Controlling Illegal Logging in a Timber Species *Shorea platyclados*: A Case for DNA Markers

*Asif Javed Muhammad – Assistant Professor (Ph.D.) (Affiliated with the Department of Forestry and Range Management, University of Agriculture Faisalabad, Faisalabad, Pakistan)
Muhammad Bilal Zia – M.Phil. (Affiliated with the Pakistan Forest Institute (PFI), Peshawar Pakistan)
Atif Ali – Ph.D. (Affiliated with the Department of Forestry and Range Management, University of Agriculture Faisalabad, Faisalabad, Pakistan)
Nauman Ghafoor – M.Phil. (Affiliated with the Department of Forestry and Range Management, University of Agriculture Faisalabad, Faisalabad, Pakistan)
Izhar Ul Haq – M.Phil. (Affiliated with the Department of Forestry and Range Management, the University of Agriculture Faisalabad, Faisalabad, Pakistan)
Cannon Chuck – Director (Ph.D.) (Center for Tree Science, The Morton Arboretum, Illinois, USA)
Wickneswari Ratnam – Professor (Ph.D.) (Affiliated with the Faculty of Science and Technology, National University of Malaysia, Malaysia).

Abstract

Shorea platyclados is a Southeast Asian tropical montane species with great potential for timber utilization. There is no information about the genetic structure and population variability available for this species, yet the region harbors most of the unlogged populations of this timber wood species. The relative abundance and intactness of most of the natural populations offer a potential use of this species as a modal species for developing tests useful for illegal logging of timber and tracing the population of origin. In this study, we used six highly polymorphic microsatellite loci on 159 *S. platyclados* individuals. The ability to confidently identify or exclude a population as the source of an individual was exploited by using different well-established assignment methods i.e., Bayesian, Frequency, and Distance-based methods. Although several alternative methods have been developed for this purpose are yet to be fully evaluated with empirical data. Specimens from the unknown population (n=20), its known source population (n=34) and two other populations from the same geographical region

(Kelantan=54; Fraser=51) were genotyped for 6 loci. The results obtained from self-assignment of individuals from all four populations identified Bayesian and Frequency methods as the most consistent and accurate, which correctly assigned 85-96% of individuals. Performance was variable among the distance-based methods, with Das (shared allele distance), Nei DA, and CavalliSforza performing best, whereas Goldstein $(\delta\mu)^2$ consistently performing poorly. Using these methods, we attempted to determine the source of *S. platyclados* individuals. Results obtained based on Frequency and Bayesian methods assigned 50-55% of individuals to the source population, the remainder of individuals were either assigned to other populations or were not assigned to any of the sampled populations. Low levels of assignment observed reasoned mainly due likely to low levels of genetic differentiation ($Fst = 0.028 – 0.052$) because of a high rate of gene flow observed between contiguous populations. Population genetic differentiation seems directly related to the high accuracy of individual assignment. Strategies to improve assignment frequency and power of discrimination were further discussed.

Keywords: *Illegal Logging; DNA markers; Population Genetics; Timber Tracking; Timber Certification*

Introduction

Foresters have been leading the charge for 'sustainable' management certification and chain-of-custody technology. The ability to track timber resources from forest to furniture is critically important for successful management and proper regulation of the marketplace. The inability of tropical timber-producing countries to capture the actual value of their timber, both economically and ecologically, has long been identified as a driving force behind rapid conservation and over-exploitation. For these governments to be successful, legally binding transparent incentives must be provided to the 'fair' players in the marketplaces. Timber producers need to be able to make wise decisions at the marketplace. The ability to provide

accurate and verifiable information about the species identity and geographic origin of a log would be a powerful tool in this marketplace.

Timber trade is an important resource of the national economy in the Southeast Asian region. Malaysia is one of the mega centers of diversity and one of the largest exporters of timber wood. Illegal logging in Malaysia is one of the major causes of deforestation and forest degradation (Gani, 2013). This harvest primarily comes from 'wild' populations under federal or state ownership. It generally contains a high level of species diversity, encompassing many of the major groups in the flowering plants. Little is known about the genetics of these organisms. The geographic setting is a vast island archipelago, with an enormous amount of coastline, numerous ports, and lots of boat traffic. These factors add up quickly into a very complex situation. International Tropical Timber Organization has recently emphasized on phased approach to certification, the need for external 'third-party' verification of species identity and legal harvest was emphasized by trade officials, forest managers, and conservation groups (Nussbaum and Simula, 2005). The Forest Stewardship Council is currently issuing certificates for paper trails as the timber moves through the chain of custody. Most tracking technologies have proposed implanting a physical marker into the logs as they are harvested, and the markers can then be traced. This technology requires constant oversight, either by governmental or third-party agencies, at the point of harvest. The best type of marker would be something inherent to the material and could be sampled at any point along the chain of custody by producers, consumers, governmental bodies, or third-party agencies. DNA marker technology is the method of choice. The use of DNA fingerprinting in the timber trade has been discussed frequently and viable DNA molecules can be harvested from fresh, long-dead, and processed wood (Asif and Cannon, 2005., Deguillous et al., 2003., Gugerli et al., 2005).

Currently, many DNA markers are available, however simple sequence repeats (SSRs) are the markers of choice. They are attractive as they combine several features of the ultimate marker

system; highly polymorphic to allow precise discrimination of closely related individuals, co-dominant and multi-allelic, highly abundant, evenly dispersed across the genome, and efficiently amenable to a simple and rapid PCR assay. They have successfully been used in a wide range of studies including population and conservation genetics (Estoup and Angers, 1998., Beaumont and Bruford, 1999), gene flow within and between populations, mapping, and breeding (McCouch et al., 1997), and extent and maintenance of genetic diversity (Byrne et al. 1996) and for paternity testing (Dow and Ashley, 1996). Similarly, DNA markers can be used for species identification (Asif and Canon, 2005), verification of source, either at the regional scale (through phylogeographic methods) or concession (population genetic assignment), and for tracking individual logs or wood products. Asif and Canon (2005) reported the identification of an endangered tropical species (*Gonystylus bancanus*) using DNA markers and developed a DNA extraction method to extract DNA from processed wood thus increasing the scope of DNA testing for processed wood products. Recently, Lowe et al. (2010) successfully developed and implemented SSR based DNA marker system to control illegal logging in Indonesia by using Merbau (*Intsia palembonics*) a high-priced timber species, from logging concession to the mill, where the DNA profile of individual logs is difficult or impossible to falsify. They required 14 SSR markers to provide an exact genotype match between forest and sawmill samples. They proposed that DNA marker-based technology is suitable for broad-scale industry applications to track legally harvested timber and check for illegal substitution along supply chains. They are currently the first company offering DNA commercial timber testing services (Double Helix Tracking Technologies). For Dipterocarpaceae species nine different primer sequences were designed to amplify SSRs containing mostly perfect and imperfect CT repeats (Ujino et al., 1998). These sets of primers were shown to be useful across various *Shorea* taxa (Ujino et al., 1998) including *S. platyclados* (Asif et al., 2015).

