Development and technical application of SSR-based individual identification system for *Chamaecyparis taiwanensis* against illegal logging convictions

Chiun-Jr Huang1,2,3, Fang-Hua Chu1, Yi-Shiang Huang4, Yu-Mei Hung3, Yu-Hsin Tseng2, Chang-En Pu3, Chi-Hsiang Chao3, Yu-Shyang Chou3, Shiau-Chian Liu6, Ya Ting You2, Shuo-Yu Hsu3, Hsiang-Chih Hsieh3, Cheng Te Hsu7, Meng-Yi Chen3, Ting-An Lin3, Hsin-Yi Shyu3, Yu-Ching Tu3 & Chi-Tsong Chen3

*Chamaecyparis taiwanensis* is an endemic plant suffering illegal logging in Taiwan for its high economic value. Lack of direct evidence to correlate stump and timber remains a hurdle for law enforcement. In this report, 23 polymorphic Genomic Simple Sequence Repeat (gSSR) and 12 Expressed Sequence Tag (EST)-SSR markers were developed and their transferability was assessed. The individual identification system built from selected non-linkage 30 SSR markers has a combined probability of identity as $5.596 \times 10^{-12}$ equivalents to identifying an individual in a population of up to 18 million *C. taiwanensis* with 99.99% confidence level. We also applied the system in an actual criminal case by selecting 19 of these markers to correlate illegally felled timbers and victim trees. Our data demonstrate that molecular signals from three timbers hit with three victim trees with confidence level more than 99.99%. This is the first example of successfully applying SSR in *C. taiwanensis* as a court evidence for law enforcement. The identification system adapted advanced molecular technology and exhibits its great potential for natural resource management on *C. taiwanensis*.

*Chamaecyparis taiwanensis* Masam. & Suzuki [= *Chamaecyparis obtusa* (Sieb. & Zucc.) Endl. var. *formosana* (Hayata) Hayata] (Cupressaceae) is a gymnosperm endemic in Taiwan. *C. taiwanensis* is endemic to Taiwan and is the dominant species in the conifer and broadleaf tree mixed forest, located in middle altitude region (from 1700 m to 2600 m) of Taiwan island1. The lowest latitude boundary of cypress’ natural distribution falls into Taiwan, suggesting a great significance in biogeography2. As an indispensable resource for making elegant buildings, furniture and handicrafts, these species play a vital role in serving wood source and timber industry. *C. taiwanensis* is well-known for their wood quality and expensiveness (4400 USD/m$^3$)(woodprice.forest.gov.tw), which often lead to endless illegal felling crimes. Therefore, developing individual identification system to *C. taiwanensis* is of more importance3.

Illegal felling remains a persistent problem in the timber producing countries all over the world. For decades, illegal logging endangered precious and valuable tree species such as cypress4, ash5, mahogany tree6, and Brazilian rosewood7 all over the world. In some cases, the law enforcement authorities, such as forestry police, arrest the suspects in time. However, lack of direct scientific evidence that correlate timbers to the stumps leads the conviction processes rather difficult and ineffective. Thus, the need of individual identification is critical to the forestry industry.

1School of Forestry and Resource Conservation, National Taiwan University, Taipei 10617, Taiwan. Biodiversity Research Center, Academia Sinica, Taipei 11529, Taiwan. Department of Forensic Science, Investigation Bureau, Ministry of Justice, New Taipei City 23149, Taiwan. Institute of Biological Chemistry, Academia Sinica, Taipei 11529, Taiwan. Department of Research Committee, Investigation Bureau, Ministry of Justice, New Taipei City 23149, Taiwan. Department of Applied Science, National Taitung University, Taitung 95092, Taiwan. Hualien Forest District Office, Forestry Bureau, Council of Agriculture, Hualien 97051, Taiwan. *email: d04625001@ntu.edu.tw; chen33039@gmail.com
The problem of illegal logging has been paid attention since 1995. More and more national and international regulations mandate tracking systems that ensure traceability on wood market\(^8\). Wood anatomy and dendrochronology are common visual identification method. The former is based on the anatomical characteristics to identify the wood, and can usually be identified to the genus\(^11\); the latter is often used to illustrate past climates, but may also provide the age and origin of the trees\(^25\). Compounds synthesized by trees and other plants are often called phytochemicals and are often used to identify species or distinguish genera. Intraspecific variation can also be detected in some species through some chemical analysis such as mass spectrometry\(^12\). Near infrared spectroscopy\(^14\), detector dogs\(^15\), stable isotopes\(^16\), and radiocarbon\(^17\). Genetic analysis can provide species-level identification, which is usually achieved by DNA sequence polymorphism\(^18\). Simple sequence repeats (SSRs) and Single nucleotide polymorphisms (SNPs) can be used to identify individuals and can be used in population genetics or systematic geography to determine the geographical region of origin within a species\(^19\). DNA fingerprinting is built into each organism itself and cannot be forged\(^20\). When enough markers are developed, in principle every individual has its own unique DNA fingerprint. DNA fingerprinting has the potential to track wood products independently within complex global supply network\(^21\). Theoretically, DNA fingerprinting is the only forensic wood identification technology that could be used to connect seized timber to illegally felled stumps\(^8\).

SSR is the most common marker used in individual identification for its short length, high polymorphism, easy polymerase chain reaction (PCR) amplification, high reproducibility, and high sensitivity\(^20,22,23\). SSRs are divided into two broad categories by different sources: Genomic (g)-SSR and expressed sequence tag (EST)-SSR\(^24\). gSSR markers have been derived from amplified genomic libraries. EST-SSRs are markers mined from EST sequence collections. gSSR markers have been reported to be more polymorphic when compared with EST-SSR in gymnosperms\(^6\) and crops\(^22,26\) because of a more diversified nucleotide sequence. Since the development of high-throughput sequencing technology, the marker development technique has been continuously advanced. Wang et al., 2018 published the first report on gSSR developed by De novo genome sequencing\(^27\). In contrast, EST-SSR, derived from the expressed sequence, is fast-acting, cost-effective and labor-saving alternative for non-model organisms\(^24\). Because of the conservative nature in gene coding regions\(^24\), newly developed EST-SSRs usually can be transferred in closely related species for marker development. The first EST-SSR based on Illumina-based de novo transcriptome was also published by Zhou et al. in 2018\(^28\). A study to develop both markers would avail of their merits and functions simultaneously.

For *C. taiwanensis*, evaluation of genetic variation or population structure is necessary for its preservation\(^2,29\) because this species is used extensively. After mid-twentieth century, the number of *C. taiwanensis* plunged, which also led to a significant decrease in both genetic variation and population structure. As an important tool for genetic and subsequent breeding, SSR markers are helpful for breeding polymorphic maternal plants and increasing the diversity of progeny. The objective of this study is to establish a scientifically valid SSR mediated individual identification system for *C. taiwanensis* in order to provide court evidence to link the seized wood and the victim tree, and to provide traceability proof for wood supply network. In the beginning of the research, we used Next Generation Sequence (NGS) technology to establish the DNA and RNA libraries of *C. taiwanensis* to accelerate the development of gSSR and EST-SSR markers. A total of 96 samples from four populations were used to evaluate the polymorphism, discriminative power, and random match rate of the selected SSRs. The linkage disequilibrium between markers was calculated to estimate the availability and credibility of the individual identification system. In this study, we successfully linked 3 stolen timbers back to 3 victim trees (case number MJIB-DNA-1080413 combine 1080328), marked the first successful application of *C. taiwanensis* individual identification system. Finally, our work would deter illegal felling toward these precious species by manifesting law enforcement effectively.

