Chapter 2

Literature reviews

Rice (*Oryza sativa* L.) is a semiaquatic, annual grass that probably evolved along the foothills of the Himalayas and associated mountain ranges in South and Southeast Asia. A second cultivated species, *Oryza glaberrima* Steud., of minor importance, probably originated in the Niger River delta of West Africa (Lu and Chang, 1978). Rice is now grown on about 137 million hectares in more than 100 countries, mostly in the tropics and subtropics and predominantly in Asia. Several million hectares are cultivated in temperate zones up to a limit of 53 ^{o} Latitude. Rice is the primary staple food of more than 1.3 billion people and it is an important part of the diet of more than half of the word's population.

Rice can be grown in a wide range of soils. It grows in both flooded and upland conditions, but moisture should be at least near field capacity. Flooding the soil normally minimizes nutritional problems. Xuan and Ross (1976) discussed appropriate production practices.

Rice cultivars vary widely in sensitivity to photoperiod and temperature. The optimum temperature for growth is about 27 °C, but the optimum temperature to promote flowering depends on the photoperiod. Rice is a short-day plant; day length of about 10 hours is usually optimal for flowering. Longer days may progressively delay or inhibit flowering, depending on the sensitivity of the cultivar. Highly sensitive types will never flower if grown continuously under long day length. The critical photoperiod for sensitive cultivars varies from 12 to 14 hours and depends on the latitude of origin (Vergara and Chang, 1976).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Rice is self- pollination crop, although some outcrossing (usually less than 1 percent) may occur (Beachell *et al*., 1938). Meiosis occurs during the boot stage around the time when the auricles of the flag leaf are opposite the auricles of the next lower leaf. This stage is followed by emergence of the panicle which bears perfect flowers in single-flowered spikelets. A flower consists of six stamens, each composed of a two-lobed, four-loculed anther borne on a slender filament, and a pistil containing one ovule. The short style bears a feathery stigma fully receptive at the time that pollen sheds. At this stage, the lodicules become turgid and force the lemma and palea apart. Anther dehiscence and extrusion occur more or less simultaneously after which the lemma and palea close. This stage is usually reached in the uppermost florets when the panicle is about 50 to 60 percent emerged from the boot, but may occur earlier in certain cultivars. This is proper time to attempt artificial hybridization.

As anthesis commences in the morning after emasculation, a blooming panicle of the male parent is cut and carried to the emasculated panicle. If not pollinated, the stigma remains receptive for 4 or 5 days, but receptivity drops at an exponential rate after the first day. Pollen normally remains viable for a very short time after anther dehiscence (Jodan, 1938), probably less than 5 minutes under most conditions. However, Noguti and Hamada (1927), Noguchi (1931) and Chadraratna (1964) reported viability under non-specified condition for as long as 50 hours. Temperatures are perhaps other environmental factors which determine the length of time from pollination to fertilization, although Akemind (1914) and Terada (1928) reported that the time is generally about 12 hours. Cho (1955) found that fertilization was completed within 3 hours after pollination.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

The time of anthesis varies among locations and, to a lesser extent, among cultivars. Pollination should only be attempted during the period of peak anther dehiscence, usually about 2 hours each day. This peak usually occurs in mid to late morning in the tropics and around noon in temperate areas.

Natural cross-pollination in cultivated rice were in rate 9 (Beachell *et al*., 1938), normally less than 1 percent (Virmani and Athwal, 1973). Male sterility has been reported in rice (Shinijyo and Omura, 1966; Erickson, 1969; Athawal and Virmani, 1972; Li, 1977), and seed set on male-sterile plants is poor, even when they are surrounded by good pollinators (Stansel and Craigmiles, 1966; Athawal and Virmani, 1972; Carnahan *et al.*, 1972). Virmani and Athwal (1973, 1974) investigated the floral characters that influence out-crossing in rice, but systems for natural hybridization remain unsatisfactory for practical purposes.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.1 Hybrid rice

Hybrid rice is produced by crossing two inbred genetically–fixed varieties. Hybrids are special because they express what is called heterosis or hybrid vigor. The idea is that if cross two parents, which are genetically distant from each other, the offspring will be superior, particularly in terms of yield. However, the so-called heterosis effect always disappears after the first (F_1) generation. Hybrid vigor or heterosis is a universal phenomenon that occurs in all biological systems. In plant breeding, hybrid vigor denotes the expression of increased vigor over the better parent and commercial variety. Heterosis in crops other than rice, especially maize, has been exploited for commercial production. China began exploiting heterosis in rice in 1976 after Yuan Long Ping and his team developed the first set of stable and high-yielding three-line hybrid rice varieties. Since then, the area planted to hybrid rice in China has rapidly increased up to 17 million hectares in 1995. Hybrid varieties out-yield the commercial check varieties by 15 to 30 percent in farmers' fields. For example, the average yield of rice planted in Sichuan Province in 1989 where about 2.62 million hectares were planted to hybrid rice, was about 8.25 tonnes per hectare (Lin and Pingali, 1994). The rice team of the Rice Research Station at Beaumont in Texas, United States reported that the highest yield of 13,800 kilograms per hectare was obtained from a hybrid rice variety over a three-year testing period. The high productivity of hybrid rice has enabled China to reduce its cultivation area from about 34.4 million hectares in 1978 to about 31.9 million hectares in 1988 and, at the same time, increased its rice production from 136.9 million to 169.1 million tonnes during the same period (FAOSTAT data, 1997). This reduction of area planted to rice not only promoted the diversification in rice-based production systems but also increased

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

incomes and mitigates production risks, and also helped minimize the country's emission of greenhouse gases such as methane and nitrous oxide. In addition, the exploitation of hybrid rice technology has increased rice farmers' incomes and created rural employment opportunities, since its yield is high and its seed production labor-intensive. The exploitation of hybrid rice technology, therefore, not only improves rice production but also reduces rural poverty at the same time as being environmentally-friendly. Outside China, the hybrid rice area was about 11,000 hectares in 1992, increasing to 34,000 hectares in 1993 and 102,000 in 1996 in Vietnam with an average yield of 6.5 tonnes per hectare (15 to 30 percent higher than the best commercial varieties). Hybrid varieties are mostly planted in northern Vietnam where they have helped some provinces such as Thai Binh, Nam Dinh produce a surplus for the first time. In India, farmers cultivated about 65,000 hectares of hybrid rice in 1996. Limited commercial cultivation of hybrid rice was also reported in Bangladesh, the People's Democratic Republic of Korea and Myanmar (Tran and Nguyen, 1998).

