Kasetsart J. (Nat. Sci.) 40 : 91 - 98 (2006)

Development of *Catalase* Gene Nuclear DNA-Based Marker for Population Genetic Analysis in Thai Teak (*Tectona grandis* L.f.)

Jongkon Cheua-ngam[1] and Hugo Volkaert[1,2]*

ABSTRACT

The teak Catalase gene was amplified and cloned using consensus PCR primers designed based on sequences from other plant species. The obtained teak *Cat* DNA sequences were used to develop specific primers. The specific primer set could successfully amplify a specific DNA fragment from teak in all populations studied. Using PCR-SSCP, 5 different alleles were detected. The tested nuclear gene primer set had considerable potential as a DNA marker for population analysis in teak. This approach could be used to specifically amplify fragments in other plant species and applied to study evolution, population genetics, outcrossing rate and mating patterns.

Key words: *Catalase*, teak, *Tectona grandis*, population genetics, SSCP

INTRODUCTION

Thailand used to be an important supplier of teak timber to the world market. However, due to overharvesting, yields declined after about 1960 (De'Ath, 1992; Graudal *et al*., 1999). The unsustainable extraction of timber from natural forest has become a major cause of concern about deforestation, forest degradation and soil erosion. Therefore, a complete ban of commercial exploitation of natural forests was decreed in 1989. Plantations of quality hardwood species under sustainable management are seen as an alternative to timber extraction from natural forests. A genetic improvement program is an essential component of a successful and sustainable plantation project. Knowledge of genetic variation within and between populations of teak is important for both conservation of remaining genetic resources and breeding for plantation varieties.

Molecular genetic analysis techniques are now widely used to study genetic diversity and gene flow in plants. DNA markers are being used for fast surveys of genetic variation in several crop improvement programs, as a tool in classification, selection, and breeding. The development of appropriate DNA-based markers for teak would make it possible to study gene-flow and genetic diversity aspects of this important species. A recent development in population genetics is the application of genealogical approaches to quantify diversity and gene flow processes (coalescent), but these methods require information on both allele frequencies and genealogical relationship of the alleles. Recent progress in marker development has been based on the detection of nucleotide sequence variation within introns of a few specific protein-coding genes (e. g. Palumbi and Baker,

1 Center for Agricultural Biotechnology, Kasetsart University, Kamphaengsaen Campus, Nakhon Pathom 73140, Thailand.
2 National Science and Technology Development Center (NSTDA-BIOTEC), Science Park, KlongLuang, Pathumthani 12120, Thailand.
* Corresponding author, e-mail: hugo.v@ku.ac.th

Received date : 02/03/05 Accepted date : 31/10/05

1994; Strand *et al*., 1997). Now nuclear gene sequences are available in genome databases and promise to greatly assist the search for new nuclear markers especially expressed sequence tag libraries (ESTs). The extensive EST databases reveal homologs from various plant species at shallow or deep levels of evolutionary history from which consensus primers for PCR could be developed. Therefore, the aim of this study was to develop specific *Catalase* (*Cat*) nuclear marker tools for teak that could be used in management of genetic resources and teak breeding programs.

MATERIALS AND METHODS

Sample collection

Individual young leaves were collected from trees in a forest remnant near Potharam, Ratchaburi province in April 2003. Additional leaf or bud samples were collected from trees in 6 natural populations in Thailand during August and September 2004 (Table 1). With a few exceptions, trees sampled from the natural populations were at a distance of at least 100 m between each individual as measured by GPS in the field. Approximately 100 mg tissue was transferred to a 1.5 ml tube containing extraction buffer within 12 hours from collection for further processing in the laboratory.

DNA extraction

The fresh young leaf samples from the Potharam population were extracted using DNeasy Plant Mini Kit (Qiagen) according to the manufacturer's instructions. The samples of individual teak from natural populations were extracted using a modified CTAB method (Doyle *et al*., 1989). The DNA concentration of all samples was estimated on agarose gel by comparing to a standard of known concentration (Fermentas).