**Result and discussion**

**Developing *C. taiwanensis* individual identification system.**  Choice of template and library preparation. The gSSR are characterized by high polymorphism and is suitable for developing individual identification markers. The EST-SSR are highly conservative which could be used for developing markers to categorize species and populations\(^20,22,23\). In this study, both DNA and RNA libraries were constructed simultaneously as gSSR and EST-SSR markers, respectively (Fig. 1, Supplementary Sect. 1). From the three DNA libraries and from a RNA library prepared for the study, the sequences were compared between individual plants as well as between groups (Supplementary Sect. 1). With these two nucleic acid markers, we envisioned to differentiate samples within or among species.

**Nucleic acid sequencing and analysis.**  Next-generation sequencing technology enables the possible procurement of large number of sequences in a short time. In this study, we used the Illumina MiSeq platform (2 × 300 bp) to sequence the DNA and RNA libraries (Fig. 1). A total of 13,651,578 and 11,763,646 raw reads were produced from DNA and RNA libraries, respectively. The raw reads were deposited in the NCBI Sequence Read Archive (PRJNA506084). The sequences were then subjected to quality-trimming and merging and afterwards 4,236,284 contigs of the DNA pool and 4,392,534 RNA contigs were assembled. The base lengths of contigs ranged from 120–579 and 120–529, at an average of 420 for DNA and RNA, respectively. According to the work published by timber researchers\(^23,30\), the nucleic acid markers with fragment lengths of around 250 bases best meet our research goals. The lengths of contigs derived from the four libraries we have prepared were found to be suitable for screening markers within 250 bp length. A target band size below 250 bp implies a higher PCR success rate as the DNA of wood samples from seized timber and victim trees were mostly severely degraded.

**SSR discovery and primer design.**  A sum of 318,153 gSSR and 63,390 EST-SSR candidate sequences were screened by Simple Sequence Repeat Identification Tool (SSRIT)\(^31\) (Fig. 1). The proportions of SSR in the
genomic DNA and RNA libraries were 7.51% and 1.44%, respectively. Study by Squirrell et al. suggests that the overall success rate of SSR marker development is about 10%. With PCR, polymorphic high-quality markers could be successfully amplified resulting to a good peak pattern quality with little stuttering and absence of non-

Figure 1. Flowchart describing the procedure of developing SSR markers and aligning illegally-felled timbers to victim trees of *C. taiwanensis*. (a) 35 SSR markers specific to individual identification were selected from the DNA and RNA libraries of *C. taiwanensis*. The cumulative random matching rate of the system reaches $CP_i = 5.596 \times 10^{-12}$, which can be used to identify 18 million individuals with a credibility of 99.99% (b) 11 seized timbers were compared with 7 victim trees, and 3 timbers were matched with 3 victim trees successfully. The values of credibility in all matched cases were over 95%. (N number of individuals, P number of populations).
amplifying (null) alleles and other factors. Therefore, about 90% of the designed markers could be screened out. We designed a total of 395 gSSR and 105 EST-SSR primer pairs for testing in C. taiwanensis.

**Marker validation.** From the PCR results, 23 gSSR and 12 EST-SSR markers with polymorphism were selected (Fig. 1, Tables 1, 2), and the success rate for gSSR and EST-SSR marker was found to be 5.82% and 11.42%, respectively. Our data showed that it is easier to select SSR markers from the RNA library than from the DNA library, which is akin to previous studies. Other reports suggest SSR occurs more frequently in EST sequences than in the genome. In addition, the fact that the information content in EST is markedly lower than that in the genome promotes the calculation and analysis of EST in silico.

The samples used in marker validation came from 4 ethnic groups (TP, SY, DS, FR), with 20 to 30 individuals in each group (N = 25, 29, 21, 21), qualified the basic requirement of at least 15 individuals per group and 3 groups per study (Fig. 1, Supplementary Sect. 1). Among the 96 individuals sampled in this study, the number of alleles per gSSR is between 2 and 14 with 6.5 in average, whereas the number of alleles per EST-SSR is between 2 and 16, 7 in average (Tables 3, 4). The levels of Ho are from 0.000 to 0.802 and 0.021 to 0.604, with average of 0.399 and 0.379, respectively. The levels of Hs of gSSR and EST-SSR are ranged from 0.041 to 0.833 and 0.205 to 0.872, with average of 0.488 and 0.528, respectively. Significant (P < 0.001) deviations of Hardy–Weinberg equilibrium (HWE) in terms of heterozygosity deficiency were detected in 9 gSSR loci: CoTW76, CoTW77, CoTW539, CoTW545, CoTW554, CoTW555, CoTW561, CoTW585 and CoTW595 (9/23 = 39.13%) and also in 6 EST-SSR loci: CoTW383, CoTW502, CoTW511, CoTW513, CoTW514 and CoTW528 (6/12 = 50%). The levels of PIC of gSSR and EST-SSR are ranged from 0.058 to 0.821 and 0.187 to 0.858, with average 0.459 and 0.482. The levels of PD from 0.041 to 0.749 and 0.205 to 0.885, with average 0.494 and 0.555. The levels of PE of gSSR and EST-SSR are ranged from 0.000 to 0.479 and 0.000 to 0.312, with average 0.169 and 0.180. The levels of Pd of gSSR and EST-SSR from 0.029 to 0.939 and 0.114 to 0.794, with average 0.505 and 0.443. Two EST-SSR markers, CoTW383 and CoTW581, have putative functions found by BLAST hit (Table 2). Heterozygosity, being one of the first parameters that appear often in a data set, reveals lot of information including population structure and other historical clue. High heterozygosity means a lot of genetic variation, whereas low heterozygosity means almost no genetic variation. The heterozygosity data echo the results of PIC, PD and PI, suggesting that the marker developed in this study has moderate genetic variation. In addition, most of these markers show Hs > Hs (except CoTW495, CoTW556, CoTW559, CoTW598, CoTW424), which suggests that the population of C. taiwanensis is an inbred. A total of 15 sets of SSR marker used in this study deviated from HWE, which suggest the population may be not under the ideal status of HWE. The reason for this deviation could be artificial selection, non-panmixia or genetic drift. Generally, EST-SSR markers are less polymorphic than gSSR in plants because of high conservation in transcribed regions. Moreover, other factors such as SSR motif type, sample size, population and species may also differentiate gSSR and EST-SSR markers. However in this study, in terms of polymorphism and cross-species transferability, there was no significant difference between gSSR and EST-SSR groups (Supplementary Sects. 2, 3), but the difference rather occurred among individual markers. This fact might be explained by polymorphism and detection limit as markers with higher PD are often selected for individual identification. Also in our study, the differences in polymorphism and cross-species transferability between gSSR and EST-SSR are not significant, but those among markers are significant. It might be because of the giant genome size in taxa Chamaecyparis (20.03–27.40 pg/2C) which leads a deviation from random sampling in marker selection. When using the system to perform individual identification assay, a marker with higher PD should be considered as priority.