China is the first country to produce hybrid rice commercially. Hybrid rice research was initiated in 1964 (Yuan, 1966) and the genetic tools essential for breeding hybrid rice varieties, such as the male sterile line (A line), the maintainer line (B line) and restorer line (R line), were developed by 1973 (Yuan and Virmani, 1988). Several hybrid combinations with good heterosis and higher yield potential were identified in 1974. Hybrid seed production techniques were basically established in 1975. The first batch of rice hybrid varieties was released commercially in 1976. Since then, the area under hybrid rice has been increasing years after years (Yuan, 1998). F_1 hybrid rice cultivars were said to be cultivated on a large scale in the

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Peoples Republic of China (Lin, 1977). Cytoplasmic-genetic male sterility was employed in hybrid seed production, but natural crossing was supplemented by hand pollination and augmented by cutting the flag leaves and tearing the leaf sheath that encloses the panicle.

Hybrid rice is successfully cultivated in China and gives more than 30 percent yield advantage over conventional pure line rice varieties to the rice farmers. This technology has enabled China to increase its rice production by nearly 200 million tonnes from 1976 to 1991 (Yuan, 1994). In 1979, the International Rice Research Institute (IRRI) began research on hybrid rice. Subsequently, several Asian countries such as Bangladesh, India, Indonesia, Malaysia, Myanmar, Philippines, South Korea, Sri Lanka, Thailand and Vietnam started their own hybrid rice program in collaboration with IRRI. Several experimental hybrids out-yielded the best inbred checks by a margin of 17 percent in replicated trials conducted at IRRI from 1986 to 1992 (Virmani, 1994). In recent years, India, Vietnam and the Philippines have released commercial rice hybrids giving at least 1 tonnes per hectare higher yield than the inbred checks. In 1998, 180,000 hectares in Vietnam, 150,000 hectares in India and 500 hectares in the Philippines were planted to hybrid rice. By 2005 A.D., India plans to cover 2 million hectares and Vietnam 0.5 million hectares with hybrid rice varieties. In addition, Bangladesh, Sri Lanka and Indonesia may also start commercializing this technology during the next four years.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.2 Male sterile system

Rice hybrid yields are about 20 percent higher than the inbred rice varieties. Commercial exploitation of heterosis has been successfully demonstrated in China where nearly 18 million hectares of a total of 33 million hectares of harvested rice land were planted with F_1 hybrid by 1992 (Maclean *et al*., 2002). Seed production for hybrid rice needs to be based on male sterility system. Cytoplasmic male sterility (CMS) system is the most common system used in rice hybrid seed production, but this system is expensive and cumbersome. The thermo-sensitive genic male sterile (TGMS) system is considered to be an efficient alternative to the CMS system for hybrid seed production (Maruyama *et al*., 1991) because it can greatly simplify the procedure of hybrid rice production compared with the CMS system. TGMS lines serve not only as male sterile lines but also as maintainer lines. Additionally, TGMS system eliminates the risk of genetic vulnerability appeared in CMS system, since this trait is relatively easy to transfer to diverse genetic background.

CMS system, also known as three-line system, is presently the most widely used system for producing F_1 rice hybrids. Although effective yet it is cumbersome because CMS lines require specific maintainer and restorer lines. Currently, most of the commercially-usable CMS lines possess wild abortive (WA) cytoplasm-induced male sterility. This situation makes CMS-based rice hybrids potentially vulnerable to biological stresses. On the other hand, breeding of japonica rice hybrids by the CMS system is constrained by the fact that japonica rice cultivars do not carry restorer genes. Transfer of these genes from indica rice cultivars to japonica is a long and cumbersome process. Similar problems exist with basmati rice. Since the TGMS system does not require restorer lines, it might be extremely useful for breeding

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

japonica and basmati rice hybrids. TGMS was developed in China (Yuan, 1987), Japan (Maruyama *et al.*, 1991) and IRRI (Voc and Virmani, 1991) and is controlled by a single recessive nuclear gene which interacts with temperature to express pollen sterility or fertility (Maruyama *et al*., 1991, Borkakati and Virmani, 1995). Seeds of a TGMS line are multiplied by selfing when they are exposed to the right temperature during their critical growth stage which in the tropics is 5 – 15 days after panicle initiation (Maruyama *et al*., 1991). Commercial F_1 hybrids are produced by pollinating a TGMS line with any fertile line or variety. This hybrid rice breeding system is called the two-line method (Yuan, 1987) which increases breeding efficiency. This study is aimed to present the results on developing TGMS lines and reports performance of some two-line rice hybrids.

There are three systems in hybrid rice production including three-line system or CMS system, two-line system and one-line system or apomixes (Yuan and Fu, 1995) At present, only three-line system and two-line system are used in hybrid rice production. Two - line system or environmental genetic male-sterility (EGMS) is grouped into three sensitive types: they are TGMS and photoperiod-sensitivity genic male-sterility (PGMS) that shows male sterility due to temperature and photoperiod respectively (Ku *et al*., 2001). Last type is thermo/photoperiod-sensitivity genic male-sterility (T/PGMS) that shows male sterility due to temperature and photoperiod (Singh and Virmani, 1990; Wu *et al*., 2003). This research will emphasize particularly on TGMS.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.3 Inheritance of TGMS and tagging the TGMS gene

There are several studies on inheritance of TGMS characteristic. It was reported that one recessive gene controlled TGMS characteristic. By studying in F_2 population, fertile and sterile ratio was 3:1 (Lang *et al*., 1997, Dong *et al*., 2000, Jia *et al*., 2001, Lopez *et al*., 2003). Molecular markers linked with TGMS characteristic which were studied by many researchers.