Development of *catalase* gene primers

Development of consensus primers

PCR primer pair for *Cat* genes was designed based on information of DNA sequences obtained from other plant species. Sequences from different plant species were retrieved from publicly accessible DNA databases (GenBank/EMBL/DDBJ) by keyword searching. The sequences were aligned using the ClustalW program (http://www.ebi.ac.uk/clustalw/index.html). The resulting alignments were improved by visual inspection using the GeneDoc program. Several criteria for selecting consensus PCR primers were used. (1) All conserved regions, at least 7 codons long, were identified. The primer region should optimally be conserved in all plant sequences or at least in all dicot plant species. (2) The conserved regions were compared for minimal redundancy at the 3' end of a potential primer. (3) Two primers should be located in opposite direction, preferably one or more introns apart. (4) The length of the target DNA sequences should not be longer than 2000

Table 1 List of plant material and collecting sites.

Population	Geographic region	Sample size (tree)
Tg8-01	Photharam, Ratchaburi	23[1] + 46[2]
Tg8-02	SaiYoke National Park, Kanchanaburi	20
Tg8-03	MaeMoei National Park, Tak	34
Tg8-04	ChiangDao National Park, Chiang Mai	29
Tg8-05	PratooPhaa, Lampang	48
Tg8-06	SakYai Forest Park, Uttaradit	40
Tg8-07	Srisatchanalai National Park, Sukhothai	44

[1] individual trees sampled all over the Potharam District

[2] all teak trees in a small forest patch behind WatBoht, Potharam District

bp as determined from available genomic sequences. (5) The length of potential primer sites, their G-C content (especially at the 3' end), and the position of the 3' end relative to the reading frame were additional criteria considered. By placing the critical 3' end at a second codon position nucleotide for forward and at the first codon position for reverse primer, the possibility of a mismatch preventing annealing may be reduced. In some cases, primer pairs with sub-optimal melting temperatures may be necessary due to the paucity of conserved regions in the sequences. Additionaly, (6) the sequence of candidate primers was compared to databases using BLAST. If the result of BLAST indicated that the primer was similar to sequences from fungi or bacteria, this primer sequence was discarded.

DNA fragments were PCR amplified using the consensus primer pairs in a total reaction mixture of 25 µl, containing 200 µM dNTPs (Promega), 5 pmole of each primers, 1× PCR buffer with 2 mM $MgCl_2$ (Qiagen), 0.3 units of *Taq* DNA polymerase (Qiagen), and 20 ng of genomic DNA template. PCR products were run on agarose gel. The reproducible PCRs were selected for cloning using TOPO-TA cloning kit (Invitrogen). Individual clones were picked, and plasmid DNA was purified and sent for sequencing.

Development of teak gene specific primers

The DNA sequences obtained from the cloned fragments were integrated in the alignments containing the sequences of other plant species. The intron - exon boundaries (GT-AG) were determined in the teak sequences. The conserved region of teak sequences were chosen to design specific primers.

Application of markers to population analysis

PCR-SSCP

DNA samples were amplified from genomic DNA by using the specific primers in 96 well plates in a total reaction mixture of 15 µl, containing 200 µM dNTPs (Promega), 2.5 mM $MgCl_2$ (Fermentas), 5 pmole of each primer, 1× PCR buffer with $(NH_4)_2SO_4$ (Fermentas), 0.3 unit of *Taq* DNA polymerase (Fermentas) and 20 ng of genomic DNA template. Amplification was carried out at 94°C for 3 min, followed by 35 cycles of 45 sec at 94°C, 45 sec at 50 °C annealing temperature, 90 sec at 72°C, and a final extension at 72°C for 5 min (15 min when the fragment was to be used for cloning). PCR products were electrophoresed on 1% agarose gel using 1xTAE buffer at 50 V for 40 min, stained with ethidium bromide and photographed. Four volumes of loading dye (98% formamide, 0.025% bromophenol blue, 0.025% xylene cyanol, and 10mM NaOH) were added to the PCR products, then denatured at 95°C for 10 min, and immediately placed on ice-cold water to stabilize single strands. Electrophoresis was performed on polyacrylamide gel using non-denaturing conditions (Single-Strand DNA Polymorphism, Orita *et al*., 1989). 3.5 µl of the aliquots were loaded on a 30 cm× 40 cm × 0.4 mm polyacrylamide (Sequagel MD, National Diagnostics, U.S.A.) gels attached to glass plate in 0.6x TBE buffer using Hoefer SQ3 Sequencer (Amersham Pharmacia Biotech), run in a 4°C refrigerator at constant 8 watt for 16 h and visualized by silver staining.

