RESEARCH ARTICLE

ScienceAsia 31 (2005): **145-149**

DNA Fingerprint Database of Some Economically Important Thai Plants: *Litchi chinensis* Sonn, *Dimocarpus longan* Lour, and *Peuraria* spp.

Rungrach Wangspa[a], Robert W. Cutler[b,*], Supranee Sitthiprom[a], Ruttaporn Chundet[a], Nadtaya Dumampai[c], and Somboon Anuntalabhochai[a]

[a] Department of Biology, Faculty of Science, Chiang Mai University, Chiang Mai 50202, Thailand.
[b] Department of Biology, Bard College, Annandale-on-Hudson, New York 12504, USA.
[c] Trang Horticultural Research Center Siako, Trang 92150, Thailand.

* Corresponding author, E-mail: cutler@bard.edu

Received 26 July 2004
Accepted 12 Jan 2005

ABSTRACT: The high annealing temperature random amplified polymorphic DNA (HAT-RAPD) method can be used to generate highly polymorphic data from PCR amplification of DNA samples to determine relatedness in plant cultivars. Using the HAT-RAPD method, we created a fingerprint database of tropical plants for fourteen subspecies of *Dimocarpus longan* Lour., twelve subspecies of *Litchi chinensis* Sonn., and seven distinct varieties of the genus *Peuraria*, two of which were previously uncharacterized. Out of a total of 22 distinct primer sets, a subset of primers with reproducible DNA band patterns was characterized for each species. From this data, we developed a web-accessible database that both graphically and quantitatively depicts the existent bands and species phylogenies using character state data generated from the banding patterns. Using historical and geographic data for these plant species and subspecies, the generated phylogenies support the currently accepted species relationships for *D. longan* and *L. chinensis* and characterize the unidentified varieties of *Peuraria*. The HAT-RAPD experimental method combined with band pattern recognition is a cost effective and easily characterized methodology that can be used to identify plant varieties, as well as to advance the knowledge of biodiversity in previously uncharacterized species.

KEYWORDS: HAT-RAPD, DNA fingerprinting, Online Database, Bioinformatics.

INTRODUCTION

Since Thailand's economy is based largely on agricultural production, many plant varieties have been generated by means of selective breeding to improve crop yield or to overcome growth limitations and diseases. This gave rise to an assortment of hybrids and divergent strains in some economically important species. For example, there are over 50 varieties of *Litchi chinensis* Sonn. and some 45 distinct varieties of *Dimocarpus longan* Lour. found in orchards throughout the country. Based on physical appearance alone, uncertain or even misleading identification of an individual plant is certain. An easy and reliable means to distinguish one variety from another is needed.

Molecular techniques such as RAPD with arbitrary primers,[9,13,15,16] restriction fragment length polymorphism,[5,8,17] amplified fragment length polymorphism,[3,12] variable number of tandem repeats,[10] and sequence-tagged simple sequence repeats are powerful techniques for species and variety identification. The RAPD technique can detect DNA segments present in individual samples to infer genetic relationships between different populations. However, a serious shortcoming of RAPD is inconsistency of banding patterns.[6] Our group has further developed a technique called HAT-RAPD. The technique, instead of using lower annealing temperatures, utilizes temperatures between 40-46°C to increase the specificity of the annealing conditions. It was successfully shown that, coupled with several random decamer primers, repeatable highly polymorphic bands can be generated.[1] Independent of this work, optimization of RAPD analysis by elevating annealing temperature, which then yielded reproducible DNA profiles, has also been reported.[2] By noting the presence or absence of conserved bands in the amplified products (which we designate as a character state measure for the Phylip DISCRETE software package[7]), varieties and hybrids of *L. chinensis* Sonn.[1] and *D. longan* Lour.,[14] were effectively and reliably identified. Thus, the HAT-RAPD profiles can further be adapted for plant identification.

The purposes of this work are to: (1) develop an algorithm to identify both conserved and polymorphic bands generated with each primer, (2) assemble HAT-RAPD marker polymorphic data into a database for

quick and easy dissemination of this resource, and (3) compute genetic diversity and relatedness of plant samples from this data. Our database generation algorithms were tested on HAT-RAPD data of *L. chinensis* Sonn. and *D. longan* Lour., as well as several species of *Peuraria*. Dendrograms generated from the analysis were compared with known background data for each plant species, and correctly regenerated the geographic and transplantation history of the longan and lychee species. Two previously unidentified varieties of the *Peuraria* species , were also characterized.