Wang *et al.* (1995) studied in bulked segregant analysis of F_2 population to identify RAPD markers linking to the TGMS gene of the 400 RAPD primers and screened for polymorphisms, five produced polymorphic products. Unfortunately, most of them were repetitive sequences, only two single-copy sequences were found. One was a 0.5 kb fragment amplified by OPT-02, another was a 1.2 kb fragment amplified by OPB-19. They were named as TGMS O.5 and TGMS 1.2, respectively. These fragments were cloned into pUC18 and mapped on chromosome 8 with a double haploid (DH) line population and recombinant inbred line (RIL) mapping population. TGMSO.5 was 12.9 cM from RZ562, while TGMS 1.2 was 12.6 cM from RZ562 on the other side. In order to confirm the chromosome location of TGMS 1.2 and TGMS O.5, two RFLP markers located in this region on chromosome 8, RZ1.2 and RG978, were selected as probes to determine their linkage relation to TGMS gene by using an F_2 population segregating for TGMS character. Results showed that RZ562 was linked to TGMS gene but not RG978, suggesting that TGMS gene was probably located between TGMS 1.2 and TGMS O.5. Because TGMS 1.2 and TGMS 0.5 are linked to TGMS gene, therefore, TGMS gene is most likely to be located on chromosome 8. This is the first report on mapping rice TGMS gene.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright © by Chiang Mai University
All rights reserved

Jia *et al*. (2001) studied reverse thermo-sensitive genic male-sterile line, J207S, that had an opposite phenotype compared to the normal TGMS lines. J207S was completely sterile when the temperature was lower than 31 °C. Thus, it could be widely used in larger areas. Genetic analysis indicated that the sterility of J207S was controlled by a single recessive gene which was first named as *rtms1*. An F_2 population from the cross between J207S and E921 was developed and used for molecular mapping of the *rtms1* gene. The AFLP (amplified fragment length polymorphism) technique, combined with BSA (bulked segregant analysis), was used to screen markers linked to the target gene, and eight polymorphic AFLP loci were identified. Co-segregating analysis using the F_2 population showed that two of them, Rev1 and Rev7, were closely linked to the target gene with a recombinant rate of 3.8 percent and 7.7 percent, respectively. Both Rev1 and Rev7 were found to be single-copy sequences through Southern blotch analysis. Rev1 was subsequently mapped on chromosome 10 with a doubled-haploid mapping populations derived from the cross CT9993 x IR62266 which was available at Texas Technology University. RM222 and RG257 were linked to Rev1 at a distance of 11.8 cM and 4.6 cM, respectively. Additional SSR markers from the rice map of Cornell University, RFLP markers from the map of RGP in Japan and the map of Texas Tech University were selected from the region surrounding Rev1 on chromosome 10 to conduct the fine-mapping of the *rtms1* gene. Presently, *rtms1* was mapped between RM239 and RG257 with genetic distance of 3.6 cM and 4.0 cM, respectively. The most-closely linked AFLP marker, Rev1, 4.2 cM from the *rtms1* gene, was sequenced and converted into a SCAR (sequence characterized amplified region) marker which could facilitate marker-assisted selection of the *rtms1* gene.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Lopez *et al*. (2003) developed TGMS lines with aromatic Thai rice background by using molecular marker-aided breeding. Four microsatellite markers (RM2, RM10, RM11 and RM214) on chromosome 7 in the vicinity of the TGMS gene tms2 and showing polymorphisms between the two parents were used in genotyping the mapping population consisting of 157 F_2 plants derived from a cross between Norin PL12 (a TGMS line from Japan) and KDML 105 (a popular aromatic Thai rice cultivar). The RM11 marker was approximately 5 cM from tms2 while RM2 was approximately 16 cM from it. In this F_2 population, the accuracy of selecting sterile plants with RM2 and RM11 markers was 92 and 97 percent, respectively. In three backcrosses, the accuracy of selection with markers for either homozygous or heterozygous plants was higher than 90 percent with RM2. Using RM11, 89 percent accuracy was obtained for selecting homozygous fertile plants and 59 percent accuracy for selecting heterozygous plants. The results demonstrated that microsatellite markers were powerful in screening large breeding populations, and these markers facilitated selection for plants possessing the tms2 in an early stage of the crop and without exposing the materials to the required temperature for TGMS gene expression. Three TGMS lines with aromatic Thai rice background were developed and showed complete pollen sterility when maximum temperature was higher than 30 °C, 1 to 2 weeks after panicle initiation. Up to 77 percent spikelet fertility was observed when these lines were exposed to temperature below 30 °C during the critical stage.

Development of simple and reliable PCR-based markers is an important component of marker-assisted selection (MAS) activities for agronomically-important genes in rice breeding. In order to develop PCR-based markers for a rice thermo-

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

sensitive genic male sterility gene tms3(t), located on chromosome 6, the nucleotide sequences of four linked RAPD markers OPF182600, OPAC3640, OPB19750 and OPM7550 were used to design and synthesize several pairs of specific primers for PCR amplification of the genomic DNA of both the parents IR32364TGMS (sterile) and IR68 (fertile), involved in mapping this gene. For the RAPD marker OPF 182600, two pairs of specific primer pair combination from different positions of the sequence resulted in generation of two co-dominant STS (Sequence Tagged Sites) markers. In case of markers OPAC3640, OPB19750 and OPAA7550, the first two could generate dominant polymorphisms while the last one could not be successful in PCR amplification. Both the co-dominant STSs with primer combinations F18F/F18RM and F18FM/F18RM were found to be tightly linked to the tms3(t) gene with a genetic distance of 2.7 cM. The sizes of the different alleles in case of F18F/F18RM, F18FM/F18RM combinations were 2300 bp, 1050 bp and 1900 bp, 1000 bp respectively. The efficiency of marker-assisted selection for this trait was estimated as 84.6 percent. Polymorphism survey of 12 elite rice lines indicated that these PCR-based markers for tms3(t) could be used in selecting TGMS plants at seedling stage in the segregating populations in environment independence of controlled temperature regime (Lang *et al*., 1999).