Assignment of an individual to the population from which it most probably originated based on its multilocus genotype has been shown to have a diverse range of applications (Davies et al., 1999., Hansen et al. 2001., Manel et al., 2002., Waser and Strobeck, 1998). They have been extensively used in fisheries and animal populations for individual identification (Estoup et al., 1998., Hansen et al., 2001). Illegal logging is a major problem causing socio-economic issues in Pakistan. This has become a major cause of deforestation in Pakistan. Pakistan has only 2% land area covered by forests. The rate of deforestation is one of the highest in the world and it is still on the rise. The annual deforestation rate was reported to be 1.8% between 1990 and 2000, and 2.7% between 2010 and 2015 (FAO, 2015). The main objective of this study was to determine the potential of a new set of SSR markers as compared to one reported by (Ng et al., 2017) and the feasibility of DNA markers to differentiate between legal and illegal logs. Once developed this model can be applied to other tree species.

Materials and Methods

Shorea platyclados is widely distributed in the Malay Peninsula, Sumatra, and Borneo. In the Peninsula, it is found along the main range from Perak to Negri Sembilan and in the mountains of Kelantan, Trengganu, and Pahang. It is confined to the upper dipterocarp forest zone, having a higher altitudinal distribution than any other *Shorea*. It has been found as low as elevations of 1000 ft (Kelantan) and as high as 4,000 ft along the Main Range and usually appears toward the upper limits of the *S. curtisii*. The timber is a dark red meranti (Desch, 1941) its abundance entitles it to be a potential source of timber.

Sample Collection

A total of 159 individuals of *Shorea platyclados* (Dipterocarpaceae) from four populations were sampled at four different locations along the Titiwangsa mountain range encompassing the states of Kelantan, Perak, and Pahang in Peninsular Malaysia (Fig. 1). Fraser Hill (Pahang) is a small mountain range -situated on the southern end of the range, has long been used for

timber extraction. Two populations, i.e., PITC1 and PITC2 are from the northern end of the range contained within the PITC (Perak) concession area. The Forest Stewardship Council and Scientific Certification System have certified the PITC operation as a sustainable forest management unit. The third population (Kelantan) is also in the north and adjacent to a newly established protected area and has recently been utilized for timber extraction. The samples were all collected within a relatively narrow elevational band, from roughly 650-900 meters. Field identifications were based on the expertise of Malaysian forestry officers from all three states. Voucher samples including branches, leaves, seedlings, and fruits were also collected for further specimen identification.

DNA Extraction and Genotyping

Total genomic DNA was extracted from the inner bark of all 159 individuals using a modified CTAB protocol (Doyle and Doyle, 1987). After extraction crude DNA was purified with High Pure PCR Template Preparation Kit (Roche-Diagnostics). The genotype of each individual tree was determined by polymorphism of the microsatellite loci using six primer pairs (Shc01, Shc02, Shc03, Shc08, Shc09, Shc11) that were developed in *Shorea curtisii* (Ujino et al.,1998). We tested the nine primers and chose the six that showed polymorphism in *Shorea platyclados* in a preliminary experiment (Asif et al., 2016). Polymerase chain reaction (PCR) amplification was performed in 25 µL reaction volumes containing 5 - 10 ng genomic DNA, 1x supplied buffer, 5 pmol of each primer, 1.5 mM $MgCl_2$, 0.2 mM of each dNTP, and 0.5 units *Taq* polymerase (Eppendorf) using the Eppendorf Mastercycler Gradient Thermocycler. The standardized annealing temperature and the number of PCR cycles for the primer pairs, Shc01, Shc02, Shc03, Shc08, Shc09, and Shc11 were 54^{o}C, 55.2^{o}C, 54.2^{o}C, 51.1^{o}C, 52.2^{o}C and 55.2^{o}C for 30s and 45 cycles respectively. The denaturation and extension conditions were the same for all primer pairs and were 94^{o}C for 4 min denaturation followed by 45 cycles of 94^{o}C for 45s, and extension at 72^{o}C for 45s, respectively. The genotype was determined using a 24-

caillary Reveal Mutation Discovery System (Spectrumedix Inc.). Fragment sizes were determined using QuickSpec^{TM} software version 2.4. (Spectrumedix Inc.). Genotypes were also confirmed using denaturing PAGE gel electrophoresis as described by Kaemmer et al., (1997).

Statistical Analysis

Allele frequencies, the effective number of alleles (*Ne*), observed and expected heterozygosities (*Ho* and *He*) were calculated for four populations (Kimura and Crow 1964., Nei 1978) by using POPGENE 32 (Yeh and Boyle 1997). Deviations from Hardy- Weinberg expectations were examined for each locus by calculating Wright's fixation index (*Fis*) (Wright 1965). fixation indices were tested for significance by χ^2 tests (Li and Horvitz 1953). Differentiation among populations was assessed by calculating pairwise *Fst* (Weir and Cockerham1984) using FSTAT Version 1.2 (Goudet 1995). The statistical significance of the *F*st value was based on the Chi-square test (Workman and Niswander 1970). *S. platyclados* individuals were assigned using their most likely population of origin using the Cornuet et al. (1999) GENECLASS v. 1.0.02 program (http://www.ensam.inra.fr/URLB/geneclass/geneclass.html), which allow multiple estimation procedures. *The Bayesian method* computes the likelihood of a genotype in each population based on the probability density of population allele frequencies (Rannala and Mountain, 1997). *The frequency method* assigns a genotype to the population in which it is most likely to occur based on the allele frequencies in the candidate population (Paetkau et al., 1995). *The distance methods* assign an individual to the population showing the closest genetic relationship to it (Cornuet et al., 1999). For the distance methods, five different measures of genetic distance were used; the shared allele distance (DAS) of Chakraborty and Jin (1993) and minimum (Dmin) distance (reviewed in Nei's, 1987), the Cavalli-Sforza and Edwards' chord distance (Dc), (Cavalli-Sforza and Edwards, 1967) and the $(\delta\mu)^2$ of (Goldstein et al., 1995). These three different tests are hereafter referred to as The Bayesian, Frequency, and

