MATERIALS AND METHODS

Isolation and characterization of microsatellite sequences

1. Plant Materials

For development of microsatellite markers in neem, young leaves samples from each of 24 populations were collected from FAO International Provenance Trials Germplasm bank of neem established in Kanchanaburi, Thailand in August, 1997. Seed sources are shown below in Table 3. The map of the area where the neem seeds were collected for provenance trials is shown in Figure 9. Figure 10 shows location of the International Provenance Trials established at Kanchanaburi, Thailand.

2. Methods

In order to develop microsatellite markers for neem, the method developed by Fischer and Bachmann (1998) was used and is summarized below as flow chart in Figure 11.

Table 3 Seed sources for the international provenance trial of neem established at Kanchanaburi in 1997

No.	Seed source name	No. of Samples	Latitude	Longitude
1	Ban Bo, Kanlasin, Thailand	15	16°17'N	103°35'E
2	Ban Nong Hoi, Kanchanaburi, Thailand	16	14°09'N	99°19'E
3	Sunyani, Ghana	20	07°21'N	02°21'W
4	Doi Tao, Chiang Mai, Thailand	18	17°57'N	98°41'E
5	Vientiane, Lao P.D.R	21	18°00'N	102°45'E
6	Tung Luang, Suratthani, Thailand	12	09°09'N	99°07'E
7	Ramannaguda, Orissa, India	18	19°05'N	83°49'E
8	Yezin, Myanmar	19	19°51'N	96°16'E
9	Ban Nong Rong, Kanchanaburi, Thailand	17	14°05'N	99°40'E
10	Myene, Myanmar	24	22°03'N	95°13'E
11	Sagar, Chanatoria Madhya Pradesh, India	20	21°51'N	78°45'E
12	Balharshah, Maharashtra, India	20	19°51'N	79°25'E
13	Ghaati Subramanya, Karnataka, India	21	13°22'N	77°34'E
14	Chitradurga, Karnataka, India	15	14°02'N	76°04'E
15	Mandore, Jodhure, India	12	26°18'N	73°01'E
16	Annur, Tamil Nadu, India	19	11°17'N	77°07'E
17	Allahabad Town, Uttar Pradesh, India	13	25°28'N	81°54'E
18	Khao Laung, Nakhon Sawan, Thailand	23	15°32'N	99°57'E
19	Lamahi, Nepal	22	27°52'N	82°31'E
20	Geta, Nepal	19	28°46'N	80°34'E
21	Tibbi Laran, Rahimyar Khan, Pakistan	22	28°24'N	70°18'E
22	Multan, Cantonment Area, Pakistan	12	30°11'N	71°29'E
23	Kuliyapitiya, Sri Lanka	13	07°08'N	80°00'E
24	Bandia, Senegal	22	14°30'N	17°02'E

Figure 9 Area where the neem seeds were collected for international provenance trials (●).

Figure 10 Location of the International Provenance Trials of neem established at Kanchanaburi in August 1997 (■).

Source: FAO International Neem Network (2000).