The fertility transformation behavior, the critical fertility and sterility temperatures and the mode of inheritance of male sterility were studied for a new TGMS line, TS6, which was identified at Tamil Nadu Agricultural University, Coimbatore, India. The pollen and spikelet fertilities recorded on plants raised at fortnightly intervals revealed that this line was completely sterile for 78 consecutive days (35/22 to 32/23 °C, maximum/minimum temperatures) and reverted to fertile

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

when the temperature was 30/18 °C. It remained fertile continuously for 69 days and the maximum pollen and spikelet fertilities recorded were 75 and 70 percent, respectively. The fertility was highly influenced by daily maximum temperature, followed by average and minimum temperatures. It was not influenced by relative humidity, sunshine hours or photoperiod. The critical temperature inducing sterility and fertility was 26.7 and 25.5 °C, respectively. The male sterility in TS6 was inherited as a monogenic recessive in the F_2 and BC_1 populations of TS6 MRST9 as well as TS6 IR68281B. Using bulked segregant analysis on an F_2 population of TS6 MRST9, an RAPD marker, OPC052962, was identified to be associated with TGMS in TS6 (Latha *et al*., 2004).

F_2 population developed from a cross between a TGMS indica mutant, TGMS–VN1, and a fertile indica line, CH1, was used to identify molecular markers linked to the TGMS gene and to subsequently determine its chromosomal location on the linkage map of rice. Bulk segregant analysis was performed using the AFLP technique. From the survey of 200 AFLP primer combinations, four AFLP markers (E2/M5–600, E3/M16–400, E5/M12–600 and E5/M12–200) linked to the TGMS gene were identified. All the markers were linked to the gene in the coupling phase. All except E2/M5–200 were found to be low-copy sequences. However, the marker E5/M12–600 showed polymorphisms in RFLP analysis and was closely linked to the TGMS gene at a distance of 3.3 cM. This marker was subsequently mapped on chromosome 2, using doubled-haploid mapping populations derived from the crosses IR64 × Azucena and CT9993 × IR62666, available at IRRI, Philippines, and Texas Technology University, respectively. Linkage of microsatellite marker RM27 with the TGMS gene further confirmed its location on chromosome 2. The closest marker,

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

E5/M12–600, was sequenced so that a PCR marker could be developed for the marker-assisted transfer of this gene to different genetic backgrounds. The new TGMS gene was tentatively designated as tms4(t) (Dong *et al*., 1999).

AnnongS-1, a TGMS rice line, has a new TGMS gene. Genetic analysis indicated that a single recessive gene named tms5 controlled the sterility of AnnongS-1. In our previous studies based on an F_2 population from the cross between AnnongS-1 and Nanjing11, tms5 was mapped on chromosome 2. Recently, a RIL (recombinant inbred line) population from the same cross was developed and used for the fine mapping of the tms5 gene. Molecular marker techniques combined with BSA (bulked segregant analysis) were used. As a result, two AFLP markers (AF10, AF8), one RAPD marker (RA4), one STS marker (C365-1), one cleaved amplified polymorphic sequence (CAPs) marker (G227-1) and four SSR markers (RM279, RM492, RM327, RM324) were found to be closely linked to tms5 gene. The DNA sequences of the RFLP marker of C365 and G227 were found in Gene Bank, and on the basis of these sequences, many primers were designed to amplify the two parents and their RIL population plants. Finally, the tms5 gene was mapped between STS marker C365-1 and CAPs marker G227-1 at a distance of 1.04 cM from C365-1 and 2.08 cM from G227-1 (Wang *et al*., 2003).

The construction of a molecular genetic linkage map is considered to be the most important aspect in plant map-based gene cloning. In this study, F_2 population from the cross between a TGMS mutant line (AnnongS-1) and Nanjing11 was used to construct a genetic linkage map of rice. Total of 142 AFLP markers, 25 SSR markers and 4 RFLP markers were assigned to 12 linkage groups covering 1537.4 cM of rice genome. The average interval between these markers was 9.0 cM. The length of each

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

chromosome was from 64.2 cM to 180.2 cM. Of the 173 AFLP markers which were used to construct the molecular map, 17 markers (9.8 percent) showed segregation distortion. Most of the distortion occurred on chromosome 2. The construction of this genetic linkage map would be helpful for mapping genes of important agronomic traits. The discovery and application of TGMS made a breakthrough for whole rice production. By using TGMS rice, three-line breeding system has been changed into two-line system. More and more TGMS genes were discovered and applied to hybrid rice production which broaden the genetic bases of hybrids and provide breeders with more choice to develop two-line hybrid rice with desirable traits. In this study, one parent of the F_2 population for construction of the genetic linkage map is AnnongS-1, a TGMS rice line with a new TGMS gene (tms5). At the same time of the framework map construction, combing with BSA analysis, AFLP markers closely linked to the tms5 gene were identified and mapped. Then, RFLP and SSR markers were employed to determine and confirm the locations of AFLP markers and tms5 gene in the linkage map. Thus tms5 was mapped on the short arm of chromosome 2. Then, 33 RFLP markers from Japanese RGP map and 19 SSR markers from the map of Cornell University were used to construct the fine map of tms5 gene-encompassing region on chromosome 2. As a result, tms5 was mapped between R394 and RM71, at the distance of 2.5 cM from R394 and 0.0 cM from RM174. At the same time, two AFLP markers (AACAG1 and AGCTG12), flanking tms5 gene, were linked to tms5 at the distance of 2.4 cM and 6.6 cM, respectively. Those markers closely linked to tms5 gene were useful for the target gene cloning with map-based cloning strategy (Jia *et al*., 2003).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

One TGMS gene was investigated by a spontaneous japonica TGMS line, Sokcho ms1, which was completely sterile when the temperature was higher than 27 °C and lower than 25 °C during the differentiation stage of the spikelets, but fertile when the temperature was between 25-27 °C. Genetic analysis and molecular mapping based on simple sequence-repeat (SSR), sequence tagged site (STS) and cleaved amplified polymorphic sequence (CAPS) maps revealed that a single recessive gene locus involved in the control of genic male sterility in Sokcho ms1. A new TGMS gene, tms6(t), was mapped primarily to the long arm of chromosome 5 of *Oryza sativa* at the interval between markers RM335 (0.1 cM) and E60663 (2.0 cM) by using an F_2 population derived from a cross between Sokcho ms1 and a fertile indica variety, Neda. It was found that tms6(t) appeared to differ from the other mapped TGMS genes in rice (Suh *et al*., 2004).