Materials and Methods

Molecular Analysis of Plant Samples

Positively identified cultivars of *L. chinensis* Sonn. and *D. longan* Lour. were obtained from germplasm collection plots at Chiang Rai Horticultural Research Center, Chiang Rai, Thailand. *Peuraria* samples were collected from multiple sites in the provinces of Chiang Rai and Chiang Mai. To obtain DNA for the analysis, plant leaves were collected from 3 separate shoots of each variety in question and ground to a fine powder in liquid nitrogen. DNA extraction, PCR amplification, and DNA band separation were carried out as described elsewhere.[1] In short, a DNA sample was obtained after incubating ground plant samples in CTAB extraction buffer (4% (w/v) CTAB, 1.4 M NaCl, 20 mM EDTA, 100 mM Tris-HCl pH 8.0, 1% (w/v) polyvinylpolypyrrolidone, and 0.1% (v/v) b-mercaptoethanol) at 60°C for 60 min, followed by RNase treatment and phenol extraction. Twenty nanograms of DNA was amplified using decamer primers of arbitrary sequence (Operon Technologies, Almeda, California) at an annealing temperature of 46°C. PCR products along with a known standard, either a 100-bp ladder or a *Pst*I-digested lambda DNA, were separated in 1.5% agarose gel with TBE (40 mM Tris-borate pH 8.0, 1 mM EDTA) and visualized under UV illumination.

Size determinations of the amplified products were generated as follows. A section of HAT-RAPD profiles for a single primer with all bands repeated in at least 3 PCR amplifications were selected. These bands were found to have sizes ranging from 200 to 2000 base pairs (bps). The profiles were digitized and band size determined using Quantity one 1-D analysis software version 4.4.0 (Bio-Rad Laboratories, Hercules, California). To account for variable intensities within different gels, lane-based background subtraction was performed before band detection, and an intensity trace of each lane was used as a guide for manual band identification. The size of each band was determined by matching against the standard in the same HAT-RAPD gel.

Database construction

The banding data generated by the Quantity one 1-D analysis software package was found to give a highly accurate band size determination for bands within a single electrophoresis run, but when identical samples were run multiple times, the raw base-pair results were initially found to be incomparable beyond a simple visual inspection. To correct for intra-gel differences, the following transformations were applied to the data to make them directly comparable. Initially, the band size data in each primer set run was used to assign bands into band classes. These classes were then characterized statistically using a set of standard PERL routines to determine the mean and standard deviation (SD) for the number of bps in each group. From bands within a single gel, the accuracy of band weight detection was found to have a SD of no more than 5 bps, with the average SD for all groups of about 3 bps. This observation was used to assess the validity of band patterns and to correct for gel specific variation. For bands to be considered distinct they were required to occur at a distinguishable distance which was set as having their confidence intervals non-overlapping. Within individual groups, outlying bands, which had initially been classified visually as similar, were reclassified into distinct band classes if they fell outside this limit.

Comparing band patterns on different HAT-RAPD gel runs, the banding patterns within each run were reproducibly distinct, but between runs the absolute molecular weight assigned to each band class was often shifted by up to as much as 15 bps. To correct for this, between distinct runs for a given primer set, totally conserved bands as shown graphically in (Fig 1) were initially designated as single classes and the molecular weights for these highly conserved bands were then linearly regressed against each other. The linear regression was accomplished using standard PERL routines and the separate runs were merged into large, single comparable datasets. Then all bands with shared confidence intervals were assigned to the same band class.

These band classes were then used to create character state data. The presence or absence of a band within a band class was designated with a 0/1 character state. The database contains the complete set of all banding patterns for the 22 primers as well as a graphical overview of the 80 to 250 distinct bands within each primer run. This overview contains both the average band positions with statistical significance given the multiple electrophoresis runs (a minimum of 3 and maximum of 5) used to generate each group. For each species, in addition to the overall summaries, complete datasets for the banding character states, summary distance matrices, and predicted phylogenies

with bootstrapping values are available at the online database.

Results and Discussions

The HAT-RAPD fingerprint database of tropical plants is a collection of HAT-RAPD band size data. HAT-RAPD species data was determined previously by our laboratory[1,14] for all species except *Peuraria*. The database discussed in this manuscript can be viewed at http://science.bard.edu/thaiplant/. The database was divided into sections according to plant species and oligonucleotide primers utilized in the PCR reaction. The band size data file, character state, distance matrices, and phylogenies are all directly accessible as part of the database. In addition to the band data for each sample, a summary of band types is also displayed with the band width given by the calculated SD for that band class. (Fig 1) shows a sample graphical overview of the large amount of raw data which can also be found at the website.

