Phylogenetic analysis of six different species of *Saraca* L. (Fabaceae, Caesalpinioideae) based on chloroplast *matK* gene

**SUJIT SIL,1,2 KALYAN KUMAR DE,1,* ASOK GHOSH3**

1Department of Botany, Hooghly Muhammad Moshin College, Chinsurah, Hooghly, PIN-712101, W.B., India. Tel. +91 9832087564, *email: wesukalyan@gmail.com*

2Taxonomy of Angiosperms and Biosystematics Section, UGC-CAS Department of Botany, The University of Burdwan, Golapbag, Burdwan, PIN-713104, W.B., India

Manuscript received: 19 June 2021. Revision accepted: 25 August 2021.

**Abstract.** Sil S, De KK, Ghosh A. 2021. Phylogenetic analysis of six different species of *Saraca* L. (Fabaceae, Caesalpinioideae) based on chloroplast *matK* gene. Biodiversitas 22: 3880-3889. *Saraca* L. is one of the most important genera, with several horticultural and therapeutic values. Specific taxonomic and phylogenetic knowledge of *Saraca* through molecular data is essential for accessing its true medicinal benefits. Nineteen different Partial *matK* gene sequences of the chloroplast genome of six different species of *Saraca*, including four amplified and 15 retrieved from the NCBI gene bank, were placed in a sequence alignment. The resulting data were examined to determine their phylogenetic and evolutionary interrelationships. The comparative analysis of different sequences of each of the species revealed intra-specific molecular diversity, and the comparison of the *matK* sequences of six different species defined their inter-specific molecular diversity. The analysis of partial *matK* sequences revealed the presence of 87 variable sites, 14 parsimony informative sites, 54 singleton sites, and 237 quadri-fold degenerate sites. The approximate nucleotide composition was A=31.02%, T=37.46%, C=16.06%, and G=15.46%. The value of transition/transversion bias was 0.90. About 322 codons were analyzed and the presence of 34 variable sites, 8 parsimony informative sites, and 25 singleton sites was observed within their respective amino acid sequences. The average pair-wise distance was 0.0444, and 189 segregating sites and 0.018809 nucleotide diversity were observed. The evolution of different species of *Saraca* and their phylogenetic interrelationships were observed by analyzing their *matK* sequences. The relative homogeneity of *S. indica* is quite low. *S. dives* had the earliest evolutionary trends while *S. declinata* had the most recent. *S. asoca* and *S. indica* are quite similar on the molecular level but can be treated as different species while the difference between *S. declinata* and one of its synonyms, *S. palembanica*, indicates the possibility of separating them into different species.

**Keywords:** *Saraca*, *matK*, molecular diversity, phylogenetic, evolutionary interrelationships

**INTRODUCTION**

*Saraca* L. is one of the most important plant genera within the family Fabaceae, subfamily Detarioideae, tribe Saracaeae (Estrella et al. 2018). *Saraca* is the genus under the sub-clade Amherstieae (LPWG 2017); it was previously categorized under tribe Detarieae, but a new tribe, *Saracaeeae*, was subsequently introduced. The sub-clade Amherstieae of the family Fabaceae (LPWG 2017) has a number of both medicinally and horticulturally important species. This tribe contains different species of *Saraca* along with other two genera, *Endertia* and *Lysidice*, based on the nature of their habitats (Estrella et al. 2017), the number and nature of their stamens (Estrella et al. 2018), and the ornamentations on their pollen grains (Sil et al. 2019). The evaluation of phylogenetic relationships among different taxa of tribe Detarieae and three species of *Saraca* was done by Saha et al. (2013). *Saraca* contains about 11 species distributed throughout tropical and subtropical Asia, including Indonesia, Malaysia, Vietnam, Myanmar, Southern China, Bangladesh, India, and Sri Lanka (Sil et al. 2018; Zuijderhoudt 1967). Four of the 11 species of *Saraca* can be observed in India (Sil et al. 2018); of these, *S. asoca* (Roxb.) De Wilde (Wilde 1967), previously known as *Jonesia asoca* Roxb. (Hou et al. 1996, Figure 1.A), is native to India while the other three had been introduced later (Sil et al. 2018). *S. indica* L. (Figure 1.C) can be naturally found in different rainforests along with the populations of *S. asoca* (Preeti et al. 2012) from the central Himalayan foothills, eastern Himalayan forests, Gangetic basins, and western and eastern coasts of India (Hegde et al. 2018; Mohan et al. 2017; Preeti et al. 2012; Sil et al. 2018). Determining the differences between these two species is a challenging task (Begum et al. 2014) as their similar morphological natures and pollen morphologies make them difficult to properly distinguish (Sil et al. 2019). *S. declinata* (Jack) Miq., commonly known as ‘Red Saraca’, is native to Malaysia (Zuijderhoudt 1967) and is characterized by reddish to saffron flowers, spreading bracteoles (Hou et al. 1996), and pollen with a micro-rugulate surface (Sil et al. 2019). *S. declinata* (= *S. palembanica* Miq., Figure 1.B), previously known as *J. declinata* Miq., (Hou et al. 1996) is characterized by 3-4 stamens, spreading and deciduous to persistent bracteoles, and reddish flowers. *S. palembanica* Miq. is presently known as a synonym of *S. declinata* (Zuijderhoudt 1967). *S. dives* Pierre (1895) is a species of *Saraca* with extremely limited distribution and can be distinguished by its smaller clasping bracteoles, yellow flowers, 8±2 stamens, and absence of staminodes (Hou et al. 1996). *S.
Saha et al. 2012), anti-genotoxic (Nag et al. 2013), and anti-ulcer (Saha et al. 2012) importance. The creation of DNA barcodes to identify medicinally important plants for the authentication of herbal plants is an important task (Mishra et al. 2015). As a therapeutically potent plant, different species of Saraca need DNA barcodes in order to avail the of effective therapeutic benefits. Hegde et al. (2018) applied the rbcL gene to the making of the DNA barcode for S. asoca. The DNA barcoding of different species of Saraca must be performed for the proper identification and usage of such therapeutically valuable plants.