2.4 DNA extraction

A rapid and efficient method for isolating DNA from plants is necessary when hundreds of samples need to be analyzed rapidly, such as in genome mapping and marker-assisted selection programs. However, the preparation of high-quality DNA from plants can be time-consuming, laborious and expensive because of the multiple manipulation steps (Mc Couch *et al*., 1988, Steenkamp *et al*., 1994, Aljanabi and Martinez, 1997). To overcome this problem, several DNA extraction protocols for temperate crops have been described, based on the use of 96-well micro titer plates (e.g., Dilworth and Frey, 2000, Paris and Carter, 2000, Mace *et al*., 2003) and the increasing number of relatively expensive commercial products that are now available

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

(e.g., DNeasy 96 Plant Kit [QIAGEN], Wizard Magnetic 96 DNA Plant System [Promega]). However, a bottleneck still exists.

The DNA extraction followed Johns *et al*. (1997) protocol with the minor modification in the method for grinding tissue samples. DNA was extracted from 0.1 to 0.2 g of fresh sample tissue. The tissue was ground with 500 µL potassium ethyl xanthogenate modified extraction buffer in Fast Prep FP120 machine (BIO 101 Inc., Carlsbad, CA) using a ceramic bead. The samples were then heated in a 65 °C water bath for 30 minutes. The samples were centrifuged at 8161 g for 10 minutes and the supernatants were transferred to a clean 1.5-mL micro-centrifuge tube. Nucleic acids were precipitated by means of 6:1 95 % (v/v) ethanol and 7.5 mole ammonium acetate for 30 minutes at room temperature. The nucleic acids were then pelleted at 1306 g for 10 minutes. Tris EDTA + RNase was added to the pellet and incubated at 37 °C for 1 hour. Next, plant debris was pelleted, and the supernatant was transferred to a clean 1.5-mL tube and the DNA was precipitated from the supernatant with 10:1 95 % ethanol and 3 mole sodium acetate for 30 minutes at room temperature. The samples were spun at 2040 g for 5 minutes to pellet the DNA. The DNA was washed with ethanol and the DNA pellets were collected through centrifugation.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.5 Polymerase chain reaction (PCR)

PCR is a technique widely used in molecular biology. It derives its name from one of its key components, a DNA polymerase used to amplify (i.e., replicate) a piece of DNA by in vitro enzymatic replication. As PCR progresses, the DNA thus generated itself as template for replication. This sets in motion a chain reaction in which the DNA template is exponentially amplified. With PCR, it is possible to

amplify a single or few copies of a piece of DNA across several orders of magnitude, generating millions or more copies of the DNA piece. PCR can be performed without restrictions on the form of DNA, and it can be extensively modified to perform a wide array of genetic manipulations.

Almost all PCR applications employ a heat-stable DNA polymerase, such as Taq polymerase. This DNA polymerase enzymatically assembles a new DNA strand from DNA building blocks, the nucleotides, using single-stranded DNA as template and DNA oligonucleotides (also called DNA primers) required for initiation of DNA synthesis. The vast majority of PCR methods use thermal cycling, i.e., alternately heating and cooling the PCR sample to a defined series of temperature steps. These different temperature steps are necessary to bring about physical separation of the strands in a DNA double helix (DNA melting), and permit DNA synthesis by the DNA polymerase to selectively amplify the target DNA. The power and selectivity of PCR are primarily due to selecting primers that are highly complementary to the DNA region targeted for amplification, and to the thermal cycling conditions used.

PCR was developed in 1983 by Kary Mullis (Bartlett and Stirling, 2003). PCR is now a common and often indispensable technique used in medical and biological research labs for a variety of applications. These include DNA cloning for sequencing, DNA-based phylogeny, or functional analysis of genes; the diagnosis of hereditary diseases; the identification of genetic fingerprints (used in forensics and paternity testing); and the detection and diagnosis of infectious diseases. Mullis won the Nobel Prize in chemistry in 1993 for his work on PCR.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