**Probability of identity and power of discrimination analysis.** Continued multiplication can be used to calculate the cumulative random probability of identity (CP) and the combined power of discrimination (CPD) for non-linked markers, where CP is the probability of two individuals most likely the same genotype, CPD is the probability of individuals being identified, and CP + CPD = 1. The credibility of the system is calculated based on “Random match probability in population size and confidence levels” published by Budowle et al. Confidence levels = (1 – CP) where N = Population size.

While applying the system in criminal cases, for the sake of objective and impartiality, practically the court will use the credibility of 95%, 99%, or 99.99% as acceptence criteria (Wall 2002, ISO/IEC 17025). In this study, only one marker in the same linkage group is used for CP analysis, and up to 30 markers can be continuously accumulated (Table 5). Also, the individual identification system’s CP is as small as 5.596 × 10^-12, and the CPD is as high as 0.99999999994040 (extremely close to 1). Applying the court’s strictest credibility standard of 99.99%, when the number of markers reached up to 30, the system can identify 18 million individuals, which actually exceed the whole C. taiwanensis population of 7.39 ± 0.73 million in Taiwan. While applying to the lowest acceptable credibility standard of 95%, the system could identify at least 2300 plants with a minimum number of 6 markers (Table 5).

**Aligning seized timbers to victim trees.** In this case (MJIB-DNA-1080413 combine 1,080,328), we successfully matched 3 seized timbers back to 3 victim trees by using 19 pairs of non-linkage SSR markers (Fig. 2, Table 6, Supplementary Sect. 4). The credibility values of the 3 cases are all above 99.99%. In our experiments, DNA samples were extracted twice or more from each sample in order to optimize the DNA concentration. Since 2007, forestry researchers have noticed that molecular markers can be used to provide direct evidence linking stolen timber and victim trees. Although many techniques have developed for extracting DNA from fresh and dried leaves (including published literature and commercial reagents), yet few studies have reported on extracting DNA from dried wood, which is still considered the most challenging part in this field of research. In forensic science field of study, it has been established that the validated DNA concentration range is between 20.03–27.40 pg/2C, which leads a deviation from random sampling in marker selection.
| Locus | Primer sequences (5′–3′) | Repeat motif | Fluorescent label | Allele size (bp) | Tm (°C) | GenBank accession no. |
|-------|--------------------------|--------------|------------------|-----------------|----------|----------------------|
| CoTW76 | F: TCTCATTCA GTTGGTATGGTT R: TCATCTTCA CGAACAAAGA | (TATC)_{16} | JOE | 161 | TD58-55 | MK213959 |
| CoTW77 | F: TGACGGTTC AATTTTTTGG R: AAAGAAAAG TTGCAATGGT | (TATC)_{7} | FAM | 170 | TD58-55 | MK213960 |
| CoTW99 | F: GGGAGCTGT AGGGAGATGAA R: ACATTGCCAA ATAGGGGATG | (TATG)_{6} | JOE | 165 | TD58-55 | MK213961 |
| CoTW314 | F: TTGTGCATGCAAAAGATCA CTA R: GGGCATGAT AATGAAAAGATG | (ACA)_{2} | FAM | 105 | 52.5 | MK213962 |
| CoTW330 | F: CCAGGTGCTTACGGGATGAG R: TCTTCCACTC TAGACCTTAGT TTTAT | (GAT)_{10} | FAM | 128 | 52.5 | MK213963 |
| CoTW337 | F: CCACCTTGT ACTCTGATCCCTC R: GATCCACTA AAGCCCTTTCCT | (CAT)_{13} | FAM | 107 | 57.5 | MK213964 |
| CoTW349 | F: GCTTGTCA TTTTTCCATT R: TCACTGGCA TACCACAAACT | (CTT)_{10} | FAM | 91 | 57.5 | MK213965 |
| CoTW495 | F: TTCCGAAAT CACATTGACAA R: TCACCTTCCT TCAAAAGGT TGA | (ATAC)_{4} | FAM | 160 | 45 | MK213966 |
| CoTW531 | F: CCACCCAC ATTTACAAAAATA R: TTTTGGGCT TTTTGAATAGG | (AAAT)_{5} | FAM | 166 | 55 | MK213967 |
| CoTW539 | F: AACCTCTTC CAACTTTGTAAT R: TACAGTTTT CTTGTCATGCA | (AAAT)_{5} | JOE | 152 | 60 | MK213968 |
| CoTW545 | F: GGAGGACAA GTCGACACCCAAATC R: CATAGTTGG GTTTCCACCTTC | (AAAT)_{5} | JOE | 151 | 60 | MK213969 |
| CoTW554 | F: ATTTAAAA GCATAACCAACAC R: TCAAGCTTAG AGCTGGTCAAG | (AAAT)_{5} | JOE | 139 | 55 | MK213970 |
| CoTW556 | F: GGAGGAGTAC CTTGCTTTATCCCT R: TTTAGGATT GCCGATATT TAGA | (AAAT)_{4} | JOE | 153 | 55 | MK213971 |
| CoTW559 | F: CAACCTGAC TAGAGGACAAAA R: TGTTCACTGT AGCTCACTCTA | (AAAT)_{4} | JOE | 150 | 57.5 | MK213972 |
| CoTW561 | F: AAAAAAGG ATCTAATGCCATA R: GGCTCCTTATTGTTGGATAA | (TAAT)_{4} | JOE | 169 | 55 | MK213973 |
| CoTW582 | F: CTAATGGGAC CTGATTTTCTAG R: GGGAAAAACA CATAGCATAA | (AC)_{2}(ATGT)_{2} | JOE | 200 | 57.5 | MK213974 |
| CoTW585 | F: TCTTCTAAA CTAATGACCA AGGA R: AGGGAAGCT ATCCTTAGAGA | (CA)_{3}(ATGT)_{4} | NED | 160 | 57.5 | MK213975 |
| CoTW588 | F: CAGGTCCCT GTAAAACCCCTC R: TGGGTCGAC ATACACACAT | (AT)_{2}(TATG)_{3}(ATGT)_{3} | NED | 167 | 57.5 | MK213976 |