Flow Chart of Development Microsatellite Markers

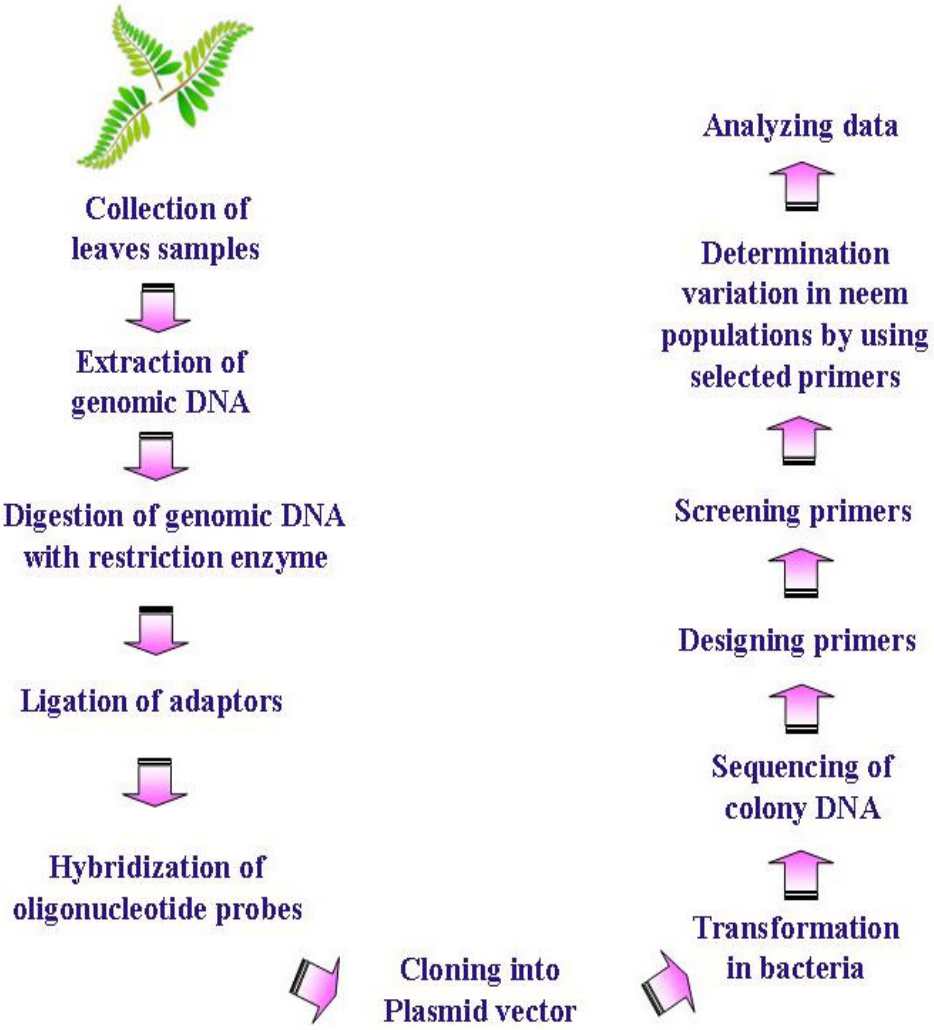

Figure 11 Flow chart for development microsatellite markers in neem.

Extraction of genomic DNA

Total genomic DNA was extracted by pooling 7 samples of young leaves of *A. indica* var. *indica* and *A. indica* var. *siamensis*, and which were then grounds to powder with liquid nitrogen following the procedure of DNeasy? Plant Mini Kit (Qiagen, Hilden) as described in Appendix 1. Very good quality DNA is required. DNA should be free from any contaminants, e.g. residual polysaccharides, proteins, RNA, remnants of phenol, chloroform, salt, etc (Surzycki, 2000, Pandey, 2005). An aliquot of DNA was applied into electrophoresis in 1.0% agarose gel in 1X TAE buffer (4.84 g Tris-acetate, 1.15 ml glacial acetic acid, 2 ml 0.5 M EDTA) and one Kb ladder (BioLab, New England) was used as standard marker.

Digestion of genomic DNA

In order to identify microsatellite sequences, total genomic DNA was digested by blunt-end-generating restriction endonuclease *Rsa*I (Roch, Germany) as shown in Figure 12. The reaction mixture and the conditions required for the digestion of genomic DNA are described in Appendix 2. The digested products were obtained and separated in 1.0% agarose gel electrophoresis in 1X TAE buffer.

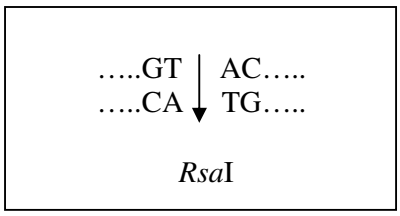

Figure 12 The *Rsa*I restriction site in the DNA sequence.

Ligation of adaptors

Digested DNA was ligated to *Rsa* adapters (21 mer and 25 mer) as shown in Figure 13 and incubated at 37 ^{0}C for 2 hours. To inactivate the enzyme, it was incubated at 95 ^{0}C and kept at room temperature.

A 21 mer	5'CTCTTGCTTACGCGTGGACTA3'
A 25 mer	5'TAGTCCACGCGTAAGCAAGAGCACA3'

Figure 13 Sequences of 21-mer and 25-mer adapters.