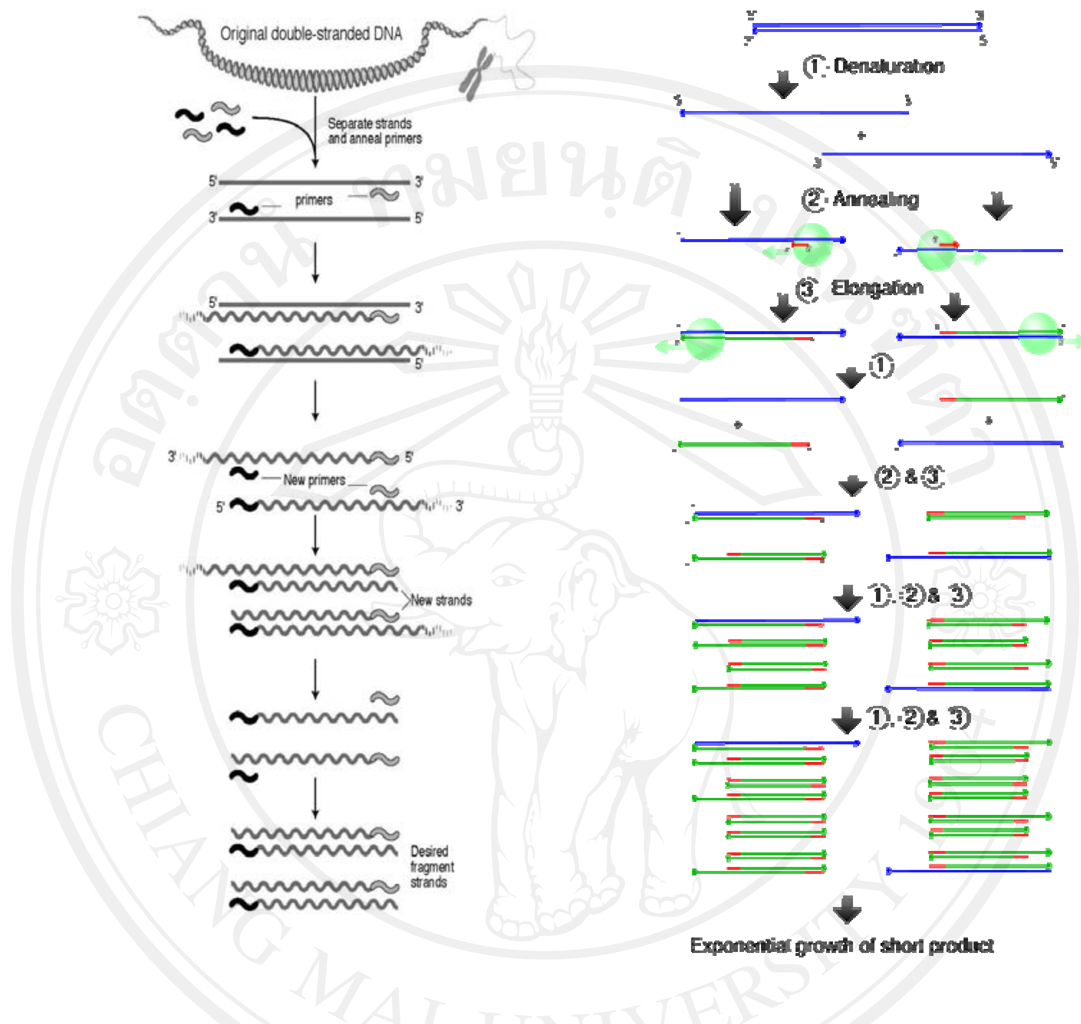

Figure 2.1 Schematic drawing of the PCR cycle. (1) Denaturing at 94-96 °C (2) Annealing at ~65°C (3) Elongation at 72°C Four cycles are shown here.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

PCR is used to amplify specific regions of a DNA strand (the DNA target). This can be a single gene, a part of a gene, or a non-coding sequence. Most PCR methods typically amplify DNA fragments of up to 10 kilo base pairs (kb), although some techniques allow for amplification of fragments up to 40 kb in size (Chang *et al.*, 1994)

The PCR usually consists of a series of 20 to 35 repeated temperature changes called cycles; each cycle typically consists of 2-3 discrete temperature steps. Most commonly, PCR is carried out with cycles that have three temperature steps (Fig. 2.1). The cycling is often preceded by a single temperature step (called hold) at a high temperature (>90 °C), and followed by one hold at the end for final product extension or brief storage. The temperatures used and the length of time they are applied in each cycle depend on a variety of parameters. These include the enzyme used for DNA synthesis, the concentration of divalent ions and dNTPs in the reaction, and the melting temperature (Tm) of the primers (Rychlik *et al.*, 1990).

Initialization step: This step consists of heating the reaction to a temperature of 94-96 °C (or 98 °C if extremely thermo-stable polymerases are used), which is held for 1-9 minutes. It is only required for DNA polymerases that require heat activation by hot-start PCR (Sharkey *et al*., 1994).

Denaturation step: This step is the first regular cycling event and consists of heating the reaction to 94-98 °C for 20-30 seconds. It causes melting of DNA template and primers by disrupting the hydrogen bonds between complementary bases of the DNA strands, yielding single strands of DNA.

Annealing step: The reaction temperature is lowered to 50-65 °C for 20-40 seconds, allowing annealing of the primers to the single-stranded DNA template. Typically, the annealing temperature is about 3-5 degrees Celsius below the Tm of the primers used. Stable DNA-DNA hydrogen bonds are only formed when the primer sequence very closely matches the template sequence. The polymerase binds to the primer-template hybrid and begins DNA synthesis.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Extension/elongation step: The temperature at this step depends on the DNA polymerase used; Taq polymerase has its optimum activity temperature at 75-80 °C, (Chien *et al*., 1976, Lawyer *et al*., 1993) and commonly a temperature of 72 °C is used with this enzyme. At this step the DNA polymerase synthesizes a new DNA strand complementary to the DNA template strand by adding dNTP's that are complementary to the template in 5' to 3' direction, condensing the 5'-phosphate group of the dNTPs with the 3'-hydroxyl group at the end of the nascent (extending) DNA strand. The extension time depends both on the DNA polymerase used and on the length of the DNA fragment to be amplified. As a rule-of-thumb, at its optimum temperature, the DNA polymerase will polymerize a thousand bases in one minute.

Final elongation: This single step is occasionally performed at a temperature of 70-74 °C for 5-15 minutes after the last PCR cycle to ensure that any remaining single-stranded DNA is fully extended.

Final hold: This step at 4-15 °C for an indefinite time may be employed for short-term storage of the reaction.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.6 Types of markers

2.6.1 Morphological markers

These are the traditional markers as mentioned before. Morphological mutant traits in a population which are mapped and linked to a desirable or undesirable trait are determined and indirect selection is carried out using the physically-identifiable mutant for the trait. There are several undesirable factors that are associated with morphological markers. The first is their high dependency on environmental factors. Often the conditions that a plant is grown in can influence the expression of these markers and lead to false determination. Second, these mutant traits often have undesirable features such as dwarfism or albinism. And lastly, performing breeding experiments with these markers is time-consuming, laborious, intensive and the large populations of plants are required, need large plots of land and/or greenhouse space in which to be grown (Stuber *et al.*, 1999).

2.6.2 Biochemical markers

Isozymes are used as biochemical markers in plant breeding. Isozymes are common enzymes expressed in the cells of plants. The enzymes are extracted, and run on denaturing electrophoresis gels. The denaturing component in the gels (usually SDS) unravels the secondary and tertiary structure of the enzymes and they are then separated on the basis of net charge and mass. Polymorphic differences occur on the amino acid level allowing singular peptide polymorphisms to be detected and utilized as a polymorphic biochemical marker.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

Biochemical markers are superior to morphological markers in that they are generally independent on environmental growth conditions. The only problem with isozymes in marker-assisted selection (MAS) is that most cultivars (commercial breeds of plants) are genetically very similar and isozymes do not produce a great amount of polymorphisms and polymorphisms in the protein primary structure may still cause an alteration in protein function or expression.