Genetic distance methods respectively for clarity. Using these assignment methods, a probability of 'belonging' to each of the three potential source populations was calculated for every individual sampled from the PITC2 population. In order to compute this probability, the distribution of an individual's assignment criterion (likelihood or distance) was compared to the assignment criteria distribution based on 100000 simulated individuals for each potential source population (Cornuet et al., 1999). Individuals were then assigned to the population in which their probability of belonging was highest and in which it exceeded one of three arbitrary threshold values $P \geq 0.05$, $P \geq 0.01$, or $P \geq 0.001$. Using these threshold values, the probability of exclusion from other populations is therefore $P \geq 0.05$, $P \geq 0.01$, or $P \geq 0.001$, respectively.

Results

All six microsatellite loci analyzed were polymorphic and the genetic variation observed was sufficient to characterize different populations of *Shorea platyclados*. *S. platyclados* is a diploid (2n = 14) species having hermaphrodite flowers therefore the probability of natural hybridization with other contiguous populations is very high. Six markers used showed different numbers of alleles and high observed heterozygosity (Table 1). *Fst* values range from 0.01 to 0.11 and depend upon the gene flow (*Nm*). The variation and abundance of gene flow of all six markers depend upon the allele frequency in each population (Table 1). The highest gene flow was observed for Sch08, Sch09, and Sch01 respectively. A restricted gene flow was observed for Sch02 and Sch011, also characterized by low allelic diversity (Table 1). Genetic diversity measures for heterozygosity indicate that all populations have relatively high observed heterozygosity (Table 2) except PITC 1. The average heterozygosity (He) among the three potential source populations (Kelantan, Fraser, and PITC 1) was 0.833±0.096, an average of 9.97±4.97 alleles were detected at each locus (Table 2). At most of the SSR loci alleles were detected at frequencies of <0.10, both within and among three populations. Nevertheless, at the loci Shc02, some alleles showed frequencies >0.50. Unique alleles were observed at

relatively low frequencies ranging from 0.009 to 0.061. The fixation indices (*Fis*) calculated for all polymorphic loci showed positive values for three loci (Shc02, Shc08, and Shc11), in all populations (Table 3). *Fis* showed significant, positive, or negative, deviation from Hardy-Weinberg equilibrium in one locus (Shc11) at Kelantan, two deviations (Shc02, Shc11) at Fraser Hill, similarly two (Shc02, Shc11) at PITC 1 whereas only significant deviation was observed for Shc08 in PITC2 respectively. At the population level, a non-significantly different excess of homozygotes was observed, where PITC1, PITC2, and Kelantan showed an excess of inbreeding. The estimation of the frequency of null alleles (rb) in four populations ranges from 0.02 - 0.09, indicating null alleles contributed about 2 - 9% more than homozygosity. All four populations of *S. platyclados* were in HWE at all loci ($P > 0.05$) based on population mean *Fis* values across all loci (Table 3). *Fst*, or the standardized variance in allele frequencies between populations, is the measure of genetic distance between populations. Pairwise differentiation analysis among population values is given in Table (4). These four populations showed moderate levels of differentiation (Table 4), with an average *Fst* = 0.037±0.012, maximum differentiation was observed between Fraser and Kelantan (Table 4). All *Fst* values were significantly different from zero. The translocated population at PITC-2 showed significant differentiation from most of the reference populations except its source population PITC1 (*Fst* = 0.031). The average *Fst* calculated for reference populations was 0.037 which means approximately 3.7% of all variation found was portioned between populations which is moderately low. Since overall, *He* is 0.827, this means that the maximum *Fst* that is mathematically possible for this comparison is less than 0.173. Another way to look at differentiation is to estimate the mean number of migrants per generation (*Nm*). Mean expected migration, *Nm* value for all populations is 4.6 per generation is relatively high. *Fst* values among all the populations corresponded well to their geographical distribution i.e., distance and location.

Most assignment tests performed well in the absence of a significant threshold, where individuals were assigned to the population in which their probability was highest (Table 5). However, only Bayesian and frequency methods performed well compared to distance-based methods (> 85% correct assignment). Self-assignment tests using three methods i.e., the Bayesian method assigned individuals with high probability in all four populations (Kelantan 92.6%; Fraser 94.1%; PITC1 88.2% and PITC2 90.0%) followed by Frequency (96.3%; 94.1%; 93.8and 85.0%) respectively (Table 5). Among the distance-based methods, Nei standard distance (DA) only performed well (>80%). Self-assignment tests performed also assigned several individuals to other populations than the source population indicating genetic association of the populations. The genotypic profile for the twenty individuals from PITC 2 appeared variable with each individual having a unique multilocus genotypic profile. Allelic diversity was relatively low (Table 2) with an average of 10.2±2.4 alleles per locus (range 7-13). For the Bayesian and Frequency methods, 45-50% of individuals consistently had the highest probability of being derived from the PITC1 population (Table 6). When significance thresholds were applied to the assignment probabilities, most individuals from PITC2 were consistently assigned to the PITC1 population when P≥0.05 (Table 6). Results at the more stringent threshold (P≥ 0.01 and P≥0.001) were more consistent, where individuals were either assigned to other populations or failed to be assigned to any of the populations (e.g., P≥0.01). Nonetheless, only Frequency and Bayesian methods performed well with or without applying the stringent level. However, results from the more stringent threshold (P≥0.01 and P≥0.001) were less clear with most individuals either not assigned to any population or were assigned to multiple populations. In all cases, the probability of each individual having originated from PITC1 was higher. Both Frequency and Bayesian methods excluded most of the PITC2 individuals from having originated from Fraser and Kelantan populations (Table 7).

