF (2002) Identification of microsatellite markers for fragrance in rice by analysis of the rice genome sequence. *Mol Breed* **9**: 245-250.
30. Jin Q, Waters D, Corderio GM, Henry RJ and Reinke RF (2003) A single nucleotide polymorphism (SNP) marker linked to the fragrance gene in rice (*Oryza sativa L.*). *Plant Sci* **165**, 359-64.