Continued
felling is the most valuable impact of this study: DNA types of these precious trees have been filed. The illegal individual identification system which could be applied to various precious species. A warning to forestall illegal that can be used directly as court evidence in illegal felling cases. This is the first time study reporting the SSR

| Locus     | Primer sequences (5’-3’)                      | Repeat motif                                                                 | Fluorescent label | Allele size (bp) | Ta (°C) | GenBank accession no. |
|-----------|-----------------------------------------------|-------------------------------------------------------------------------------|-------------------|------------------|---------|----------------------|
| CoTW595   | F: ATGTATGCA GGTGATATGTG R: CCCCCTTGCG CTCTTATCTA | (TGTA)$_{13}$                                                               | NED               | 125              | 60      | MK213977             |
| CoTW597   | F: GGTAGTGATGTA GGTATGTGG R: TGATGGACC CTATAGGTGG | (GTAT)$_{4}$(ATGT)$_{1}$(TATG)$_{1}$                                        | NED               | 169              | 55      | MK213978             |
| CoTW598   | F: TCTTGTCTCT TCAATTATCGTG R: GCGAAAGTGA GATTACCAAT | (ATGT)$_{1}$                                                              | NED               | 264              | 55      | MK213979             |
| CoTW599   | F: TGCACAT TGGAAATGGACT R: TACATGGT GGATGCTCTTA  | (GT)$_{1}$(ATGT)$_{1}$(ACAT)$_{1}$                                          | NED               | 166              | 60      | MK213980             |
| CoTW600   | F: TGTTGATGT TTAGTGT GAATTTG R: ACAAAATCT TCTTTATCAACCG  | (TGATG)$_{1}$(ATGT)$_{1}$                                                    | NED               | 184              | 55      | MK213981             |

Table 1. Characteristics of 23 gSSR loci developed in Chamaecyparis taiwanensis. Ta annealing temperature, TD touchdown PCR.

0.625 and 10 ng/μL. 46. False negative result cannot be ruled out from over-concentrated sample and vice versa. Therefore, it is necessary to extract DNA two or more times for dry timber, as abovementioned, because its DNA extraction is challenging. Several studies suggested that the error rate increases along with PCR cycles47,48. Base misincorporation incurred by PCR occurs randomly throughout the sequence without hot spots48. The probability of base misincorporation is $1.85 \times 10^{-5}$ per base per cycle48. After comparing the results of positive and negative endpoint, we discovered 36 cycles is the upper limit which leads to positive PCR product without false-positive result. From comparing the results of positive and negative endpoint, we discovered 36 cycles is the upper limit which leads to positive PCR product without false-positive result. Therefore, the cycles were controlled below 35 cycles in our study, but not increasing cycles without limit. In addition, the SSR types of each marker were analyzed at least twice with ABI 3130XL. Signals below 150 RFU peak height threshold were considered not detectable. We developed a protocol of two sessions of instrumental operation and setup threshold value from pilot test result for illegal felling investigation cases. By comparing the profiles from positive and negative controls with test samples, we can obtain objective data with least erroneous possibility to conclude our investigation for court evidence.

Thirty-seven victim trees were reported by Luodong Forest District Office in December 2018. According to census data, the crime scene forest area is 281.03 hectare and the density of C. taiwanensis is 16 ± 1.6 individuals/ hectare. However, in order to protect suspects’ rights, we took an excess of the maximum possible population size: 10,000 into the calculation. Among 22 samples in this case, 7 succeeded in analysis, which is, by our definition, showing positive result in just 35 PCR cycles. The rest were denoted “Not detected” due to low positive PCR result (all tests comply the standard of accredited laboratory ISO/IEC 17025) or CL < 95%. It is worthwhile to note that seized illegally-felled timber 6TC matched the victim tree 6 TB (CPI = 3.342 × 10$^{-13}$, CPD = 0.999999999999666, CL = 99.999999%). In addition, seized illegally-felled timber 7TC matched with victim tree 7TA and 7 TB (CPI = 1.631 × 10$^{-13}$, CPD = 0.999999999999837, CL = 99.9999999%), and seized illegally-felled timber 8 TB matched with victim tree 8TA (CPI = 4.468 × 10$^{-10}$, CPD = 0.999999999553151, CL = 99.999532%). In this individual identification system test case (Table 6, Fig. 2, Supplementary Sect. 4), the minimal amount of matching marker was 17 among the positive-matched groups (CL = 99.999532%). The credibility increased along with the matching marker amount. The credibility is dependent on population size and matching marker amount. In addition, while aligning the evidence to the victim individuals, it is a common scenario that the sample DNA might have been degraded. Successful extraction is one of the crucial steps to identify same individual using DNA matching techniques. The extractable DNA in desiccated timber is low in quantity and poor in quality. The extracted DNA can only be used for individual identification using markers developed for specific species. In this regard, SSR marker is a traditional marker for individual identification, which has been widely used in human and gradually extended to other species. All the SSR markers designed in this study are shorter than 300 bp, which would be suitable for amplifying the lysed DNA fragments from desiccated timbers. Although SSR marker has the merits above mentioned, the overall success rate of DNA extraction and genotyping from timber is relatively low (31.81%, 7 out of 22 samples tested successfully). The low quantity and quality of DNA in timber sample might have limited our success rate. An improvement on DNA extraction method would enhance our success rate on timber samples. In addition, increasing the number of SSR markers capable of individual identification would decrease the overall CPD and increase CL. Overall, we provide scientific proof that can be used directly as court evidence in illegal felling cases. This is the first time study reporting the SSR individual identification system which could be applied to various precious species. A warning to forestall illegal felling is the most valuable impact of this study: DNA types of these precious trees have been filed. The illegal
The felling crime rate is dropping after public propagation of cypress individual identification system. Moreover, the individual identification system would also provide certificate for legal timber trading. This system would also deter dishonest businessman piggybacking illegal material in legal timber auction, which would further forestall illegal logging. In addition, these markers can be also used in population genetic analysis studies that facilitate the conservation and breeding of *C. taiwanensis*.

**Conclusions**

In this study, we developed an individual identification system for *C. taiwanensis* and provided the scientific evidence. This methodology can be adopted by the courts to link seized timber and victim trees. The *C. taiwanensis* individual identification system of this study includes 23 gSSR and 12 EST-SSR markers revealing polymorphism. When the 30 non-linkage markers were applied to *C. taiwanensis* identification, the lowest CPI was 5.596 × 10⁻¹² and the highest CPD was 0.99999999994404, which was sufficient to identify 18 million random samples of *C. taiwanensis* (CL = 99.99%). While applied in the criminal cases of *C. taiwanensis* illegal logging, this SSR marker system successfully matched five seized illegally-felled timbers to three victim trees with minimal 99.99% CL. To the best of our knowledge, this is the first time the SSR technology is being applied to provide molecular evidence for court conviction on *C. taiwanensis* illegal logging. Our study would provide not only the scientific evidence correlating seized timber and victim tree, but also could inherent unique serial number to identify every single *C. taiwanensis* timber. We demonstrated the feasibility of matching seized/legally-felled timber with victim tree by modern SSR technology, which would prevent illegal logging by warning the criminals that the woodland trees