Discussion

Four populations were sampled from the two extremes (north and south tips) along the Titiwangsa mountain range. Precautious population selection was primarily practiced obtaining a reasonable differentiating power in the absence of prior genetic information on *S. platyclados* species. Microsatellite primer systems were successfully applied within the same species group in which they were developed, the primary system that was developed in *S. curtisii* (Ujino et al., 1998) worked successfully in *S. platyclados*. The number of alleles detected at all six loci was different than observed in 40 individuals of *S. curtisii* in a primary forest population (Ujino et al., 1998). These differences in allele numbers between species have also been found in phylogenetically closely related tropical Leguminosae species (Dayanandan et al., 1997). The observed heterozygosity observed over six loci in *S. platyclados* was closely related to *S. curtisii* over eight loci (Ujino et al., 1998). This discrepancy could also be attributed to different population sizes. These primers have been successfully used for paternity analysis in *N. heimii* (Konuma et al., 2000) with high paternity exclusion probability. This result indicates that these primers are applicable for individual assignment studies in dipterocarp species as amplification of microsatellites has been reported in other dipterocarp taxa (Ujino et al., 1998).

Overall mean heterozygosity estimated from six microsatellites among four *S. platyclados* populations reflects high genetic variability. Population statistics indicate that genetic variation between populations is partitioned homogenously. *S. platyclados* showed relatively high levels of genetic variability within populations compared to between populations and displayed little genetic differentiation among populations. Allozyme studies of *S. leprosula* populations in Malaysia revealed high levels of genetic diversity and most of the diversity was partitioned within populations (Lee et al., 2000b). Weak genetic structure and high similarity revealed the absence of genetic drift and major bottlenecks in the past. A low level of population differentiation and a high outcrossing rate were also reported in *S. leprosula* (Lee et al., 2000). The present distribution of genetic diversity within species is influenced by both nature and

evolutionary history. For example, populations of species with long-distance seed dispersal may have been founded by several individuals and are thus likely to contain a representative sample of genetic diversity. Furthermore, species with continuous distribution or the potential for long-distance pollen and seed dispersal may subsequently acquire genetic diversity from neighboring populations. Such species usually have relatively little population-to-population variation in genetic diversity. *S. platyclados* populations studied showed a large within-population variation compared to between and gene flow occurs through pollen compared to seed and genetic similarity among populations present in proximity may have acquired genetic diversity from neighboring populations through gene flow.

The individual assignment was achieved using GENECLASS, as it offers the advantage of using several test statistics simultaneously. Moreover, it is difficult to compare results derived using different statistical tools e.g., STRUCTURE. Comparing probabilities from the outputs of the two is difficult due to the different methodologies and assumptions required for each program. That is, the 'probability of belonging' estimated through GENECLASS is not directly equivalent to the 'posterior probability of an individual having originated' as estimated by STRUCTURE, because in the approach of Pritchard et al. (2000) the sum of the posterior probabilities that each individual has originated from the tested series of potential source populations will always be equal to one, whereas using the method of Cornuet et al. (1999), the sum of the probabilities of belonging for each individual need not equal to one. We achieved a relatively high degree of accuracy in assigning *Shorea* individuals to the population of origin using six polymorphic loci. When the performance of different assignment methods was compared, the frequency-based method, and Bayesian methods performed better as suggested by Cornuet et al. (1999). The Bayesian method was the most efficient, while the Goldstein et al., (1997) distance method showed a markedly low performance. This low efficiency may be explained by the fact that, as indicated by Goldstein and Pollock (1997), the $(\delta\mu)^2$ distance

performs better for largely divergent than for closely related populations. Among the distance-based methods only DAS performed well whereas others did moderately well as also reported by Cornuet et al. (1999). A possible explanation for this difference might be the distinct nature of the data analyzed. Cornuet et al. (1999) used simulated data, obtained from the assumption of exact Hardy-Weinberg proportions at all loci and linkage disequilibrium. Genotypes were sampled in our study from real populations. Hardy-Weinberg contrasts revealed deviations from equilibrium for several microsatellite' loci analyzed in our study. In this situation, distance methods, which don't rely on HWE assumption, may produce better results. For the PITC2 population, only 55% of individuals could be assigned to their population of origin with high probability whereas the remaining individuals were either could not be assigned to any population or to adjacent populations. The power of assignment tests depends on a number of factors i.e. genetic differentiation among populations, number of populations sampled, degree of polymorphism at the loci, number of loci studied, and sample sizes (Cornuet et al., 1999., Hansen et al., 2001). Two factors, i.e., population differentiation and the number of loci surveys per population, are worth mentioning. It is suggested that assignment accuracy positively correlated with genetic divergence among populations (Manel et al., 2001) All individual could be assigned to the reference population with 99.9% certainty for two highly differentiated populations (*Fst* = 0.15-0.2) when 10 loci (H > 0.6) and samples of 30 - 50 individuals are used per population. *Fst* values reported in this study range from 0.028-0.052 and observed heterozygosity (0.662-0.752), *Fst* is far lower than suggested for accurate assignment. Topchy et al. (2004) suggested inclusion of populations distantly related increased genetic differentiation and thus increased assignment accuracy. Adding more populations from Terengganu, Sumatra and Java in the future analysis would be useful to achieve higher differentiation. However, this strategy might be useful only if populations are closely related, sharing niche with high gene flow (low differentiation) pooled together and treated as a single

population. This would increase the overall probability of population exclusion as highly differentiated populations would be excluded as a potential source. Ng et al. (2017) recently used seven chloroplast and fifteen SSR markers in *Shorea platyclados* for forensic identification. They used independent assignment tests and assigned a high number of individuals to their population of origin. Chloroplast markers proved less effective as they have weak power of discrimination due to low mutation rates. High discrimination power can be achieved by adding more SSR markers. The set of markers previously used by Ng et al. (2017) can be used to increase discrimination power. However, there is no prior information about the exact number of markers required to increase discrimination power as the number of trees per population and number of populations will vary in a particular geographical region. Therefore. screening populations with more SSR loci will be more cumbersome and will strongly be hampered by the need to develop more loci as currently only a few primers are available for this genus. This perhaps would be a strong drawback of using STMS markers in these cases as a greater number of markers would not necessarily resolve population differentiation problems. The other option would be utilizing other highly polymorphic markers e.g., AFLP (amplified fragment length polymorphism). Campbell et al. (2003) compared the effectiveness of microsatellite-based method for individual assignment to that of AFLP. AFLP was utilized successfully even in cases with low levels of population differentiation. They reported 95% success in a case study on whitefish (*Coregonus clupeaformis*) using AFLP markers and this high fidelity is achievable under many situations. AFLP's made a promise, however, they are technically demanding and time-consuming. Blanc–Jolivet et al. (2018) successfully demonstrated the use of single nucleotide polymorphism (SNP) in resolving the geographical origin of *Larix* species. They also reported a high success rate of assignments, however, it became possible to generate many SNP markers in a short time. Genotyping by sequencing (GBS) emerged as a powerful tool for sampling genome-wide genetic variation and thus has

the potential to generate thousands of SNP markers in a short period of time (Sun et al., 2023). This approach has the advantage of being rapid, high throughput, cost-effective, and applicable to organisms without sequenced genomes. It has been increasingly applied to generate SNP genotype data for plant genetic and genomic studies (Asif et al., 2017., Fu and Yang, 2017., Lounder et al., 2023., Sun et al., 2023).