Cloning and sequencing

From the banding patterns that appeared on the SSCP gel, the nuclear allelotypes were scored. Representative individuals were selected for PCR cloning. The PCR products were purified using QIAquick® PCR Purification Kit (Qiagen) according to the manufacturer's instructions then ligated into pGEM-T Vector Systems I (Promega) according to the manufacturer's instructions. The ligations were transformed into competent cell (*Escherichia coli*, 'DH10B') by electroporation.

The clones were grown overnight at 37 °C in LB (Luria-Bertani) medium agar with 100 μg/ml of antibiotic (ampicilin), and then screened by blue-white colony selection. Each cloned copy of PCR amplification represented a single allele, and unambiguous sequence of both alleles in an individual could be obtained by sequencing multiple clones. Single colonies were selected for PCR amplification (reaction in each primer used same as corresponding PCR-SSCP method) to check the presence of insert and type of allele. The remainder of the same single colony was grown overnight in an incubator shaker at 37°C, 150 rpm in 700 μl of LB medium broth with 100 μg/ml of ampicilin in 1.5 microcentifuge tube, then kept in 25 % glycerol and stored at -80°C. The PCR products of the clone were electrophoresed on polyacrylamide gel (same method was used PCR-RF-SSCP). Two μl of bacterial culture containing plasmids representing different alleles were used to inoculate 5 μl of LB medium broth with 100 μg/ml of ampicilin in 15 μl tube and grown overnight in incubator shaker at 37°C, 150 rpm. The plasmids were extracted using Wizard® *Plus* SV Minipreps DNA Purification System (Promega) according to manufacturer's instruction, and then sent for sequencing at Macrogen, Inc. (Seoul, Korea).

RESULTS

Development of consensus primer

73 DNA sequences coding for *Cat* genes representing 38 plant species were obtained from databases. The retrieved sequences consisted of partial and complete cDNAs, ESTs, and genomic DNA fragments. The sequence alignment allowed the design of a primer set of general usability among plants. Comparison of expressed and genomic DNA sequences indicated that the intron-exon boundaries were not 100 % conserved among loci and plant species.

The positions of the consensus primers were chosen in exon 3 and exon 4. The primer sequences were 5-GGT TTC TTT GAR GTY ACN CAY GA-3 for forward and 5-TG ATG AGC ACA YTT NGG NGC RTT-3 for reverse primer (with R = A or G, Y = C or T, N = A, C, G or T). In *Arabidopsis thaliana* (At1g20630) these primers would amplify a fragment of approximately 1100 bp.

The PCR reactions using the CAT primer set amplified a fragment in all populations, producing a single band on agarose gel. Two positive clones were obtained from this primer set and sequenced. The length of the inserts was 856 bp.

Development and analysis of specific nuclear gene primer sets

The initial sequence from two teak clones were compared to GenBank databases using BLAST for verification of gene identity. The position of the specific primer for teak were chosen in exon 3 and 4, spanning the intervening intron (based on comparison with the complete sequence of *A. thaliana*, At1g20630). The sequences of the teak specific primers were 5'-CGATTCTCCAC TGTCATCCA-3' for forward and 5'-GGAA GTTGTTTCCCACCAAA-3' for reverse primer.