Overexploitation due to its high therapeutic values has made different species of Saraca globally vulnerable (http://dx.doi.org/10.2305/IUCN.UK.1998.RLTS.T34623A9879360.en). Evaluating the molecular diversity of these plants can help estimate both the strength of the gene pool and the potentiality of regeneration of depleted populations. RAPD (Mohan et al. 2017; Saini et al. 2018) and ISSR (Hegde et al. 2018) have been applied to evaluate the molecular diversity of S. asoca in previous studies but the evaluation of the inter and intra-specific diversity of different species of Saraca is a separate, challenging task that needs to be performed with the help of matK sequences. The purpose of the present study is to determine the molecular diversity of six species of Saraca at both the intra and inter-specific levels and emphasize their phylogenetic correlation.

MATERIALS AND METHODS

Collection of samples
Young, tender, and green leaves of studied species of Saraca, namely S. asoca, S. declinata (= S. palembanica), S. indica, and S. thaipingensis, were collected from the captivity. Voucher specimens of collected samples were preserved and submitted in BURD and accessioned in Herbarium of the University of Burdwan (Acronym BURD, Thiers 2021).

DNA extraction and purification
DNA was extracted and purified with the help of the power plant pro DNA isolation kit (MOBIO, USA); the purified DNA was collected and preserved at -20°C.

Figure 1. Flowering twigs of four species of Saraca. A: S. asoca, B: S. declinata, C: S. indica, D: S. thaipingensis
Table 1. Details of primers used during the PCR amplification

| Primer sequences | Designed by |
|------------------|-------------|
| matK-RP          | Saha et al. (2013) |
| matK-FP          | Johnson and Soltis (1994) |
| matK-RP          | designed by Authors |
| matK-FP          | designed by Authors |
| matK-RP          | designed by Authors |
| matK-FP          | designed by Authors |
| matK-RP          | designed by Authors |
| matK-FP          | designed by Authors |
| matK-RP          | designed by Authors |

Quality checking

Portions of extracted DNA were run through 1% agarose gel to test their purity. For the spectrometric analysis of the extracted DNA, 5 μl of extracted DNA was dissolved in 500 μl of double-distilled water. The optical densities were observed using a spectrophotometer (UV vis spectrophotometer 1700DC, Simarju, Japan).

Selection of primers

Johnson and Soltis (1994) discussed several primers for the PCR amplification of the matK gene product but such conserved regions also have rare variations. Thus, this experiment used some designed primers and some degenerate primers (Table 1).