Isozyme electrophoresis methods were used to identify creeping bentgrass cultivars as reported by Warnke *et al*., (1997). However, Jones (1983) reported isozyme instability after storage, resulting in different banding patterns. Another draw- back of isozyme analysis was the limited number of markers available (Golembiewski *et al*., 1997).

2.6.3 Molecular markers

Molecular markers are based on naturally-occurring polymorphisms in DNA sequences (i.e., base pair deletions, substitutions, additions or patterns) (Gupta *et al.*, 1999). There are various methods to detect and amplify these polymorphisms so that they can be used for breeding analysis and these techniques will be the focus of this study. Molecular markers are superior to other forms of MAS because they are relatively simple to detect, abundant throughout the genome even in highly-bred cultivars, completely independent of environmental conditions and can be detected at virtually any stage of plant development.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

There are 5 conditions that characterize a suitable molecular marker (Gupta *et al*., 1999):

1) Must be polymorphics

2) Co-dominant inheritance

3) Randomly and frequently distributed throughout the genome

4) Easy and cheap to detect

5) Reproducible

Molecular markers can be also used for several different applications including; germplasm characterization, genetic diagnostics, characterization of transformants, study of genome organization and phylogenic analysis.

2.7 Molecular markers

All living organisms are made up of cells that are programmed by genetic material called DNA. This molecule is made up of a long chain of nitrogen-containing bases (there are four different bases – adenine [A], cytosine [C], guanine [G] and thymine [T]). Only a small fraction of the DNA sequence typically makes up genes, i.e., that code for proteins, while the remaining and major share of the DNA represents non-coding sequences, the role of which is not yet clearly understood. The genetic material is organized into sets of chromosomes (e.g., five pairs in *Arabidopsis thaliana*; 30 pairs in *Bos taurus* [cow]), and the entire set is called the genome. In a diploid individual (i.e., where chromosomes are organized in pairs), there are two alleles of every gene – one from each parent.

Molecular markers should not be considered as normal genes as they usually do not have any biological effect. Instead, they can be thought of as constant landmarks

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

in the genome. They are identifiable DNA sequences, found at specific locations of the genome, and transmitted by the standard laws of inheritance from one generation to the next. They rely on a DNA assay, in contrast to morphological markers that are based on visible traits, and biochemical markers that are based on proteins produced by genes.

Different kinds of molecular markers exist, such as restriction fragment length polymorphisms (RFLPs), random amplified polymorphic DNA (RAPDs) markers, amplified fragment length polymorphisms (AFLPs), simple sequence repeat (SSR) and single nucleotide polymorphisms (SNPs). They may differ in a variety of ways – such as their technical requirements (e.g., whether they can be automated or require use of radioactivity); the amount of time, money and labour needed, the number of genetic markers that can be detected throughout the genome, and the amount of genetic variation found at each marker in a given population. The information provided to the breeder by the markers varies, depending on the type of marker system used. Each has its advantages and disadvantages and, in the future, other systems are likely to be developed.

Molecular markers are valuable tools in both basic and applied research, such as fingerprinting genotypes, analyzing genetic diversity, determining variety identity, marker-assisted breeding, phylogenetic analysis and map-based cloning of genes (McCouch *et al.*, 1997; Joshi *et al.*, 2001; Nagaraju *et al.*, 2002; Ni *et al.*, 2002). In rice (*Oryza sativa*), more than 10,000 molecular markers have been developed (Saito *et al.*, 1991; Causse *et al.*, 1994; Kurata *et al.*, 1994; Harushima *et al.*, 1998; Wu *et al.*, 2002). Most of them are expressed sequence tags and RFLPs. In addition to

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

expressed sequence tags and RFLPs, SSR markers and SNP markers have been developed in recent years.

2.7.1 Restriction fragment length polymorphic (RFLP) markers

A restriction fragment length polymorphism (or RFLP, often pronounced as "rif-lip") is a variation in the DNA sequence of a genome which can be detected by a laboratory technique known as gel electrophoresis. Analysis of RFLP variation is an important tool in genome mapping, localization of genetic disease genes, determination of risk for a disease, genetic fingerprinting and paternity testing.

The basic technique for detecting RFLPs involves the fragmentation of genomic DNA by a restriction enzyme, which can recognize and cut DNA wherever a specific short sequence occurs, in a process known as a restriction digest. The resulting DNA fragments are then separated by length through a process known as agarose gel electrophoresis, and transferred to a membrane via the Southern blot procedure. Hybridization of the membrane to a labeled DNA probe then determines the size of the fragments which are complementary to the probe. An RFLP occurs when the size of a detected fragment varies between individuals. Each fragment size is considered an allele, and can be used in genetic analysis

RFLP markers have been used to distinguish some creeping bentgrass cultivars, but only a limited number of cultivar-specific markers have been identified (Caceres *et al*., 2000).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.7.2 Random amplified polymorphic DNA (RAPDs) markers

RAPD stands for random amplification of polymorphic DNA. It is a type of PCR reaction, but the segments of DNA that are amplified are random. The scientists performing RAPD create several arbitrary, short primers (8-12 nucleotides), then proceed with the PCR using a large template of genomic DNA, hoping that fragments will amplify. By resolving the resulting patterns, a semi-unique profile can be gleaned from a RAPD reaction.

No knowledge of the DNA sequence for the targeted genes is required, as the primers will bind somewhere in the sequence, but it is not certain exactly where. This makes the method popular for comparing the DNA of biological systems that have not had the attention of the scientific community, or in a system in which relatively few DNA sequences are compared (it is not suitable for forming a DNA databank). Due to the fact that it relies on a large, intact DNA template sequence, it has some limitations in the use of degraded DNA samples. Its resolving power is much lower than targeted, species specific DNA comparison methods, such as short tandem repeats. In recent years, RAPD is used to characterize, and trace, the phylogeny of diverse plant and animal species.