After ligation, the products were purified by using MinElute™ Gel Extraction kit (QIAGEN, Hilden) following the protocol of the kit, described in Appendix 3. After purification the restriction-ligation DNA fragments were eluted in 10 ?l. To test the success of restriction-ligation, PCR was performed as described in Appendix 4 and checked in 1% agarose gel.

Hybridization of oligoneucleotide probe

The restriction-ligation DNA fragments were enriched by hybridizing with biotinylated microsatellite oligonucleotide $(CT)_{10}$ probe and captured on MagneSphere? Magnetic Separation (Promega). At the end of this step, 10 ?l of enriched DNA samples were recovered and checked by PCR as described in Appendix 5 by using the 21 mer oligonucleotide (Figure 13) as a primer.

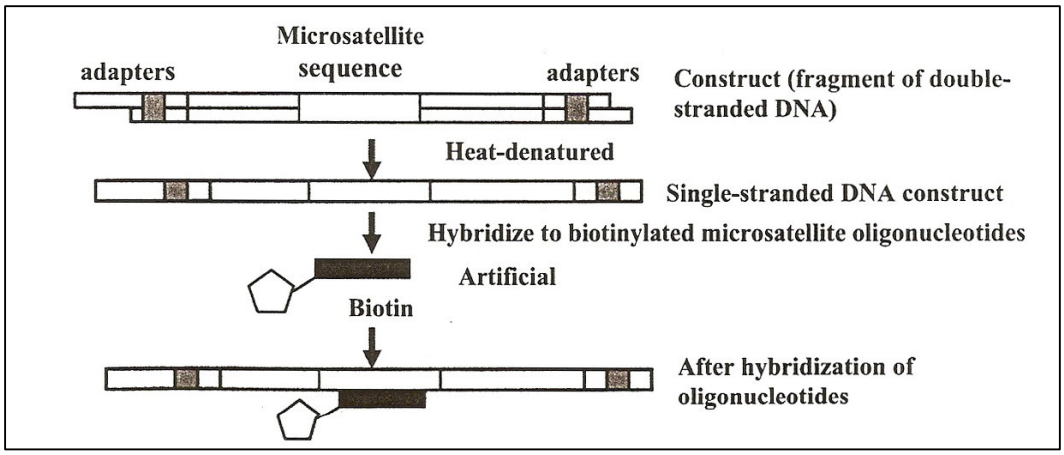

Figure 14 Different steps of the hybridization process.

Source : Pandey (2005).

Cloning into plasmid vectors

These DNA fragments were cloned into pGEM-T Easy vector system (Promega) using manufacturers' recommendations as described in Appendix 6. The vector map is shown in Figure 15.

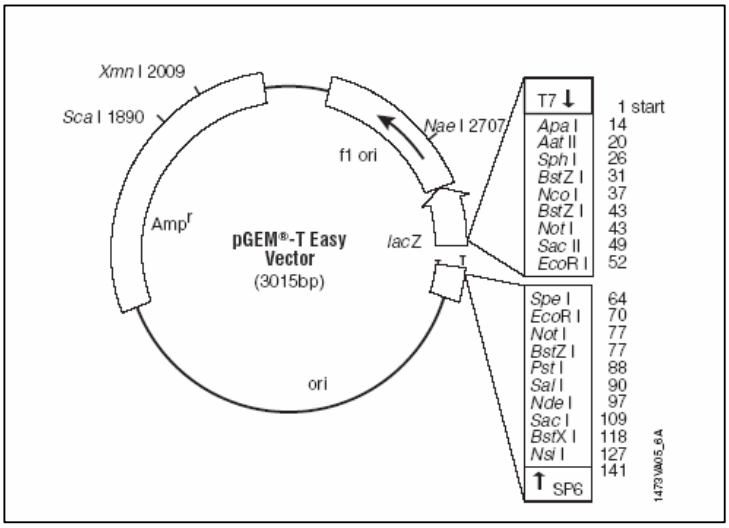

Figure 15 pGEM-T Easy Vector Circle Map.

Source: Promega? (2005).