PCR-based amplification

PCR-based amplification of the chloroplast matK region, along with some of the 5’ and 3’ introns (trnK), was done with the help of the following technique (Johnson and Soltis 1994). Two of the above set of primers (Table 1) were selected according to their amplification ability. The resulting mixture included 25 μl of the reaction mixture, containing 0.5 μl of purified DNA, 1.0 μl of dNTPs (10mM), 2.0 μl of the primer (100ng/μl), 0.5 μl of Taq DNA polymerase (3U/μl), 2.5 μl of 10X Taq assay buffer, and 18.5 μl of millipore water. The DNA amplification was performed using thermocycler technology (BioRad). The thermal conditions of the PCR were fixed as one cycle of initial denaturation for 5 min at 95°C and then for 45 sec at 94°C, followed by 35 cycles for 1 min each at 55°C; the annealing was maintained at 72°C for 1 min and the final amplification lasted for 10 min at 72°C.

1% Tris-acetate-EDTA agarose gel electrophoresis

1% Tris-acetate-agarose gel electrophoresis was established to visualize the productivity of the polymerized chain reaction. The stain used was ethidium bromide, and a ladder (100bp) was used as well. The gel plate was scanned and analyzed with Bio-Rad, USA-made Gel-DocTM XR+ system.

Purification of PCR product

The PCR product was purified with the help of the PEG precipitation technique. The ethanol precipitate of the PCR fragments was obtained by following a procedure published in Lever et al. (2015) with some modifications.

Sequencing of PCR product

Sequencing of the extracted DNA was conducted using a specific sequencing mixture with a total volume of 10 μl. It contained 0.5 μl of purified PCR product (100ng/μl), 3.5 μl of Big dye terminator ready reaction mix, 1.5 μl sequence buffer, 2 μl primer (100pmol/λ), and 2.5 μl of Milli Q water. This process followed that of Li et al. (2011). Four sequences of four species of Saraca (S. asoca, S. declinata, S. indica, and S. thaeipingensis) were accessed using the NCBI gene bank.

Sequence analysis

Nineteen different sequences (Table 2) of the matK genes of six different species, namely S. asoca, S. declinata, S. dives, S. indica, S. palembanica, and S. thaeipingensis containing 15 retrieved along with the four amplified sequences created were analyzed with the help of MEGAX software (Kumar et al. 2018). Different sequences were arranged according to different species, which assisted in analyzing intra- and inter-specific molecular diversity.

Comparative alignments of different sequences were conducted and the variable sites, parsimony informative sites, singleton sites, and quadric-fold degenerate sites were evaluated. From the nucleotide sequences, base substitution matrices (Tamura 1992), transition/transversion bias (Tamura and Nei 1993), nucleotide frequencies, composition and pair-wise distances (Nei and Kumar 2000), and disparity indices (Kumar and Gadagkar 2001) were also evaluated. Tajima’s neutrality test (Tajima 1989) was performed to quantify the segregation sites and nucleotide ratios. Amino acid sequences were obtained from the matK sequences with the help of ExPasy online software, and the diversity of the amino acids was evaluated.

Dendrogram

A dendrogram was obtained by applying sequences from six different species to a maximum likelihood comparative method using 1000 bootstrap with Endertia spectabilis as the out-group; an evolutionary tree was then obtained by applying the ML dendrogram.
Table 2. Details of the matK sequences understudy

| Species                        | Source            | Gene bank accessions |
|-------------------------------|-------------------|----------------------|
| Saraca asoca (Roxb.) De Wilde | Present authors   | MT535510             |
| Saraca asoca (Roxb.) De Wilde | NCBI gene bank    | KJ994830.1           |
| Saraca asoca (Roxb.) De Wilde | NCBI gene bank    | KF592389.1           |
| Saraca asoca (Roxb.) De Wilde | NCBI gene bank    | MG735764.1           |
| Saraca asoca (Roxb.) De Wilde | NCBI gene bank    | KY162281.1           |
| Saraca asoca (Roxb.) De Wilde | NCBI gene bank    | MG816814.1           |
| Saraca asoca (Roxb.) De Wilde | NCBI gene bank    | MG816793.1           |
| Saraca asoca (Roxb.) De Wilde | NCBI gene bank    | KX538519.1           |
| Saraca declinata (Jack) Miq.  | Present authors   | MT535512             |
| Saraca declinata (Jack) Miq.  | NCBI gene bank    | EU362035.1           |
| Saraca declinata (Jack) Miq.  | NCBI gene bank    | HM049553.1           |
| Saraca declinata (Jack) Miq.  | NCBI gene bank    | KX162282.1           |
| Saraca declinata (Jack) Miq.  | Present authors   | MT526218             |
| Saraca indicia L.             | Present authors   | EU362034.1           |
| Saraca indica L.              | Present author    | MT535511             |
| Saraca thaipingensis Cantley ex Prain | NCBI gene bank | KX162286.1 |
| Saraca indica L.              | Present authors   | KX162285.1           |