The specific primers of *Cat* gene successfully amplified a fragment from all teak populations. On agarose gel, a single band of approximate 400 bp was visible. Different alleles could not be distinguished (Figure 1). However, using SSCP, several alleles could be identified in all populations (Table 2). Different alleles from the Potharam population are shown in Figure 2. Using sub-optimal annealing temperature during the PCR, a second locus was revealed on SSCP. Therefore, PCR amplification at the highest possible annealing temperature (52°C) was necessary. After cloned fragments were sent for sequencing, the size of the DNA amplified fragments varied between 388 to 413 bp.

DISCUSSION

Development of primer set

The specific teak catalase primer set could successfully amplify a DNA fragment from the nuclear DNA, and produced a single band on agarose gel, but no differentiation of allele was observed. However different alleles could be identified in each population when electrophoresis was done on polyacrylamide gel using SSCP conditions. The sequences obtained from cloned fragments indicated that the target gene was amplified. A similar approach has been used to amplify specific DNA fragments from other plants such as, *Musa* spp. *Xylia xylocarpa* and *Andrographis paniculata* (data not shown). These results demonstrated that sequences from other species could provide enough information to design primers that amplified low copy number nuclear genes from various plant taxa. Therefore these primers could be called "universal primers" and useful for population studies in other plant species as well. Because each amplification product should include an intron, the differences in size among the bands could, in general, be attributable to variation in intron length among duplicated loci. Once a particular band has been sequenced, it is possible to design new primers that are specific to that band in the teak genome. Single amplification products can then be easily screened for variation by SSCP direct sequencing.

Because intron sequences evolved at a much higher rate than exon sequences, it was expected that nuclear gene based markers would show higher polymorphism than protein based

Table 2 Number of alleles from 7 populations amplified from the Cat locus.

Population	Number of alleles
Potharam	4
SaiYoke	3
MaeMoei	4
ChiangDao	3
PratooPha	3
SakYai	3
Srisatchanalai	4
All populations	5

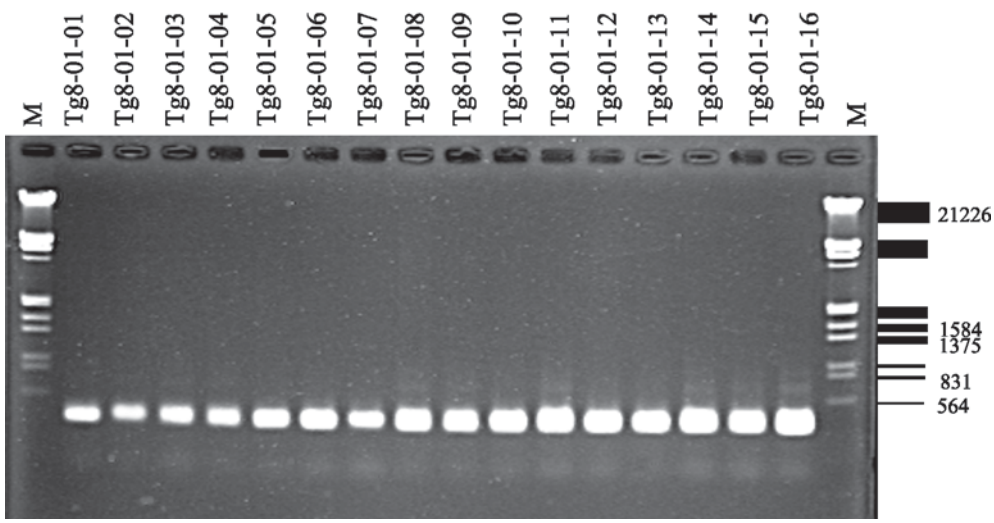

Figure 1 PCR products amplified from Potharam teak samples using the teak specific CAT primer set. Lane M is a size standard (λ/Hind III+EcoRI), the approximate size of the fragments is 400 bp.

isozymes. The copy number for nuclear biosynthetic genes is generally low and the rate of mutation is higher than for chloroplast (Fineschi *et al*., 2004) and mitochondrial genes, therefore they can be used to quantify diversity and study evolutionary processes, population genetics, outcrossing and mating patterns in teak tree or other plants.