Transformation in Bacterial

The competent cells were prepared as described in Appendix 7. Recombinant vectors were transformed into DH5α competent cells by the heat shock transformation method as shown in appendix 8 and grown on LB agar (1% Bacto Tryptone, 0.5% Yeast Extract, 1% NaCl and 1.5% Agar) plates containing ampicilin (100 mg/ml) and X-gal (20 mg/ml) for blue-white selection. After blue-white selection, white colonies were picked directly into new LB plate and insert plasmids were amplified by PCR using the universal primers T7 and SP6 (Reaction Mix and PCR program were described in Appendix 9).

Plasmids were extracted by using GenElute^{TM} Plasmid Miniperp kit (Sigma) as described in Appendix 10. The plasmids obtained from the extraction were used as template for the sequencing reactions by using the vector universal T7 or SP6 primers. The PCR conditions for sequencing reactions were as recommended by Applied Biosystems and used a fluorescent dye terminator.

Sequencing of DNA fragments

Sequencing of the fragments was done using BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) and carried out with an ABI Prism 3100-Avant Genetic Analyzer (Applied Biosystem Foster City, CA). The regions corresponding to the cloning vector and adapters were deleted using the BioEdit Program (Hall, 1999). The detail of reaction mix and PCR program was described in Appendix 11.

Identification of microsatellites

Microsatellites were categorized according to type, with slight modification to the derived by Webber (1990). A perfect repeat was defined as a run of tandem repeats without interruption. Imperfect repeats were defined as two or more runs of uninterrupted repeats where the terminal repeats consisted of at least three full repeat lengths for dinucleotide repeats and two full repeats lengths for repeats of all larger motifs. Compound repeats were defined as runs of repeats separated by no more than three consecutive non repeat nucleotides from a perfect or imperfect repeat or a different motif.

Designing primers

Primers were designed for the flanking region of the SSR using the Primer3 computer software (Rozen and Skaletsky, 2000). Primers were designed according to the default parameters of the program with: product size = 100 to 300, annealing temperatures between 50 ^{0}C and 60 ^{0}C, G+C content of ~ 50% to yield amplification products between 100 and 300 bp. Oligonucleotide primers were synthesized by KU vector (Thailand).

Application of microsatellite markers in determination of genetic variation in neem populations

For testing the primers previously developed, 24 populations of neem as describe in Table 3 were used.

DNA extraction

For each population young leaves from individual trees were independently collected. About five grams of leaf tissue was ground to powder with liquid nitrogen, using a mortar (Surzycki, 2000). It was suspended in 800 ?l of 2X CTAB (cethyltrimethyl ammonium bromide) extraction buffer methods. The extraction buffer contained 4% CTAB, 2.8 M NaCl, 40 mM EDTA, 200 mM Tris-HCl, pH 8.0, and 0.4% 2-mercaptoethanol. Samples were incubated at 65 ^{o}C for approximately one hour, mixed with an equal volume of chloroform-isoamyl alcohol (24:1), and centrifuged at 13,000 rpm for 10 min at room temperature. The supernatant was transferred to a clean microcentrifuge tube, repeating this step twice. The aqueous layer was collected and precipitated by adding 0.6 volumes of cold isopropanol. The DNA was pelleted by centrifugation at 13,000 rpm for ten minutes. The supernatants were removed and the DNA pellet washed with 700 ?l of 70% ethanol, repeating this step a second time. The pellet was dried for 20 minutes at 50 ^{o}C and then dissolved in 100 ?l of ultrapure water.

DNA quality and concentration were examined by electrophoresis in 1.0% ethidium bromide stained agarose gel in 1X TAE buffer and one Kb ladder was used as standard marker.

PCR amplification

For the amplification of DNA samples from leaves of Indian neem and Thai neem, 26 SSR markers (Table 4) previously developed were used as primers.

PCR reactions were carried out in a final volume of 10 ?l, containing approximately 20 nanograms of genomic DNA, 0.5 units of *Taq* polymerase (Invitrogen) 1X *Taq* polymerase buffer (Invitrogen), 2mM $MgCl_2$ (Invitrogen), 2.5 mM of dNTPs (Eppendorf) and 10 pmol of each primer. Amplification reactions were carried out on a PTC-200 Peltier Thermal Cycler (MJ Research) using the following cycling profile: 94 ^{0}C for 3 min followed by 30 cycles at 94 ^{0}C for 45 s, 55-58 ^{0}C for 30 s, and 72 ^{0}C for 1.30 min, and a final extension step at 72 ^{0}C for 10 min and the PCR products were stored at 4^{o}C before analysis.