### Table 2. Characteristics of 12 EST-SSR loci developed in *Chamaecyparis taiwanensis*. $T_a$ = annealing temperature.

| Locus   | Primer sequences (5′–3′) | Repeat motif | Fluorescent label | Allele size (bp) | $T_a$ (°C) | GenBank accession no | Putative function [organism] |
|---------|--------------------------|--------------|-------------------|------------------|-----------|----------------------|-----------------------------|
| CoTW383 | F: GATGACCAGCACTATTTG    | (GAAGG)$_{10}$ | FAM               | 113              | 57.5      | MK213982             | Hypothetical protein (chloroplast) *Hydrodictyon reticulatum* |
|         | R: GATCAAAATGAGCCCTAAT   |              |                   |                  |           |                      |                             |
| CoTW407 | F: AGCAGACAGCAGTTATGAGAG | (GATG)$_{5}$  | FAM               | 75               | 55        | MK213983             | No hit                      |
|         | R: TCTATTTTCTGAGTCATTCA  |              |                   |                  |           |                      |                             |
| CoTW409 | F: TTGAGTGGGAACACAGAGAG  | (ATAG)$_{10}$| FAM               | 81               | 60        | MK213984             | No hit                      |
|         | R: AACAGAACCACATACATTC   |              |                   |                  |           |                      |                             |
| CoTW420 | F: CATCTAGTGCTGACCACAAG  | (ATAG)$_{6}$  | FAM               | 101              | 55        | MK213985             | No hit                      |
|         | R: CAGGAGACAGATTACAG     |              |                   |                  |           |                      |                             |
| CoTW424 | F: CACTGGTGTACCTTGAAC    | (ATCT)$_{10}$(AT)$_{3}$ | FAM | 100              | 55        | MK213986             | No hit                      |
|         | R: CAACACATCATGGTCAAGCA  |              |                   |                  |           |                      |                             |
| CoTW502 | F: GTTGTAGTGGTTTATTAGGG  | (AAAT)$_{9}$  | FAM               | 154              | 55        | MK213987             | No hit                      |
|         | R: TGGGTTACTATGATGTGTTG  |              |                   |                  |           |                      |                             |
| CoTW504 | F: CCCGAGACAGACATATATTA  | (AAAT)$_{5}$  | FAM               | 159              | 55        | MK213988             | No hit                      |
|         | R: TCTAATGTGGTGGTGTG    |              |                   |                  |           |                      |                             |
| CoTW511 | F: AAGAAGCAGAGATGCTAATTGC | (AAAT)$_{3}$ | FAM               | 138              | 50        | MK213989             | No hit                      |
|         | R: CTAGCTACAGGGATAATTCTCG |              |                   |                  |           |                      |                             |
| CoTW513 | F: TGAGAATAATCAACATCCTAT | (AT)$_{10}$  | FAM               | 281              | 55        | MK213990             | No hit                      |
|         | R: AGTGTGATAACCACTACATGATGG |              |                   |                  |           |                      |                             |
| CoTW514 | F: GCAGGAGATTTGATGATAAATTA | (AAAT)$_{5}$ | FAM               | 291              | 55        | MK213991             | No hit                      |
|         | R: TTCTGGTGCCAAGCATATTATA |              |                   |                  |           |                      |                             |
| CoTW528 | F: CTTGTAATACATACATCTCA | (AAAT)$_{5}$  | FAM               | 147              | 55        | MK213992             | No hit                      |
|         | R: GCGAAAAACATAAAAATCTCA |              |                   |                  |           |                      |                             |
| CoTW581 | F: TGAAGAGGTGTAGTATGCTAC | (TA)$_{10}$(ATGT)$_{4}$ | JOE | 196              | 57.5      | MK213993             | Hypothetical protein *Bacillus subtilis* [Philippines] |
|         | R: ACATTTCTACTGGATG        |              |                   |                  |           |                      |                             |
Table 3. Genetic characterization of 23 polymorphic gSSR loci of *Chamaecyparis taiwanensis*. A number of alleles, $Ho$ observed heterozygosity, $He$ expected heterozygosity, PIC polymorphism information content or power of information content, PD power of discrimination, PE power of exclusion or probability of exclusion, $PI$ probability of identity, PD is equal to $1 – PI$, $N$ number of individuals. *Highly significant from Hardy–Weinberg equilibrium ($P < 0.001$).

| Locus   | A  | $Ho$ | $He$  | PIC   | PD    | PE    | $PI$ | $N$ |
|---------|----|------|-------|-------|-------|-------|------|-----|
| CoTW76  | 9  | 0.411| 0.670*| 0.644 | 0.670 | 0.151 | 0.329| 95  |
| CoTW77  | 4  | 0.022| 0.477*| 0.476 | 0.476 | 0.000 | 0.523| 93  |
| CoTW99  | 6  | 0.333| 0.378 | 0.354 | 0.378 | 0.102 | 0.621| 96  |
| CoTW314 | 5  | 0.396| 0.492 | 0.420 | 0.492 | 0.141 | 0.507| 96  |
| CoTW330 | 9  | 0.604| 0.642 | 0.576 | 0.642 | 0.312 | 0.357| 96  |
| CoTW337 | 10 | 0.625| 0.662 | 0.641 | 0.662 | 0.333 | 0.337| 96  |
| CoTW349 | 9  | 0.615| 0.626 | 0.591 | 0.626 | 0.322 | 0.373| 96  |
| CoTW495 | 5  | 0.427| 0.411 | 0.371 | 0.411 | 0.163 | 0.585| 96  |
| CoTW531 | 4  | 0.083| 0.119 | 0.115 | 0.119 | 0.006 | 0.880| 96  |
| CoTW539 | 2  | 0.000| 0.170*| 0.155 | 0.169 | 0.000 | 0.830| 96  |
| CoTW545 | 2  | 0.000| 0.041*| 0.059 | 0.041 | 0.000 | 0.958| 95  |
| CoTW554 | 3  | 0.042| 0.452*| 0.381 | 0.452 | 0.001 | 0.547| 96  |
| CoTW556 | 4  | 0.747| 0.605*| 0.548 | 0.605 | 0.479 | 0.394| 95  |
| CoTW559 | 2  | 0.479| 0.375 | 0.304 | 0.375 | 0.202 | 0.625| 96  |
| CoTW561 | 2  | 0.000| 0.061*| 0.058 | 0.060 | 0.000 | 0.939| 96  |
| CoTW582 | 14 | 0.635| 0.737 | 0.714 | 0.736 | 0.344 | 0.263| 96  |
| CoTW585 | 7  | 0.531| 0.546*| 0.514 | 0.546 | 0.240 | 0.453| 96  |
| CoTW588 | 10 | 0.615| 0.719 | 0.684 | 0.719 | 0.322 | 0.280| 96  |
| CoTW595 | 9  | 0.802| 0.833*| 0.821 | 0.970 | 0.002 | 0.029| 96  |
| CoTW597 | 9  | 0.583| 0.681 | 0.639 | 0.680 | 0.292 | 0.319| 96  |
| CoTW598 | 4  | 0.277| 0.257 | 0.260 | 0.256 | 0.072 | 0.743| 94  |
| CoTW599 | 12 | 0.542| 0.749 | 0.715 | 0.749 | 0.245 | 0.250| 96  |
| CoTW600 | 9  | 0.426| 0.527 | 0.519 | 0.526 | 0.162 | 0.473| 94  |
| Average | 6.521| 0.399| 0.488 | 0.459 | 0.494 | 0.169 | 0.505|     |