Application of specific primer to population genetic analysis

The obtained sequences confirmed variation within the intron, while the exons were conserved (Mason-Gamer *et al*., 1998; Ishii and McCouch 2000; Natari *et al*., 2003).

The *Cat* locus showed polymorphism in all populations examined. The number of alleles that could be detected using this approach was markedly higher than the number of alleles that were detected in a previous report using isozyme assays. Kjaer *et al*. (1996) reported only 2 to 4 alleles at 10 isozyme loci for trees from a provenance trial containing populations from India, Indonesia and Thailand.

Initial analysis of the SSCP banding patterns obtained could be problematic since it was not known in advance how many bands could be expected and how the different alleles would behave during the electrophoresis. Simpler banding patterns were assumed to represent homozygous individuals while banding patterns that consisted of combinations of two different simple banding patterns were assumed to represent heterozygous individuals (Figure 2). In some samples obtained from an international teak provenance trial the number of different banding patterns observed was too large for a straightforward analysis without testing for segregation in progenies (data not shown). Further studies, including analysis of segregation in progenies, will be needed to clarify the allelic relationships among the observed SSCP banding patterns.

For sequencing of the different alleles, several problematic sequences were obtained from cloned fragments, resulting from *Taq* polymerase errors and recombination (Cronn *et al*., 2002) between different alleles from a heterozygous individual during the PCR amplification step. Therefore, multiple clones were sent for sequencing with at least 2 clones from homozygous trees and four clones from

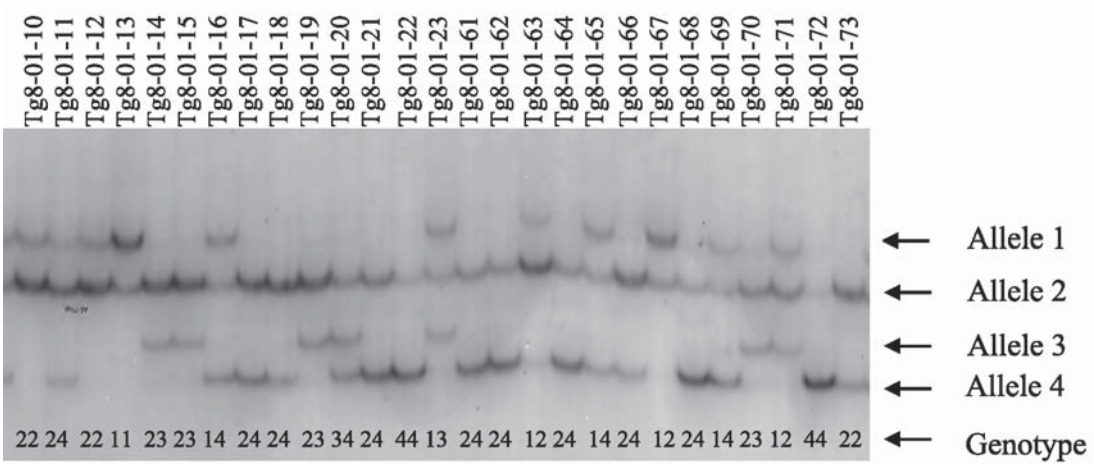

Figure 2 SSCP polyacrylamide gel of the *Cat* marker amplified from individual teak trees from the Potharam population. Different alleles are assigned. Note that without progeny testing the different types of banding pattern observed make allelic inferences tentative.

heterozygous trees to confirm the sequence identity of each allele. Palumbi and Baker (1994) estimated the frequency of *Taq* polymerase error rate, at ~ 1 transition substitution per 1000 bases, by sequencing multiple clones in heterozygotes and comparing several individual sequences from the same allele. To avoid the problem of DNA polymerase errors retained in individual clones, PCR products could be sent for directly sequencing.