Although a more limited range of band types was utilized, the dendrograms of *L. chinensis* Sonn. (Fig 2) and *D. longan* Lour. (Fig 3), found in earlier research,[1,14] were similar to our findings. These dendrograms, which were generated by bootstrapping the character state data can be viewed at the web site indicated above and were printed using the program TREEVIEW.[11] The newly characterized *Peuraria* dendrogram consists of two main branches (Fig 4). The first branch contains the subspecies *P. stricta* and *P. wallichii*, which are both found as bush-like growths in the northern region of

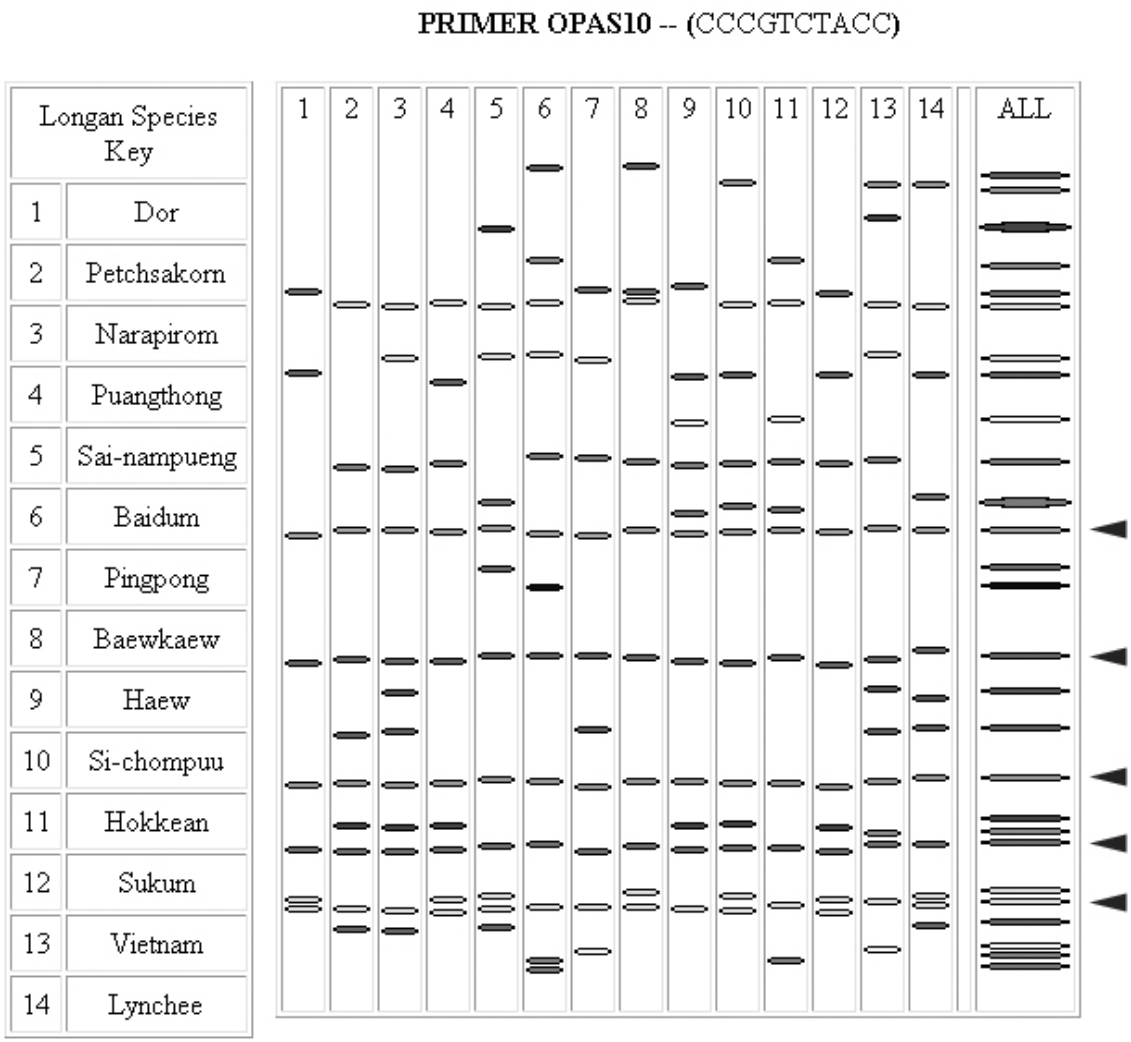

Fig 1. Representative HAT-RAPD data set for primer OPAS10. The five completely conserved bands which were used to linearly regress the separate lanes are indicated by arrow heads. Additional analyses can be reached at http://science.bard.edu/thaiplant/.