Table 4. Genetic characterization of 12 polymorphic EST-SSR loci of *Chamaecyparis taiwanensis*. A number of alleles, $Ho$ observed heterozygosity, $He$ expected heterozygosity, PIC polymorphism information content or power of information content, PD power of discrimination, PE power of exclusion or probability of exclusion, $PI$ probability of identity, PD is equal to $1 – PI$, $N$ number of individuals. *Highly significant from Hardy–Weinberg equilibrium ($P < 0.001$).

| Locus   | A  | $Ho$ | $He$  | PIC   | PD    | PE    | $PI$ | $N$ |
|---------|----|------|-------|-------|-------|-------|------|-----|
| CoTW383 | 12 | 0.104| 0.872*| 0.858 | 0.872 | 0.010 | 0.127| 96  |
| CoTW407 | 6  | 0.448| 0.475 | 0.389 | 0.474 | 0.178 | 0.525| 96  |
| CoTW409 | 16 | 0.458| 0.629 | 0.611 | 0.629 | 0.186 | 0.370| 96  |
| CoTW420 | 4  | 0.323| 0.388 | 0.365 | 0.387 | 0.096 | 0.612| 96  |
| CoTW424 | 3  | 0.552| 0.505 | 0.505 | 0.504 | 0.263 | 0.495| 96  |
| CoTW502 | 3  | 0.219| 0.327*| 0.298 | 0.326 | 0.046 | 0.673| 96  |
| CoTW504 | 4  | 0.604| 0.622 | 0.577 | 0.622 | 0.312 | 0.377| 96  |
| CoTW511 | 3  | 0.021| 0.205*| 0.187 | 0.205 | 0.000 | 0.794| 96  |
| CoTW513 | 5  | 0.563| 0.721*| 0.672 | 0.720 | 0.263 | 0.279| 96  |
| CoTW514 | 4  | 0.292| 0.481*| 0.443 | 0.481 | 0.079 | 0.518| 96  |
| CoTW528 | 3  | 0.448| 0.499*| 0.419 | 0.498 | 0.178 | 0.501| 96  |
| CoTW581 | 12 | 0.521| 0.615 | 0.583 | 0.614 | 0.231 | 0.385| 96  |
| Average | 7.000| 0.379| 0.528 | 0.482 | 0.555 | 0.180 | 0.443|     |
could be identified on the basis of molecular level. Additionally, these markers can be also used in population genetic analysis studies that facilitate the conservation and breeding of *C. taiwanensis*.

| Loci # | CPI  | Confidence levels (CL) | Comment                      |
|--------|------|------------------------|------------------------------|
| 1      | 2.900×10⁻² | 99.99% | 99% | 95% |                             |
| 2      | 3.683×10⁻³ | 1,300 | 7,000 |               | Miniature identifiable population size |
| 3      | 9.208×10⁻⁴ | 3,800 | 19,000 |               | Small identifiable population size |
| 4      | 2.422×10⁻⁴ | 10,000 | 53,000 |               | Moderate identifiable population size |
| 5      | 6.780×10⁻⁵ | 27,000 | 140,000 |               |                                             |
| 6      | 2.163×10⁻⁵ | 4,800 | 2,400,000 |               | Large identifiable population size |
| 7      | 7.289×10⁻⁶ | 9,000 | 480,000 |               |                                             |
| 8      | 2.602×10⁻⁶ | 19,000 | 980,000 |               |                                             |
| 9      | 9.628×10⁻⁷ | 53,000 | 2,400,000 |               |                                             |
| 10     | 3.591×10⁻⁷ | 140,000 | 6,600,000 |               |                                             |
| 11     | 1.354×10⁻⁷ | 370,000 | 16,600,000 |               |                                             |
| 12     | 5.213×10⁻⁸ | 980,000 | 46,000,000 |               |                                             |
| 13     | 2.054×10⁻⁸ | 1,900,000 | 89,000,000 |               |                                             |
| 14     | 9.304×10⁻⁹ | 5,500,000 | 24,000,000 |               |                                             |
| 15     | 4.401×10⁻⁹ | 11,000,000 | 53,000,000 |               |                                             |
| 16     | 2.178×10⁻⁹ | 23,000,000 | 108,000,000 |               |                                             |
| 17     | 1.091×10⁻⁹ | 47,000,000 | 231,000,000 |               |                                             |
| 18     | 5.653×10⁻¹⁰ | 90,000,000 | 454,000,000 |               |                                             |
| 19     | 2.957×10⁻¹⁰ | 170,000,000 | 809,000,000 |               |                                             |
| 20     | 1.552×10⁻¹⁰ | 330,000,000 | 1,580,000,000 |               |                                             |
| 21     | 8.491×10⁻¹¹ | 1,550,000,000 | 7,350,000,000 |               |                                             |
| 22     | 4.967×10⁻¹¹ | 10,000,000,000 | 47,000,000,000 |               |                                             |
| 23     | 3.040×10⁻¹¹ | 1,600,000,000 | 7,770,000,000 |               |                                             |
| 24     | 1.888×10⁻¹¹ | 2,700,000,000 | 13,000,000,000 |               |                                             |
| 25     | 1.270×10⁻¹¹ | 5,700,000,000 | 27,000,000,000 |               |                                             |
| 26     | 9.439×10⁻¹² | 4,000,000,000 | 19,000,000,000 |               |                                             |
| 27     | 7.495×10⁻¹² | 5,600,000,000 | 27,000,000,000 |               |                                             |
| 28     | 6.221×10⁻¹² | 8,200,000,000 | 39,000,000,000 |               |                                             |
| 29     | 5.84×10⁻¹² | 8,700,000,000 | 40,000,000,000 |               |                                             |
| 30     | 5.596×10⁻¹² | 9,100,000,000 | 45,000,000,000 |               |                                             |

Table 5. The discrimination power in SSR marker combination. CPI, cumulative random probability of identity, CL = (1 − CPI)², N number of individuals.
Figure 2. The photos and matrices of the three groups in which seized timbers and victim trees are successfully linked via 19 non-linked SSR markers matches. (a) Group 6TB/6TC, in wild. (b) Group 7TA/7TB/7TC, sampled. (c) Group 8TA/8TB, chopped. *A1 Freq = Allele 1 Frequency. **A2 Freq = Allele 2 Frequency. Allele frequencies are from 96 individuals of *C. taiwanensis*. P = p^2 or 2pq. CPI, combined probability of identity. CL, confidence levels.
Materials and methods