On the basis of PCR, RAPD has the advantage of being rapid, cost-effective, and has been used for identification of crops such as buffalograss (*Buchloe dactyloides*, Nutt.), Kentucky bluegrass (*Poa pratensis* L.) and bentgrass (*Agrostis capillaries*) (Huff *et al*., 1993; Huff, 2001; Golembiewski *et al*., 1997). However, the technique can be difficult to reproduce because of its high sensitivity to reaction conditions (Skroch and Nienhuis, 1995).

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.7.3 Simple sequence length polymorphism (SSLP) or Simple sequence repeat (SSR) or microsatellite markers

Microsatellites are simple, tandemly-repeated di-nucleotide to tetra-nucleotide sequence motifs flanked by unique sequences. They are valuable as genetic markers because they are co-dominant, detect high levels of allelic diversity, and are easily and economically assayed by the PCR. Results from screening a rice genomic library suggest that there are estimated 5,700-10,000 microsatellites in rice, with the relative frequency of different repeats decreasing with increasing size of the motifs. A map consisting of 120 microsatellite markers demonstrates that they are well distributed throughout the 12 chromosomes of rice. Five multiple copy primer sequences have been identified that could be mapped to independent chromosomal locations. The current level of genome coverage provided by these SSLPs in rice is sufficient to be useful for genotype identification, gene and quantitative trait locus (QTL) analysis, screening of large insert libraries, and marker-assisted selection in breeding. Studies of allelic diversity have documented up to 25 alleles at a single locus in cultivated rice germplasm and provided evidence that amplification in wild relatives of *Oryza sativa* is generally reliable. The availability of increasing numbers of mapped SSLP markers can be expected to complement existing RFLP and AFLP maps, increasing the power and resolution of genome analysis in rice (Susan *et al.*, 1997). Approximately 2,740 SSR markers have been genetically mapped in rice, about one SSR marker every 157 kb (Chen *et al.*, 1997; Temnykh *et al.*, 2000, 2001; McCouch *et al.*, 2002)

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.7.4 Sequence characterized amplified region markers (SCAR)

SCAR DNA analysis was developed to produce more specific and reproducible results (Paran and Michelmore, 1993; Jung *et al*., 1999). SCAR markers are created by using a longer primer (extended sequence of a RAPD primer) that has a specific sequence of approximately 20 bases. By increasing the specificity, the results are less sensitive to changes in reaction conditions and are more reproducible (Hernandez *et al.*, 1999). Reliable SCAR markers have already been successfully derived from RAPD fragments in Lettuca, Vicia and Triticum (Paran and Michelmore, 1993; Vidal *et al*., 2000; Hernandez *et al.*, 1999).

2.7.5 Single-strand conformation polymorphisms (SSCP)

SSCP is the electrophoretic separation of single-stranded nucleic acids, based on subtle differences in sequence (often a single base pair) which results in a different secondary structure and a measurable difference in mobility through a gel. The mobility of double-stranded DNA in gel electrophoresis is dependent on strand size and length but is relatively independent of the particular nucleotide sequence. The mobility of single strands, however, is noticeably affected by very small changes in sequence, possibly one changed nucleotide out of several hundred. Small changes are noticeable because of the relatively unstable nature of single-stranded DNA; in the absence of a complementary strand, the single strand may experience intrastrand base pairing, resulting in loops and folds that give the single strand a unique 3D structure, regardless of its length. A single nucleotide change could dramatically affect the strand's mobility through a gel by altering the intrastrand base pairing and its resulting 3D conformation (Melcher, 2000). Single-strand conformation polymorphisms

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

analysis takes advantage of this quality of single-stranded DNA. First announced in 1989 as a new means of detecting DNA polymorphisms, or sequence variations, SSCP analysis offers an inexpensive, convenient and sensitive method for determining genetic variation (Sunnucks *et al*., 2000). Like RFLPs, SSCPs are allelic variants of inherited, genetic traits that can be used as genetic markers. Unlike RFLP analysis, however, SSCP analysis can detect DNA polymorphisms and mutations at multiple places in DNA fragments (Orita *et al*., 1989). As a mutation scanning technique, though, SSCP is more often used to analyze the polymorphisms at single locus, especially when used for medical diagnoses (Sunnucks *et al*., 2000).

Single-stranded DNA mobilities are dependent on temperature. For best results, gel electrophoresis must be run in a constant temperature. Sensitivity of SSCP is affected by pH. Double-stranded DNA fragments are usually denatured by exposure to basic conditions: a high pH. Kukita *et al*. (1997) found that adding glycerol to the polyacrylamide gel lowers the pH of the electrophoresis buffer (more specifically, the Tris-borate buffer) and the result is increased SSCP sensitivity and clearer data. Fragment length also affects SSCP analysis. For optimal results, DNA fragment size should fall within the range of 150 to 300 bp, although SSCP analysis of RNA allows for a larger fragment size (Wagner, 2002). The presence of glycerol in the gel may also allow a larger DNA fragment size at acceptable sensitivity (Kukita *et al*., 1997). Under optimal conditions, approximately 80 to 90 percent of the potential base exchanges are detectable by SSCP (Wagner, 2002). If the specific nucleotide responsible for the mobility difference is known, a similar technique called single nucleotide polymorphism (SNP) may be applied.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved

2.7.6 Single nucleotide polymorphisms (SNPs)

One of the major objectives of genetics is the association of sequence variations with heritable phenotypes. Traditional strategies, such as linkage analysis, in which pedigree analysis tracks transmission of a disease through a family, have been successfully applied in the detection of Mendelian disorders. In recent years, a more powerful approach involving the detection of SNPs has become increasingly popular. By conventing, a nucleotide polymorphisms must be presented in at least one percent of the human population to be called an SNP. SNPs are the most common type of DNA sequence variations and occur once in every 100-300 bases.

Researchers are looking for associations between a disease and specific sequence differences in a population by using this high degree of variations. As it is easier to obtain DNA samples from a random set of individuals in a population than from every member of a family over several generations, it is conceivable that researchers may be able to increase the identification of genes responsible for pathological traits. Due to the high frequency of SNPs, the chances are high that the disease is predominantly caused by, or closely associated with, specific SNPs.

ลิขสิทธิ์มหาวิทยาลัยเชียงใหม่
Copyright© by Chiang Mai University
All rights reserved