Conclusions

This study demonstrated the potential and problems of practical applications of the individual assignment procedures at various stages of timber production for log identification. Once various DNA markers are identified specific to populations, unknown individuals can be assigned to their origin by surveying a few loci. These tools can thus successfully be used for logging, wood certification, poaching, and smuggling of highly threatened timber species by using genetic markers coupled with assignment analysis on confiscated materials. Similarly, these tests can be performed to check and maintain the source of origin of timber along the supply line by running random sample tests. These strategies could result in solving a multitude of socio-economic and political problems by producing strong evidence/s against the law offenders to help curb associated problems.

Acknowledgments

We would like to express our greatest appreciation to the Perak and Pahang Forestry Departments, and the Economic Planning Unit of the Malaysian Government for assisting us in the collection of timber tree samples. Extraction of total DNA was carried out at UKM's Forest Genetics Laboratory in the School of Environmental and Natural Resource Sciences. SSR analysis was provided by the Centre for Biotechnology and Genomics at Texas Tech University, which acknowledges the TTU Department of Plant and Soil Sciences' investment in the Reveal Mutation Discovery System.

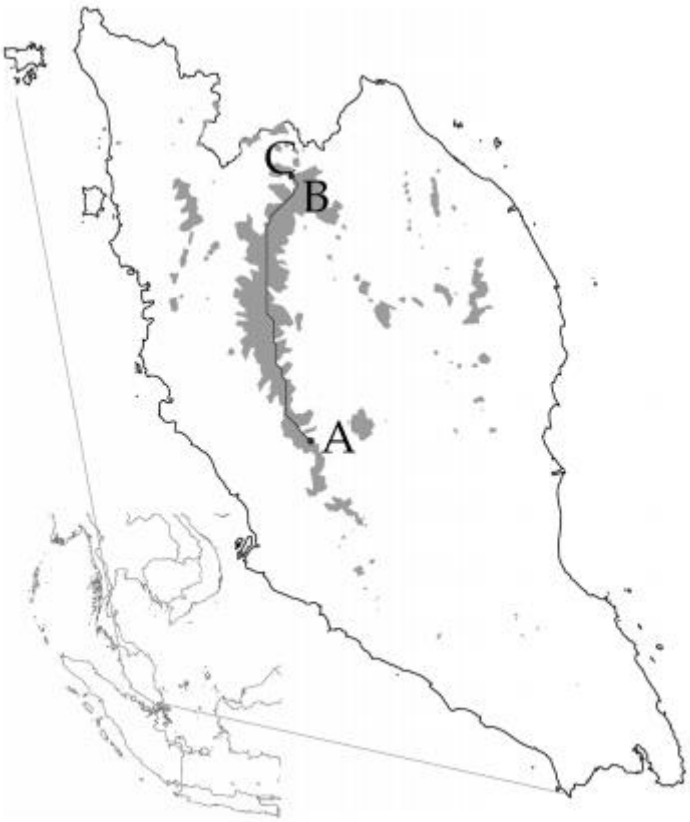

Figure 1. Map of Peninsular Malaysia showing sample locations and upland areas above 750m elevation (shaded in grey). Locations: A) Fraser's Hill; B) Belum; and C) PITC. The line shows the least cost-path between PITC and Fraser's (227 km long).

Table 1. Characteristics of microsatellite loci used in *Shorea platyclados*: basic information, genetic diversity, and population structure.

	°C	bp	*Ho*	*He*	*Fis*	*Fit*	*Fst*	*Nm*
Sch01	54	117-204	0.80	0.87	0.06	0.08	0.02	12.7
Sch02	55.2	132-166	0.63	0.81	0.20	0.29	0.11	2.7
Sch03	54.2	120-146	0.72	0.85	-0.08	-0.04	0.03	9.9
Sch08	51.1	279-340	0.79	0.88	0.25	0.26	0.01	21.1
Sch09	52.2	178-228	0.80	0.88	0.11	0.12	0.01	14.4
Sch011	55.2	159-182	0.76	0.88	0.39	0.45	0.10	2.9

°C – optimized annealing temperature; bp – size range of fragments in base pairs; *Fis* – inbreeding index in total population; *Fit* – inbreeding within populations; *Fst* – inbreeding among populations; *Nm* – gene flow estimated from *Fst*; estimates were calculated according to standard techniques (Raymond, Rousset 1995).

Table 2. Summary statistics for six microsatellite loci in four populations of *Shorea platyclados* from the high montane range of Malaysia.

Population	**N**	**A**	***He***	***Ho***
Kelantan	108	14.5±5.3	0.869±0.051	0.750±0.187
Fraser	102	13.7±6.0	0.804±0.155	0.752±0.163
PITC 1	64	11.7±3.6	0.827±0.081	0.662±0.280
PITC2	40	10.2±2.4	0.808±0.084	0.708±0.128
Average	157	12.5±1.96	0.827±0.029	0.718±0.042

Table 3. Fixation indices (*Fis*) in four *S. platyclados* populations based on six microsatellite loci.

Loci	Kelantan	Fraser Hill	PITC 1	PITC 2
Shc01	-0.012	0.093	0.103	0.083
Shc02	0.214	-0.221*	0.645**	0.147
Shc03	-0.051	-0.165	-0.101	0.107
Shc08	0.283	0.196	0.126	0.518**
Shc09	0.024	0.186	0.101	0.109
Shc11	0.441**	0.270*	0.523**	0.191
Mean	0.147	0.075	0.214	0.196

*P<0.05 and **P<0.01

Table 4. Population differentiation is represented by the fixation index (*Fst*) of *Shorea platyclados* genetic groups between each other. The range of *Fst* is from 0 to 1, 1 meaning that the two populations don't share any genetic diversity and are thus genetically identical.