Developing *C. taiwanensis* individual identification system. Library preparation and SSR enrichment. In this study, we constructed both DNA and RNA libraries of *C. taiwanensis* (Fig. 1). Three DNA libraries were created from individuals of TP (Voucher no. Chung 2448) and 100R (Voucher no. Chung 2603, 2621) (Supplementary Sect. 1). To build the DNA libraries, genomic DNA was extracted from fresh leaves using the cetyltrimethylammonium bromide (CTAB) method. The quality and concentration of DNA were measured by NanoDrop 2000 (Thermo Fisher Scientific, San Diego, California, USA) and Qubit 2.0 Fluorometer (Thermo Fisher Scientific). From the total genomic DNA, microsatellites enriched in SSR markers was followed the magnetic bead enrichment method of Glenn and Schable. Briefly, DNA was digested using AluI/XmnI and HaeIII/XmnI (New England Biolabs, Ipswich, Massachusetts, USA). The double-stranded SuperSNX linkers (SuperSNX24 Forward: 5′-GTT TAA GGC CTA GCT AGC AGA ATC -3′; SuperSNX24 + 4p: 5′-pGAT TCT GCT AGC TAG GCC TTA AAC AAA-3′) were ligated to the digested DNA fragments. The linker-conjugated DNA fragments were hybridized with Biotin-labeled microsatellite probes containing Mix 2: (AG)_{12}, (TG)_{12}; Mix 3: (AAC)_{6}, (AAG)_{8}, (AAT)_{12}, (ATC)_{8}, (ACT)_{12}, and Mix 4: (AAAC)_{6}, (AAAG)_{6}, (AATC)_{6}, (AATG)_{6}, (ACAG)_{6}, (ACCT)_{6}, (ACTG)_{6}, (ACCT)_{6}, (AAAT)_{8}, (ACAT)_{8}, (AAGT)_{8}, and (AGAT)_{8}. The SSR hybridized fragments were extracted using Streptavidin M-280 Dynabeads (Invitrogen, Carlsbad, California, USA) and recovered by PCR using the SuperSNX24 Forward primers. The concentration and quality of SSR-enriched libraries were measured by Nanodrop 2000 (Thermo Fisher Scientific, Carlsbad, San Diego, California, USA) and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA).

One individual of *C. taiwanensis* (Voucher no.: Chung 2627) from XI was used to prepare RNA library. RNA was extracted from fresh leaves by using the CTAB method. The double-stranded SuperSNX linkers (SuperSNX24 Forward: 5′-GTT TAA GGC CTA GCT AGC AGA ATC -3′; SuperSNX24 + 4p: 5′-pGAT TCT GCT AGC TAG GCC TTA AAC AAA-3′) were ligated to the digested DNA fragments. The linker-conjugated DNA fragments were hybridized with Biotin-labeled microsatellite probes containing Mix 2: (AG)_{12}, (TG)_{12}; Mix 3: (AAC)_{6}, (AAG)_{8}, (AAT)_{12}, (ATC)_{8}, (ACT)_{12}, and Mix 4: (AAAC)_{6}, (AAAG)_{6}, (AATC)_{6}, (AATG)_{6}, (ACAG)_{6}, (ACCT)_{6}, (ACTG)_{6}, (ACCT)_{6}, (AAAT)_{8}, (ACAT)_{8}, (AAGT)_{8}, and (AGAT)_{8}. The SSR hybridized fragments were extracted using Streptavidin M-280 Dynabeads (Invitrogen, Carlsbad, California, USA) and recovered by PCR using the SuperSNX24 Forward primers. The concentration and quality of SSR-enriched libraries were measured by Nanodrop 2000 (Thermo Fisher Scientific, Carlsbad, San Diego, California, USA) and Qubit 2.0 Fluorometer (Thermo Fisher Scientific, USA).

Sequencing and analysis. Three DNA and one RNA libraries were sequenced using the Illumina MiSeq System (2 × 300 bp paired-end; Illumina, San Diego, California, USA) at Tri-I Biotech (New Taipei City, Taiwan). The raw reads were prescreened to remove adapter sequences and reads with greater than 0.1% error or with an

---

Table 6. Individual PI and CPI of four matched groups. PI probability of identity, CPI combined probability of identity, CL confidence levels. Database: 96 *C. taiwanensis* individuals. Markers failed in PCR amplification or their peak heights smaller than RFU150 in typing were not subjected into PI calculation and are shown in blank. 6TB, 7TA, 7TB and 8TA are victim tree. 6TC, 7TC, 8 TB are seized timber.

| Locus    | Matched samples | 6TB | 6TC | 7TA | 7TB | 7TC | 8TA | 8TB |
|----------|-----------------|-----|-----|-----|-----|-----|-----|-----|
| CoTW77   |                 |    .615  | .029 |    |    |    |    |    |
| CoTW99   |                 | .624 | .624 | .080 |    |    |    |    |
| CoTW314  |                 | .084 | .007 | .084 |    |    |    |    |
| CoTW337  |                 | .017 | .114 | .294 |    |    |    |    |
| CoTW409  |                 | .005 | .407 | .151 |    |    |    |    |
| CoTW420  |                 | .144 | .594 |    |    |    |    |    |
| CoTW424  |                 | .300 | .300 | .300 |    |    |    |    |
| CoTW495  |                 | .580 | .580 | .580 |    |    |    |    |
| CoTW502  |                 | .038 | .648 | .648 |    |    |    |    |
| CoTW504  |                 | .491 | .491 | .491 |    |    |    |    |
| CoTW511  |                 | .981 | .981 | .981 |    |    |    |    |
| CoTW514  |                 | .726 | .251 | .726 |    |    |    |    |
| CoTW528  |                 | .490 | .420 | .420 |    |    |    |    |
| CoTW531  |                 | .889 | .889 | .889 |    |    |    |    |
| CoTW556  |                 | .471 | .380 | .471 |    |    |    |    |
| CoTW595  |                 | .056 | .042 | .132 |    |    |    |    |
| CoTW597  |                 | .037 | .037 | .031 |    |    |    |    |
| CoTW598  |                 | .724 | .724 | .724 |    |    |    |    |
| CoTW600  |                 | .019 | .031 |    |    |    |    |    |
| PI       |                 |     3.342 × 10^{-15} | 1.631 × 10^{-13} | 4.648 × 10^{-16} |    |    |    |    |
| CPI      |                 | 99.9999999% | 99.9999999% | 99.999532% |    |    |    |    |

CL confidence levels.
average quality less than QV30. High-quality filtered DNA and cDNA reads were merged by CLC Genomics Workbench version 7.5 (QIAGEN, Aarhus, Denmark).

SSR screening and primer design. SSRIT was applied to screen the gSSR and EST-SSR containing sequences from contigs. To design gSSR and EST-SSR primers, sequences with at least five di-, tri-, tetra-, penta-, and hexa-nucleotide repeats were selected using BatchPrimer3,66, with optimized conditions set length at 18–23 bp, melting temperature 45–62 °C, and a product size of 80–300 bp.