RESULTS AND DISCUSSIONS

The DNA extracted from the leaves of the four species of Saraca (S. asoca, S. declinata, S. indica, and S. thaipingensis) was run through 1% agarose gel (Figure 2) to test its purity. The selected primers were 5’-CGATTTATGCAATTGCTCGATA-3’ and 5’-CGAATGTCTCTACAGCTTGTGAT-3’; the latter ultimately showed better PCR productivity in the extracted DNA. The amount of DNA synthesized was satisfactory in the 1% Tris-acetate-EDTA agarose gel electrophoresis (Figure 3). The sequence data obtained from the experiment and additional sequence data obtained from the NCBI gene bank were then analyzed further.

The sequence alignment showed the presence of multiple insertions and deletions in the nucleotide sequence of the matK gene, showing important constraints in the process of evolution (Kim and Kim 2011); such gaps were treated as missing data and such data was excluded as ambiguous. The comparative accounts of the different matK sequences of the six different species are described below.

Comparative alignment (Figure S1) of the six sample sequences of S. asoca indicated the presence of 15 variable sites, seven parsim-info sites, two singleton sites, and 153 quadric-fold degenerate sites. The substitution matrix revealed (Table 3) that the different samples sequences of S. asoca had higher transitional substitution rates in C-T (18.4398) and A-G (15.6749). The nucleotide composite data (Table S4) and frequencies were 30.50%, 37.83%, 16.33%, and 15.34% for A, T/U, C, and G. The transition/transversion bias (Table S3) was 0.87. The average composition distance (Table S1) was 0.0022146888 and the average pair-wise distance was 0.0044910195 (Table S2). Tajima’s neutrality test (Table 7) showed the presence of 15 segregation sites; the nucleotide diversity was 0.001525 and the evolution and evolutionary rates (Table S5) were 0.90, 0.96, 1.00, 1.04, and 1.10. The Codon usage pattern was seen to be extremely diverse (Table S6) and amino acid sequencing (Figure S2, Table S7) of the protein retrieved from the different matK sequences of S. asoca showed five variable sites and one singleton site.

The comparative alignment (Figure S3) of the six sample sequences of S. declinata revealed the presence of eight variable sites, one parsim-info site, seven singleton sites, and 193 quadric-fold degenerate sites. The maximum likelihood estimate of the substitution matrix (Table 4) among the four samples of S. declinata and the two sequences of S. palembanica showed that A-G substitution is the highest (21.4819) while T-C substitution is the lowest (5.6815). A total of 1746 variable positions were found in the six sequences and the frequencies of A, T/U, C, G, are 30.77%, 37.34%, 16.41%, and 15.43%, respectively. The transition/transversion bias (Table S10) was 0.60 and the nucleotide frequencies were 25.00% for each of the four nucleotides. The average composite distance (Table S8) across the different specimens was 0.0026726890 while the average pair-wise distance (Table S9) was 0.0016635407. Eight segregation sites and 0.001031 nucleotide diversity (Table S11) were found via Tajima’s neutrality test of the six sequences (Table 7). The discrete Gamma distribution parameter (Table S12) was 200.0000 and the evolutionary rates are 0.90, 0.96, 1.00, 1.04, and 1.10. Codon usage bias showed higher diversity (Table S13). The amino acid sequences (Figure S4) (Table S14) of the matK gene sequences of S. declinata contained four variable sequences, one parsim-info site, and three singleton sequences.
Table 3. Maximum likelihood estimate of the substitution matrix of the sample sequences of *S. asoca*

| From/To | A     | T     | C     | G     |
|---------|-------|-------|-------|-------|
| A       | -     | 9.4674| 4.0859| 7.8849|
| T       | 7.6345| -     | 7.9539| 3.8404|
| C       | 7.6345| 18.4298| -     | 3.8404|
| G       | 15.6749| 9.4674| 4.0859| -     |

Note: Transitional substitution rates are in bold

Table 4. Maximum likelihood estimate of substitution matrix of sample sequences of *S. declinata*

| From/To | A     | T     | C     | G     |
|---------|-------|-------|-------|-------|
| A       | -     | 9.1771| 4.0335| 10.7741|
| T       | 7.5637| -     | 5.6815| 3.7935|
| C       | 7.5637| 12.9267| -     | 3.7935|
| G       | 21.4819| 9.1771| 4.0335| -     |