	Kelantan	Fraser	PITC1
Fraser	0.029		
PITC1	0.028	0.050	
PITC2	0.036	0.052	0.031

Table 5. Assessing the performance of eight different assignment methods by using individuals from all four populations and assigning them to their population of origin (self–assignment). Frequency and Bayesian methods performed better than distance-based methods where a high percentage of individuals were assigned to the population of origin.

Method	Kelantan	Fraser	PITC 1	PITC 2
Frequency	Kelantan= 63.0% Fraser = 3.7% PITC 1 = 0.0% PITC 2 = 0.0%	Kelantan = 5.9% Fraser = 94.1% PITC 1 = 0.0% PITC 2 = 0.0%	Kelantan = 2.9% Fraser = 0.0% PITC1 = 93.8% PITC 2 = 2.9%	Kelantan = 5.0% Fraser = 5.0% PITC 1 = 5.0% PITC2 = 85.0%
Bayesian	Kelantan = 92.6% Fraser = 3.7% PITC 1 = 3.7% PITC 2 = 0.05	Kelantan= 5.9% Fraser = 94.1% PITC 1 = 0.0% PITC2 =0.0%	Kelantan= .8% Fraser = 2.9% PITC = 88.2% PITC2 =0.0%	Kelantan = 0.0% Fraser = 5.0% PITC 1 = 5.0% PITC2 =90.0%
DAS	Kelantan = 90.8% Fraser = 5.5% PITC = 3.7% PITC 2 = 0.0%	Kelantan = 9.8% Fraser = 78.5% PITC 1 = 1.9% PITC 2 = 1.9%	Kelantan = 1.8% Fraser = 2.9% PITC1 = 73.5% PITC 2 =11.8	Kelantan = 0.0% Fraser = 5.0% PITC 1 = 10.0% PITC2 =85.0%
Nei standard	Kelantan = 87.1% Fraser = 5.5% PITC = 3.7% PITC 2 = 3.7%	Kelantan =17.7% Fraser =0.0% PITC1 =80.4% PITC 2 =1.9%	Kelantan=14.7% Fraser = 5.9% PITC 1 = 67.6% PITC 2 = 11.8%	Kelantan = 5.0% Fraser = 5.0% PITC 1 = 15.0% PITC2 =75.0%
Nei minimum	Kelantan = 83.3% Fraser = 7.4% PITC 1 = 5.6% PITC 2 = 3.7%	Kelantan= 11.8% Fraser = 84.4% PITC 1 = 1.9% PITC 2 = 1.9%	Kelantan = 4.7% Fraser = 5.9% PITC 1 = 64.7% PITC 2 = 14.7%	Kelantan = 0.0% Fraser = 1.0% PITC 1 = 15.0% PITC 2 =75.0%
Nei DA	Kelantan = 90.7% Fraser = 5.6% PITC 1 = 3.7% PITC 2 = 0.0%	Kelantan = 5.9% Fraser = 94.1% PITC 1 = 0.0% PITC 2 = 0.0%	Kelantan = 8.8% Fraser = 0.0% PITC 1 = 5.3% PITC 2 = 5.9%	Kelantan = 0.0% Fraser = 10.0% PITC 1 = 10.0% PITC2 =80.7%
Cavalli-Sforza	Kelantan = 88.9% Fraser = 7.4% PITC 1 = 3.7% PITC 2 = 0.0%	Kelantan = 3.9% Fraser = 96.1% PITC 1 = 0.0% PITC 2 = 0.0%	Kelantan 11.8% Fraser = 2.9% PITC 1 = 76.5% PITC 2 = 8.8%	Kelantan = 5.0% Fraser = 0.0% PITC 1 = 10.0% PITC2 = 85.0%
$(\delta\mu)^2$	Kelantan = 38.8% Fraser = 16.7% PITC 1 = 1.9% PITC 2 = 42.6%	Kelantan= 22.2% Fraser = 35.3% PITC 1 = 3.9% PITC 2 = 37.3%	Kelantan 35.3% Fraser = 17.6% PITC 1 = 14.7% PITC 2 = 32.4%	Kelantan = 0.0% Fraser = 20.0% PITC 1 = 0.0% PITC2 =80.0%

Table 6. The percentage of individuals from PITC 2 correctly assigned to their population of origin (PITC 1) using eight different assignment methods and significance threshold criteria.

Threshold	F	B	DAS	Ds	Dmin	Dc	Cavalli-Sforza	$(\delta\mu)^2$
None	K=20 F=30 P=50 N=0	K=20 F=30 P=45 N=5	K=10 F=35 P=45 N=10	K=25 F=30 P=45 N=0	K=20 F=35 P=45 N=0	K=20 F =30 P =30 N =20	K=15 F=30 P=45 N=10	K=0 F=100 P=0 N=0
P≥0.05	K=20 F=30 P=50 N=0	K=20 F=25 P=55 N=0	K=5 F=25 P=50 N=20	K=25 F=25 P=35 N=15	K=15 F=25 P=30 N=30	K=10 F=25 P=20 N=45	K=10 F=25 P=20 N=45	K=0 F=100 P=0 N=0
P≥0.01	K=15 F=30 P=45 N=10	K=10 F=20 P=30 N=40	K=10 F=35 P=55 N=0	K=20 F=35 P=55 N=0	K=20 F=25 P=40 N=15	K=15 F=30 P=45 N=10	K=15 F=25 P=55 N=5	K=0 F=100 P=0 N=0
P≥0.001	K=20 F=30 P=50 N=0	K=15 F=30 P=45 N=10	K=10 F=35 P=50 N=5	K=25 F=30 P=45 N=0	K=20 F=25 P=45 N=10	K=20 F=30 P=50 N=0	K=15 F=30 P=55 N=0	K=0 F=100 P=0 N=0

F= frequency method; B = Bayesian method (Rannala & Mountain 1997); Das; shared allele distance; Ds = Nei's standard genetic distance; Dmin, Nei's minimum genetic distance; Dc= Cavalli-Sforza and Edwards chord distance; $(\delta\mu)^2$ Goldstein *et al. 's* genetic distance (see Materials and Methods for other references). K, F, P, and N were Kelantan, Fraser, PITC1 and failed to assign to any of the population. Where no threshold was set, individuals were simply assigned to the population where their probability of originating was highest. N Individuals not assigned to any population as probabilities were below threshold.