Marker validation. A total of 75 markers including 23 gSSR and 12 EST-SSR markers newly designed in this study, and 40 published SSR67,68 (Supplementary Sect. 2) were subjected to validation test on 96 samples from four C. taiwanensis populations (TT SY, DS and FR, see Supplementary Sect. 1). In addition, we also tested cross-species transferability of the designed gSSR and EST-SSR markers (Supplementary Sect. 3). The samples used in marker validation and cross-species transferability of DNA were extracted using the VIOGENE plant DNA extraction kit (VIOGENE, New Taipei City, Taiwan). The PCR reaction was conducted with a final volume 20 μL containing 2 ng of genomic DNA, 0.25 μL of 10 μM each primer and 10 μL of Q-Amp 2 × Screening Fire Taq Master Mix (Bio-Genesis Technologies, Taipei, Taiwan). The following PCR conditions were used: an initial denaturation of 95 °C for 2 min; 30 cycles of 95 °C for 45 s, a primer-specific annealing temperature (Tables 1, 2) for 45 s, and 72 °C for 45 s; followed by a 15-min extension at 72 °C (Tables 1, 2). The amplified products were evaluated on the ABI 3130XL (Applied Biosystems, Waltham, Massachusetts, USA) with GeneScan 500 ROX Size Standard (Applied Biosystems). Fragment size was determined by using Genemapper version 3.2 (Applied Biosystems).

Marker analysis. GenAlex 6.51b259 was used to calculate number of alleles (A), observed heterozygosity (Ho), expected heterozygosity (He), Hardy–Weinberg equilibrium (HWE) of the newly developed gSSR and EST-SST markers. PowerMarker V3.2560 was used to calculate polymorphism information content or power of information content (PIC)61. Power of discrimination (PD)62, PD = 1 − Σ2, where P is the frequency of genotype i. Power of exclusion or probability of exclusion (PE)62, PE = h[1 − 2 h(1 − h)], where h is the frequency of heterozygotes. Probability of identity (Pi)63, Pi = 1 − PD. Combined power of discrimination (CPD)64, here we calculated CPD of 30 markers. CPD = 1 − [(1 − PD1)(1 − PD2)...(1 − PD30)]. Combined probability of identity (CPI)65. Microsoft Excel (Microsoft Office 2016) was used to calculate PD, PI, PE, CPD, CPI. GENEPOP 4.260 was used to test for linkage disequilibrium.

Aligning seized timbers to victim trees. Samples from five seized timbers of Taiwan Yilan District Prosecutors Office, six illegally-felled timbers found at crime scene woodland and seven victim trees (Supplementary Sect. 4) were collected. Duplicates of a victim tree (7TA and 7TB) was sourced out in order to ensure the reproducibility of the identical SSR type in individual tree. Two grams of each sample was powdered in liquid nitrogen and the total genomic DNA was extracted following the protocol of VIOGENE plant DNA extraction kit (VIOGENE, New Taipei City, Taiwan). Nineteen non-linkage markers were selected for DNA typing. The sample succeeded in typing were further combined to the aforementioned database to calculation the CPI.

Received: 12 February 2020; Accepted: 24 November 2020
Published online: 16 December 2020

References

1. Hwang, S. Y., Lin, H. W., Kuo, Y. S. & Lin, T. P. RAPD variation in relation to population differentiation of Chamaecyparis formosa and Chamaecyparis taiwanensis. Bot. Bull. Acad. Sin. 42, 173–179 (2001).
2. Wang, W. P., Hwang, C. Y., Lin, T. P. & Hwang, S. Y. Historical biogeography and phylogenetic relationships of the genus Chamaecyparis (Cupressaceae) inferred from chloroplast DNA polymorphism. Plant Syst. Evol. 241, 13–28. https://doi.org/10.1007/s00606-003-0031-0 (2003).
3. Chen, Y. J. & Chang, S. T. Distribution and characteristic comparisons of the endemic cypress in Taiwan. Taiwan J. For. Sci. 32, 71–86 (2017).
4. Huang, C. J. et al. Isolation and characterization of SSR and EST-SSR loci in Chamaecyparis formosensis (Cupressaceae). Appl. Plant Sci. 6, e01175. https://doi.org/10.1002/aps3.1175 (2018).
5. Tereba, A., Woodward, S., Komecka, A., Borys, M. & Nowakowska, J. A. Analysis of DNA profiles of ash (Fraxinus excelsior L.) to provide evidence of illegal logging. Wood Sci. Technol. 51, 1377–1387. https://doi.org/10.1007/s00226-017-0942-5 (2017).
6. Cabral, E. C. et al. Wood typification by Venturi easy ambient sonic spray ionization mass spectrometry: The case of the endangered Mahogany tree. J. Mass Spectrom. 47, 1–6. https://doi.org/10.1002/jms.2016 (2012).
7. Kite, G. C. et al. Dalnigrin, a neoflavonoid marker for the identification of Brazilian rosewood (Dalbergia nigra) in CITES enforcement. Phytochemistry 71, 1122–1131. https://doi.org/10.1016/j.phytochem.2010.04.011 (2010).
8. Dormontt, E. E. et al. Forensic timber identification: It’s time to integrate disciplines to combat illegal logging. Biol. Cons. 191, 790–798. https://doi.org/10.1016/j.biocon.2015.06.038 (2015).
9. Vlam, M. et al. Developing forensic tools for an African timber: Regional origin is revealed by genetic characteristics, but not by isotopic signature. Biol. Cons. 220, 262–271. https://doi.org/10.1016/j.biocon.2018.01.031 (2018).
10. Celani, C. P., Lancaster, C. A., Jordan, J. A., Espinoza, E. O. & Boekh, K. S. Assessing utility of handheld laser induced breakdown spectroscopy as a means of Dalbergia speciation. Analyt. Chem. 144, 5117–5126. https://doi.org/10.1039/c9an00984a (2019).
11. Gasson, P. How precise can wood identification be? Wood anatomy’s role in support of the legal timber trade, especially CITES. IAWA J. 32, 137–154 (2011).
12. Speer, J. H. Fundamentals of Tree-Ring Research (University of Arizona Press, Tucson, 2010).
Acknowledgements

The authors would like to thank Dr. Kuo-Fang Chung (Biodiversity Research Center, Academia Sinica, Taiwan) for his support in research funding and facilities. The authors would like to thank Dr. Chaolun Allen Chen and Dr. Shu-Miaw Chaw (Biodiversity Research Center, Academia Sinica, Taiwan) for their support in research facilities. The authors would like to thank Forestry Bureau Council of Agriculture of Executive Yuan, Taiwan for its support in samples collections. The authors also thank to the officials of Taiwan Yilan District Prosecutors Office, Forth Division of the Seventh Special Police Corps National Police Agency of Ministry of the Interior and Luodong Forest District Office for generously providing samples of the test cases. This work was financially supported by Ministry of Science and Technology, Taiwan (grant no. MOST 104-2321-B-002-056) and Ministry of Justice, Taiwan (grant no. 109-1301-05-17-02).

Author contributions

C.J.H. conceived, designed and conducted the experiments, and wrote the main manuscript text, drew the figures and tables, sample collection, funding application and manuscript submitted. E.H.C. edited the manuscript. Y.S.H. edited the manuscript and assisted to draw figures and tables. Y.M.H. performed the experiments. Y.H.T. edited the manuscript. C.E.P. data analysis and edited the manuscript. C.H.C. edited the manuscript. Y.C.T. performed the experiments. M.Y.C., T.A.L., H.Y.S., and Y.C.T. performed the experiments. C.T.C. funding application. All authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1038/s41598-020-79061-z.

Correspondence and requests for materials should be addressed to C.-J.H. or C.-T.C.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher’s note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2020