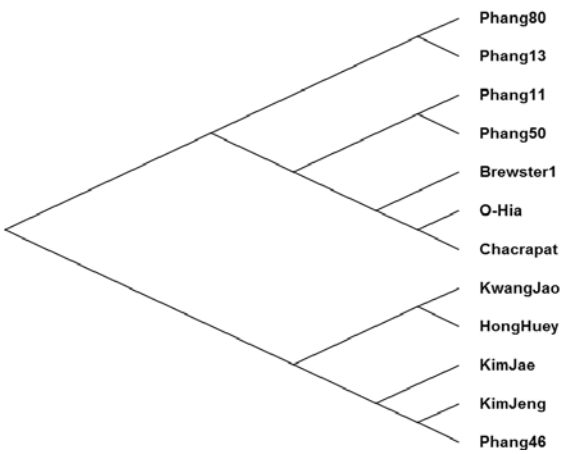

Fig 2. Genetic relationships among twelve subspecies of *Litchi chinensis* Sonn. This dendrogram was constructed from character state data using the Phylip software package using discrete character states and bootstrapped 1000-fold. Bootstrap values are available at the online database.

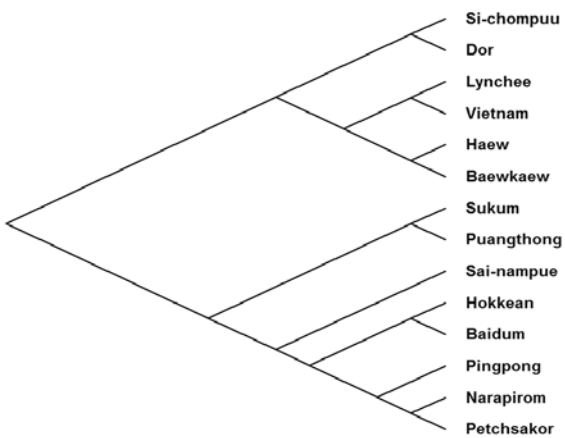

Fig 3. Genetic relationship among fourteen Longan cultivars. This dendrogram was constructed from character state data using the Phylip software package using discrete character states and bootstrapped 1000-fold. Bootstrap values are available at the online database.

Thailand. The second branch is composed of the *P. allopecuroides*, *P. mirifica*, *P. phescoloides*, and *P. candollei* varieties. These plants are found as vine-like growths. The root for the tree shown in (Fig 4) was set using the two main classes of growth presentation. This rooting places the previously unidentified subspecies of *P. candollei* as most closely related to *P. mirifica*. It should be noted that *P. mirifica* produces a biomedically important phytoestrogen in its roots called miroestrol, which has been reported to have rejuvenating properties.[4] Additional research on the *P. candollei* subspecies may shed some light on its use in traditional folk medicine.

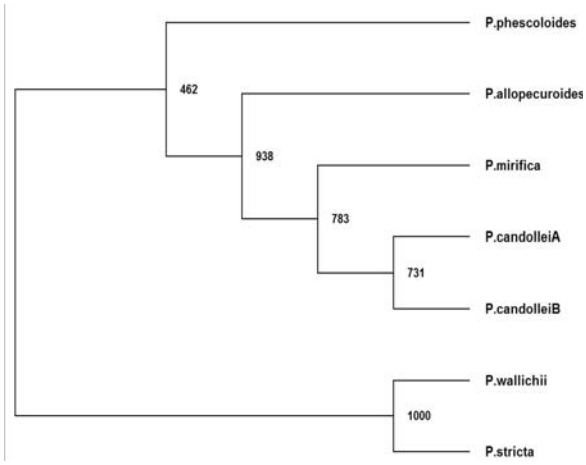

Fig 4. Genetic relationship among six Peuraria species and two previously unidentified subspecies of Peuraria candollei A & B. The internal nodes display the bootstrapping value using a 1000 replicate bootstrapping analysis.