CONCLUSION

Although no sequence information was available for the particular target species, enough information on related species was stored in public databases allowed to design PCR amplification primers. The primer set of *Cat* was designed from consensus sequence regions in other plant species and could successfully amplified a fragment of the teak genome. The obtained sequences could be used to design specific primers that amplify a single locus from the teak genome. A high level of polymorphisms was observed when analyzed using the SSCP separation method. The different alleles in each population could be distinguished and counted. Therefore, DNA based markers developed from low copy nuclear genes have potential for use in population genetic studies in teak and other plant species.

ACKNOWLEDGEMENTS

We would like to thank European Commission DG XII Research (project number: ICA4-CT-2001-10094) for financial support and the Center for Agricultural Biotechnology, Kasetsart University for providing a research assistantship. We would like to thank the National Parks, Wildlife and Plant Conservation Department for granting permission to collect samples in some National Parks. We acknowledge the assistance of Mr. Jertpong Makaramani, Dr. Somran Sutdee, and the staffs at the National Parks, Mr. Prayart Chaichon from Pratoo Phaa, Lampang province, and the villagers from WatBoht, Potharam, Ratchaburi province for assistance during the collection of plant materials.

LITERATURE CITED

Cronn, R., M. Cedroni, T. Haselkorn, C. Grover and J. F. Wendel. 2002. PCR-mediated recombination in amplification products derived from polyploid cotton. **Theor. Appl. Genet**. 104: 482-489.

De' Ath, C. 1992. A history of timber exports from Thailand with emphasis on the 1870-1937 period. **Nat. Hist. Bull. Siam Soc.** 40: 49-65.

Doyle, J.J. and J.L. Doyle. 1989. Isolation of plant DNA from fresh tissue. **Focus** 12: 13-15.

Fineschi, S., S. Cozzolino, M. Migliaccio and G. G. Vendramin. 2004. Genetic variation of relic tree Species: the Case of Mediterranean *Zelkova abelicea* (Lam.) Boisser and *Z. sicula* Di Pasquale, Garfì and Quézel (Ulmaceae). **Forest Ecology and Management** 197: 273-278.

Graudal, L., E.D. Kjaer, V. Suangtho, P. Saardavut and A. Kaosa-ard. 1999. **Technical Note 52, DANIDA Forest Seed Center, Humlebaek.** 36.

Ishii, T. and S.R. McCouch. 2000. Microsatellite and microsynteny in the chloroplast genome of *Oryza* and eight other gramineae species. **Theor. Appl. Genet**. 100: 1257-1266. Kjaer, E.D., H.R. Siegismund and V. Suangtho. 1996. A Multivariate study on genetic variation in teak (*Tectona* (L)). **Silv. Gen.** 45(5-6): 361-368

Marson-Gamer, R.J., C. F. Weil and E. A. Kellogg. 1998. Granule-Bound starch synthase: structure, function, and phylogenetic utility. **Mol. Biol. Evol.** 15(12): 1658-1673.

Natari, L., T. Giordani and A. Cavallini. 2003. Sequence variability of a dehydrin gene within

Helianthus annuus. **Theor. Appl. Genet**. 106: 811-818.

Orita, M., H. Y. Suzuki, H. Kanazawa, T. Sekiya and K. Hayashi. 1989. Rapid and sensitive Detection of point mutations and DNA polymorphisms using the polymerase chain reaction. **Genomics** 5: 874-879.

Palumbi, S. R. and C.S. Baker. 1994. Contrasting population structure from nuclear intron sequences and mtDNA of Humpback whales. **Mol. Biol. Evol**. 11(3): 426-435.

Strand, A.E., J. Leebens-Mack and B.G. Milligan. 1997. Nuclear DNA-based markers for plant evolutionary biology. **Mol. Ecol.** 6: 113-118.