Table 7. Average probability for the twenty PITC2 individuals originating from each of the three potential populations using two assignment methods. The range of probabilities are given in parentheses.

	Frequency	**Bayesian method**
Kelantan	0.1370 (0.0144-0.4399)	0.0256 (0.0017-0.0913)
Fraser	0.1223 (0.0198-0.3931)	0.1442 0.0044-0.3860
PITC1	0.2024 (0.0015-0.6149)	0.2423 (0.0015-0.6456)

References

1. Asif, M. J., Abdullah, M.Z., Muhammad, N. and Ratnam, W., 2017. Detecting mislabeling and identifying unique progeny in Acacia mapping population using SNP markers. *Journal of Forestry Research*, *28*(6), pp.1119-1127.
2. Asif, M.J. and Cannon, C.H., 2005. DNA extraction from processed wood: a case study for the identification of an endangered timber species (*Gonystylus bancanus*). Plant Molecular Biology Reporter, 23, pp.185-192.
3. Beaumont, M.A., and Bruford, M.W., 1999. Microsatellites in conservation genetics. In: Goldstein DB, Schlotterer C, editors. Microsatellites: evolution and applications. Oxford: Oxford University Press. p 165–182.
4. Blanc-Jolivet, C., Yanbaev, Y., Kersten, B. and Degen, B., 2018. A set of SNP markers for timber tracking of Larix spp. in Europe and Russia. Forestry: An International Journal of Forest Research, 91(5), pp.614-628.
5. Bunney, E., McInerney, F.A., Dormontt, E., Malik, A., Welti, N., Wilkins, D., Plant, M., Hettiarachchi, D.S., Thomas, D., Dowell, A. and Hamalton, T., 2023. Safeguarding sandalwood: A review of current and emerging tools to support sustainable and legal forestry. *Plants, people, planet*, *5*(2), pp.190-202.
6. Byrne, B.M. and Shavelson, R.J., 1996. On the structure of social self-concept for pre-, early, and late adolescents: a test of the Shavelson, Hubner, and Stanton (1976) model. Journal of Personality and Social Psychology, 70(3), p.599.
7. Campbell, D., Duchesne, P. and Bernatchez, L., 2003. AFLP utility for population assignment studies: analytical investigation and empirical comparison with microsatellites. Molecular Ecology, 12(7), pp.1979-1991.
8. Cavalli-Sforza, L.L. and Edwards, A.W., 1967. Phylogenetic analysis. Models and estimation procedures. American journal of Human Genetics, 19(3 Pt 1), p.233.
9. Chakraborty, R. and Jin, L., 1993. A unified approach to study hypervariable polymorphisms: statistical considerations of determining relatedness and population distances. DNA fingerprinting: state of the science, pp.153-175.

10. Cornuet, J.M., Piry, S., Luikart, G., Estoup, A. and Solignac, M., 1999. New methods employ multilocus genotypes to select or exclude populations as origins of individuals. Genetics, 153(4), pp.1989-2000.
11. Davies, N., Villablanca, F.X. and Roderick, G.K., 1999. Determining the source of individuals: multilocus genotyping in nonequilibrium population genetics. Trends in ecology & evolution, 14(1), pp.17-21.
12. Deguilloux, M.F., Dumolin-Lapègue, S., Gielly, L., Grivet, D. and Petit, R.J., 2003. A set of primers for the amplification of chloroplast microsatellites in Quercus. Molecular Ecology Notes, 3(1), pp.24-27.
13. Desch, H.E., 1941. Dipterocarp timbers of the Malay Peninsula. Malayan forest records; no. 14.
14. Dev, S.A., Unnikrishnan, R., Prathibha, P.S., Sijimol, K., Sreekumar, V.B., Azhar Ali, A., Anoop, E.V. and Viswanath, S., 2023. Artificial intelligence in timber forensics employing DNA barcode database. *3 Biotech*, *13*(6), p.183.
15. Dow, B.D. and Ashley, M.V., 1996. Microsatellite analysis of seed dispersal and parentage of saplings in bur oak, Quercus macrocarpa. Molecular Ecology, 5(5), pp.615-627.
16. Doyle, J.J. and Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochemical bulletin.
17. Estoup, A., and Angers, B., 1998. Microsatellites and minisatellites for molecular ecology: theoretical and empirical considerations. Pp. 55–86 in G. Carvalho, ed. Advances in molecular ecology. IOS Press, Amsterdam
18. FAO. 2015a. Global Forest Resources Assessment. 2015. Rome. (available at www.fao.org/forest-resources-assessment/en).
19. Fu, Y.B. and Yang, M.H., 2017. Genotyping-by-sequencing and its application to oat genomic research. Oat: Methods and Protocols, pp.169-187.
20. Gani, I.Q.L.M., 2013. Current situation of illegal logging in Peninsular Malaysia. International Journal of Sciences, 2(06), pp.12-17.
21. Goldstein, D.B. and Pollock, D.D., 1997. Launching microsatellites: a review of mutation processes and methods of phylogenetic inference. Journal of heredity, 88(5), pp.335-342.
22. Goldstein, D.B., Ruiz Linares, A., Cavalli-Sforza, L. L., and Feldman, M.W., 1995. Genetic absolute dating is based on microsatellites and the origin of modern humans. Proceedings of the National Academy of Sciences, 92(15), pp.6723-6727.
23. Goudet, J.F.S.T.A.T., 1995. FSTAT (version 1.2): a computer program to calculate F-statistics. Journal of heredity, 86(6), pp.485-486.
24. Gugerli, F., Parducci, L. and Petit, R.J., 2005. Ancient plant DNA: review and prospects. New Phytologist, 166(2), pp.409-418.
25. Hansen MM, Kenchington E, and Nielsen EE, 2001. Assigning individual fish to populations using microsatellite DNA markers. Fish Fish Ser 2:93–112.
26. Kaemmer, D., Fischer, D., Jarret, R.L., Baurens, F.C., Grapin, A., Dambier, D., Noyer, J. L., Lanaud, C., Kahl, G. and Lagoda, P.J.L., 1997. Molecular breeding in the genus Musa: a strong case for STMS marker technology. Euphytica, 96, pp.49-63.
27. Kimura, M. and Crow, J.F., 1964. The number of alleles that can be maintained in a finite population. Genetics, 49(4), p.725.
28. Konuma, A., Tsumura, Y., Lee, C.T., Lee, S.L. and Okuda, T., 2000. Estimation of gene flow in the tropical-rainforest tree *Neobalanocarpus heimii* (Dipterocarpaceae), inferred from paternity analysis. Molecular Ecology, 9(11), pp.1843-1852.
29. Lee, S.L., Wickneswari, R., Mahani, M.C. and Zakri, A.H., 2000. Genetic diversity of a tropical tree species, *Shorea lepr*osula Miq.(Dipterocarpaceae), in Malaysia:

implications for conservation of genetic resources and tree improvement 1. Biotropica, 32(2), pp.213-224.
30. Lee, S.L., Wickneswari, R., Mahani, M.C. and Zakri, A.H., 2000. Mating System Parameters in a Tropical Tree Species, *Shorea leprosula* Miq.(Dipterocarpaceae), from Malaysian Lowland Dipterocarp Forest1. Biotropica, 32(4), pp.693-702.
31. Li, C.C., and Horvitz, D.G., 1953. Some methods of estimating the inbreeding coefficient. American Journal of Human Genetics, 5(2), p.107.
32. Lounder, H., 2023. Machine Learning and Genomics: Using a Novel Neural Network for Population Assignment of a Threatened Seabird, Leach's Storm-Petrel (*Hydrobates leucorhous*).
33. Lowe, A.J., K.-N. Wong, Y.-S. Tiong, S. Iyerh, and F.-T. Chew. 2010. Silvae Genet. 59: 263-268
34. Manel, S., Berthier, P. and Luikart, G., 2002. Detecting wildlife poaching: identifying the origin of individuals with Bayesian assignment tests and multilocus genotypes. Conservation biology, 16(3), pp.650-659.
35. Manel, S., Williams, H.C. and Ormerod, S.J., 2001. Evaluating presence–absence models in ecology: the need to account for prevalence. Journal of Applied Ecology, 38(5), pp.921-931.
36. Mascarello, M., Lachenaud, O., Amalfi, M., Smets, E., Hardy, O.J., Beeckman, H. and Janssens, S.B., 2023. Genetic characterization of a group of commercial African timber species: From genomics to barcoding. *Plos One*, *18*(4), p.e0284732.
37. McCouch, S.R., Chen, X., Panaud, O., Temnykh, S., Xu, Y., Cho, Y.G., Huang, N., Ishii, T. and Blair, M., 1997. Microsatellite marker development, mapping and applications in rice genetics and breeding. Oryza: from molecule to plant, pp.89-99.
38. Muhammad, A.J., Cannon, C.H. and Ratnam, W., 2016. Cross-specific amplification of microsatellite DNA markers in *Shorea platyclados*. Journal of Forestry Research, 27, pp.27-32.
39. Nei, M., 1978. Estimation of average heterozygosity and genetic distance from a small number of individuals. Genetics, 89(3), pp.583-590.
40. Nei, M., 1987. Molecular evolutionary genetics. Columbia university press.
41. Ng, Y.H., Lai, S.P., Su, Z.P., Yap, J.Y., Teoh, H.Q. and Lee, H., 2017. Factors influencing accounting students' career paths. Journal of Management Development, 36(3), pp.319-329.
42. Nussbaum, R. and Simula, M., 2013. The forest certification handbook. Taylor & Francis.
43. Oliveira, M., Azevedo, L., Ballard, D., Branicki, W. and Amorim, A., 2023. Using plants in forensics: state-of-the-art and prospects. *Plant Science*, p.111860.
44. Paetkau, D., Calvert, W., Stirling, I. and Strobeck, C., 1995. Microsatellite analysis of population structure in Canadian polar bears. Molecular Ecology, 4(3), pp.347-354.
45. Pritchard, J.K., Stephens, M. and Donnelly, P., 2000. Inference of population structure using multilocus genotype data. Genetics, 155(2), pp.945-959.
46. Rannala, B. and Mountain, J.L., 1997. Detecting immigration by using multilocus genotypes. Proceedings of the National Academy of Sciences, 94(17), pp.9197-9201.
47. Sneha, M.V., Madhushree, A.H., Ranjan, S.T., Divakara, B.N., Kumara, P.M. and Prabuddha, H.R., 2023. Genome sequencing and characterization of microsatellite markers of *Pterocarpus santalinus* Lf: an economically important endangered tree of Eastern Ghats, India. *Journal of Genetics*, *102*(2), p.35.
48. Sun, J.J., Xia, X.M., Wei, X.X. and Wang, X.Q., 2023. Tracing the geographic origin of endangered plant species using transcriptome-derived SNPs: An example of *Cathaya argyrophylla*. *Molecular Ecology Resources*, *23*(4), pp.844-854.

49. Topchy, A., Scribner, K. and Punch, W., 2004. Accuracy-driven loci selection and assignment of individuals. Molecular Ecology Notes, 4(4), pp.798-800.
50. Ujino, T., Kawahara, T., Tsumura, Y., Nagamitsu, T., Yoshimaru, H. and Ratnam, W., 1998. Development and polymorphism of simple sequence repeat DNA markers for *Shorea curtisii* and other Dipterocarpaceae species. Heredity, 81(4), pp.422-428.
51. Waser, P.M. and Strobeck, C., 1998. Genetic signatures of interpopulation dispersal. Trends in ecology & evolution, 13(2), pp.43-44.
52. Weir, B.S. and Cockerham, C.C., 1984. Estimating F-statistics for the analysis of population structure. evolution, pp.1358-1370.
53. Workman, P.L. and Niswander, J.D., 1970. Population studies on southwestern Indian tribes. II. Local genetic differentiation in Papago. American Journal of Human Genetics, 22(1), p.24.
54. Wright, S., 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution, pp.395-420.
55. Yeh, F.C. and Boyle, T.J.B., 1997. Population Genetic Analysis of Co-Dominant and Dominant Markers and Quantitative Traits. Belgian Journal of Botany, 129, 157.