Note: Transitional substitution rates are in bold

Table 5. Maximum likelihood estimate of substitution matrix of sample sequences of *S. thaipingensis*

| From/To | A     | T     | C     | G     |
|---------|-------|-------|-------|-------|
| A       | -     | 7.8372| 3.4105| 11.3780|
| T       | 6.4353| -     | 7.2155| 3.1659|
| C       | 6.4353| 16.5808| -     | 3.1659|
| G       | 23.1279| 7.8372| 3.4105| -     |

Three sample sequences (Figure S5) of *S. thaipingensis* were aligned; analysis showed 24 variable sites, 24 singleton sites, and 150 quadri-fold degenerate sites. The substitution matrix (Table 5) of the three sample sequences of *S. thaipingensis* showed a maximum transitional substitution rate in A-G, though that of C-T/U was also high (16.58); the nucleotide frequencies were 30.87%, 37.59%, 16.36%, and 15.19% for A, T/U, C, and G respectively. The estimated transition/transversion bias (Table S17) was 1.19. The average overall composite distance (Table S15) was 0.117183951 and the pair-wise distance (Table S16) was 0.015074753. Nucleotide composition data is shown in Table S18. Tajima’s neutrality test evaluated 24 segregation sites with a nucleotide diversity of 0.009547 (Table 7); the evolutionary rates (Table S19) of the taxa were 0.00, 0.00, 0.01, 0.022, and 4.77, respectively. The codon usage pattern was shown in Table S20. The quantification of amino acid (Figure S6) (Table S21) showed greater diversity and higher amount of phenylalanine and serine was diagnosed.

The different sample sequences (Figure 4) of six *Saraca* species were selected from 18 sets of sequences; the comparative alignment showed 87 variable sites, 14 parsim-info sites, 54 singleton sites, and 237 quadri-fold degenerate sites. The disparity index test (Kumar and Gadagkar 2001) indicated that the homogeneity of the substitution pattern (Table S22) of *S. thaipingensis* as compared to *S. asoca, S. declinata,* and *S. palembanica* is high (1.0000), but low as compared to *S. indica* (0.0380). The maximum likelihood estimate of the substitution matrix (Table 6) of 2832 sites (Tamura and Nei 1993) showed that the transitional substitution rate of C-T/U is highest (17.45), though that between A-G was also high (17.30); the nucleotide frequencies were 31.02%, 37.46%, 16.06%, and 15.46% for A, T/U, C, and G respectively. The maximum likelihood estimate of transition/transversion bias (Table S25) (Tamura 1992) was 0.90 and the frequencies of A, T/U, C, and G were 34.24%, 34.24%, 15.76%, and 15.75%, respectively. The average composite distance (Table S23) was 0.05403 and the average pairwise distance (Table S24) was 0.0444. Tajima’s neutrality test (Tajima 1989) of six selected sample sequences showed 189 segregating sites, 0.018809 nucleotide diversity, and -2.313589 Tajima test statistic value (Table 7). Nucleotide composition data is shown in Table S26. The estimated value of the discrete Gamma distribution (Table S27) (Tamura and Nei 1993) was 1.1699 and the mean evolutionary rates were 0.14, 0.41, 0.74, 1.23, and 2.49 substitutions per site. The codon usage bias (Table S28) and the quantification of amino acids (Figure 5) (Table S29) showed greater diversities, and isoleucine and serine had the maximum abundances.
The dendrogram, obtained using the maximum likelihood method (Figure 6), showed that the *S. dives* had evolved the earliest and that *S. asoca* and *S. indica* had evolved from the same lineage. *S. thaipingensis*, *S. palembanica*, and *S. declinata* evolved successively from the same ancestral lineage and *S. declinata* and *S. palembanica* showed close genetic proximity. An average of 522 codons was analyzed and the alignment of amino acid sequences revealed the presence of 34 variable sites, eight parsim-info sites, and 25 singleton sites.