For checking polymorphism, aliquots of the amplification products were loaded on 30% polyacrylamide gel. Gels were run for 45 minutes at 1000 volts. They were then checked and visualized by Gel Scan 3000 (Corbetta Robotics). Alleles were scored according to molecular weight, and were estimated by comparing with a 50 bp ladder (Invitrogen) and check polymorphism.

Data analysis

The data achieved from polyacrylamide gel electrophoresis were interpreted using program TFPGA (Miller, 1997) and POPGENE version 1.32 (Yeh *et al*., 1999). The bands obtained from polyacrylamide gel electrophoresis were scored and genotyped for further statistic analysis as follow:

1. Number of allele per locus (N)

$$N = \frac{\text{Total alleles}}{\text{Total number of loci studied}}$$

2. Effective number of allele, n_e is estimation of the reciprocal of homozygosity (Kimura and Crow, 1964). The data were calculated using the software POPGENE version 1.31 (Yeh *et al.,* 1999).

$$n_e = 1/\Sigma p_i^2$$

Where p_i = frequency of the *i*th allele for the studied locus

3. Allele frequencies were calculated using the software POPGENE version 1.31 (Yeh *et al.,* 1999).

$$X_i = (2H_A+H_B)/2N$$

Where X_i = the frequency of the *i*th allele

H_A = number of homozygous

H_B = number of heterozygous

N = number of all samples

4. Heterozygosity is the state of having two different alleles of the same gene. It was calculated using the program POPGENE version 1.32 (Yeh *et al.,* 1999).

4.1 Observed heterozygosity (H_o), was calculate as (Nei, 1978)

$$H_o = \frac{\text{Number of heterozygous individuals}}{\text{Total number of genotypes per locus}}$$

4.2 Expected heterozygosity (H_e), were calculated according to the formula (Nei, 1978).

$$h_k = 2n(1-\Sigma x_i^2)/2n-1$$

$$He = \Sigma h_k/r$$

Where
- h_k = the value of h for the k*th* locus
- X_i = the frequency of the *i*th allele
- N = number of the *k*th allele
- r = number of study allele
- H_e = average heterozygosity per locus

5. Test of Hardy-Weinberg equilibrium

The Hardy-Weinberg equilibrium theory is explain the frequency of genotype at a given locus in a random mating population. The data were calculated using the software TFPGA version 1.3 (Miller, 1997).

6. Genetic distance (Nei, 1978), standard genetic distance is defined as

$$D = -\ln[G_{XY}/\sqrt{G_X G_Y}]$$

Where G_X = the means of Σp^2_i all over loci

G_Y = the means of Σq^2_i all over loci

G_{XY} = the means of $\Sigma p_i q_i$ all over loci

7. Phylogenetic tree

Phylogenetic tree were produced by clustering the data with the unweighted pair group method (UPGMA) (Sneath and Sokal, 1973) using the program TFPGA (Miller, 1997).

8. F-coefficient (Wright, 1978)

F-statistic is a value to determine if the variances between the means of populations are significantly different, this value determine the P-value. The different F-statistic look at different level of population structure. F_{it} is the inbreeding co-efficient of an individual; F_{is} is the inbreeding co-efficient of and individual relative to the subpopulation; and F_{st} is the effect of subpopulations compared to the total populations. The F-coefficient were calculated using TFPGA program (Miller, 1997) according to the formula

$$F_{is} = (H_T - H_S)/H_T$$

$$F_{it} = (H_T - H_O)/H_T$$

$$F_{st} = (H_S - H_O)/H_S$$

Where H_O = Observed heterozygosity in populations

H_T = Expected heterozygosity in populations

Hs = Expected heterozygosity in subpopulaitons

Places

1. DNA and Isoenzyme Laboratory, Forest Genetic and Biotechnology Research Group
   National Park, Wildlife and Plant Conservation Department
2. FAO International Provenance Trials Germplasm bank of neem, Kanchanaburi

Duration

February 2004 to June 2007