As additional HAT-RAPD data becomes available, it can be integrated into the database to extend the results described in this manuscript. A dendrogram calculated from these new distance matrices would indicate the relatedness of the new strain to the existing ones. One major benefit of having a centralized data repository for plant DNA profile data is that as new strains of previously characterized species or even new species are examined, the data can be updated in the database and therefore immediately available to researchers in the scientific community. In conclusion, we constructed a database to integrate and disseminate HAT-RAPD analysis data as a tool to identify and classify new species and sub-species of plants. Genetic relationships within sample species were comparable with historical accounts and by using this resource, two previously unidentified subspecies of Peuraria were characterized.

Acknowledgements

We would like to thank the Fulbright Foundation for providing a Fulbright Fellowship to Robert Cutler, which helped to financially support this research project, and to the Bard College Research and Travel Fund.

References

1. Anuntalabhochai S, Chundet R, Chiangda J and Apavatjru (2000) Genetic Diversity with in Lychee (*Litchi chinensis* Sonn.) based on RAPD analysis. *Acta Horticulturae* **575**, 253-9.
2. Atienzar, F, Evenden, A, Jha, A, Savva, D and Depledge, M (2000) Optimized RAPD Analysis Generates High-Quality Genomic DNA Profiles at High Annealing Temperature. *BioTechniques* **28**, 52-4.

3. Balasaravanan T, Pius PK, Raj Kumar R, Muraleedharan N and Shasany AK (2003) Genetic diversity among south Indian tea germplasm (*Camellia sinensis*, *C. assamica* and *C. assamica* spp. *lasiocalyx*) using AFLP markers. *Plant Science* **165**, 365-72.
4. Cain JC (1960). Miroestrol: an estrogen from the plant *Puerraria mirifica*. *Nature* **198**, 774-7.
5. Chen CX, Sun JS and Zhu LH (2000) Inheritance and sequence homology analysis of the maize DNA introgressed into the wheat double haploid plant through wheat×maize cross. *Acta Botanica Sinica* **42**, 728–31
6. Ellsworth DL, Rittenhouse KD and Honeycutt RL (1993) Artifactual variation in randomly amplified polymorphic DNA banding patterns. *Biotechniques* **14**, 214-7.
7. Felsenstein J (1993) PHYLIP (Phylogeny Inference Package) version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.
8. Harding K and Millam S (2000) Analysis of chromatin, nuclear DNA and organelle composition in somatic hybrids between *Solanum tuberosum* and *Solanum sanctae-rosae*. *Theoretical and applied genetics* **101**, 939–74.
9. Kaga A, Tomooka N, Egawa Y, Hosaka K and Kamijima O (1996) Species relationships in the subgenus Ceratotropis (genus Vigna) as revealed by RAPD analysis. *Euphytica* **88**, 17-24.
10. Osawa M, Kaneko M, Horiuchi H, Kitano T, Kawamoto Y, Saitou N and Umetsu K (2003) Evolution of the cystatin B gene: implications for the origin of its variable dodecamer tandem repeat in humans. *Genomics* **81**, 78-84.
11. Page RDM (1996) TREEVIEW: An application to display phylogenetic trees on personal computers. *Computer applications in the Biosciences* **12**, 357-8.
12. Rottenberg A and Parker JS (2003) Conservation of the critically endangered *Rumex rothschildianus* as implied from AFLP diversity. *Biological conservation* **114**, 299-303
13. Shimada T, Hayama H, Haji T, Yamaguchi M and Yoshida M (1999). Genetic diversity of plums characterized by random amplified polymorphic DNA (RAPD) analysis. *Euphytica* **109**, 143-7.
14. Sitthiphrom S, Chundet R and Anuntalabhochai S (2005) Investigation of Genetic relationship and hybrid detection in longan (*Dimocarpus longan* Lour.) by high annealing temperature RAPD. *Acta Horticulturae* In press.
15. Welsh, J and McClelland, M (1990). Fingerprinting genomes using PCR with arbitrary primers. *Nucleic Acid Research* **18**, 7213-8.
16. Williams JGK, Kubelik AR, Livak KJ, Rafalski A, and Tingey SV (1990). DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531-5.
17. Zhou A, Xia G, Zhang X, Chen H and Hu H (2001) Analysis of chromosomal and organellar DNA of somatic hybrids between *Triticum aestivum* and *Haynaldia villosa* Schur. *Molecular Genetics and Genomics* **265**, 387–93.