The *matK* gene sequence of the chloroplast had faced many phases of evolution in different plants; this presents as the nucleotide substitution within the sequences (Guyeux et al. 2019; Nguyen et al. 2015; Kress 2017). Phylogenetic analysis of *Chamaecrista* (Torres et al. 2011), *Cistaceae* (Aparicio et al. 2017), *Leucojum* sp., and *Galanthus* sp. (Tasci et al. 2013) was done effectively with the help of the sequence data of the *matK* gene. *matK* sequencing has been widely used in different Fabaceae plants, including *Caesalpinioideae* (Monkheang 2011), to evaluate discriminatory status (Khan et al. 2016) and identification (Gao et al. 2011).
Table 7. Results of Tajima’s neutrality tests of intra-specific and inter-specific levels of Saraca

| Species       | M  | S   | Ψt      | Θ       | Π     | D         |
|---------------|----|-----|---------|---------|-------|-----------|
| S. asoca      | 6  | 15  | 0.00591716 | 0.002591457 | 0.001525312 | -2.53584415 |
| S. declinata  | 6  | 8   | 0.004581901 | 0.002006672 | 0.00103928  | -2.866601   |
| S. thaipingensis | 3  | 24  | 0.014319809 | 0.009546539 | 0.009546539 | n/c         |
| Inter specific | 6  | 189 | 0.066737288 | 0.029228009 | 0.018808851 | -2.31358942 |

Note: M=number of sequences; S=number of segregating sites; Ψt=S/n; Θ=Ps/αl; Π=nucleotide diversity; D=Tajima’s test statistics

Figure 6. Dendrogram showing the phylogenetic interrelationships between six different species of Saraca

*matK* sequences amplified from four different species of *Saraca* were analyzed along with various previously retrieved *matK* sequences of the six different species of *Saraca*. This procedure helped both evaluate the genetic diversity and phylogenetic correlations between six species of *Saraca* and determine the intra-specific diversity of each species. The characterization of tribe Detarieae by applying the data of *matK* sequence including three species of *Saraca* was done by Saha et al. (2013) and the diversification of the new tribe Saraceae from Detarieae was done by Estrella et al. (2018), but the phylogenetic estimation of six species of *Saraca* and the evaluation of their intra-specific molecular diversity had not been done prior to the present study.

Molecular evolution is characterized by nucleotide substitution per site among different DNA sequences; in addition to transition/transversion, G+C conversion biases were applied to more specifically decipher the molecular evolution (Tamura 1992). This method has more potential than other pair-wise comparative methods, like Tajima and Nei (1984). Kumar and Gadagkar (2001) established that the differences between sequences are half of the sum of the overall nucleotide diversity. In our analysis of the intra-specific level molecular diversity of different species of *Saraca*, *S. thaipingensis* showed the maximum values of transition/transversion bias, but its nucleotide frequencies were similar to the other species. This result indicates that, though these three species are closely related, their intra-specific diversity is also pronounced. The overall composite distance was also much higher in *S. thaipingensis* than in other species. Tajima’s neutrality test showed that segregation sites and nucleotide diversity were also much higher in *S. thaipingensis* than in the other species. Overall analysis of the six species showed a higher transition/transversion bias. The presence of 189 segregating sites and higher nucleotide diversity values within the *matK* sequence of the six species of *Saraca*, as determined by Tajima’s neutrality test, indicates that the genus *Saraca* has large molecular diversity.

A *matK*-based DNA barcode can efficiently identify different species through molecular signature (More et al. 2016). The morphological differences between *S. asoca* and *S. indica* are less pronounced than molecular differences. The morphological approach (Begum et al. 2014) and the palynological approach (Sil et al. 2019) have a sufficient though limited ability to identify different species. Molecular barcoding using the *matK* gene marker is, therefore, best suited for proper identification. *S. palembanica* Miq. is synonymous with *S. declinata* (Jack) Miq. (Hou et al. 1996) and their morphological features, like their 3-4 functional stamens and the nature of their bracteoles, differ somewhat, but their molecular characteristics as per *matK* sequence diversity indicate the necessity of defining *S. palembanica* as a different species that evolved from the same ancestors.

The true therapeutic benefits of medicinally important plants can only be obtained if proper identification is conducted, making molecular identification essential. Though Mishra et al. (2015) and Hegde et al. (2018) used ISSR and *rbcL* markers to define *S. asoca* compared to its adulterants, using the *matK* sequence is the most used effective method for meeting this goal.

DNA barcoding, particularly via applying different chloroplast genes like the *matK* sequence, can help determine genetic divergence (Duan et al. 2019) and protect endangered plant species (Gogoi and Bahu 2018).
Different species of *Saraca* are threatened due to overexploitation and reductions in the gene pool. *matK*-based analysis can help estimate the genetic variability of populations and the strength of the gene pool; this data can then be used to assist in the regeneration of depleted populations of different species of *Saraca*.

*Saraca* is a diverse genus and different species therein have different levels of molecular complexity. The relative homogeneity among different samples of *S. indica* was found to be the lowest, *S. dives* occurred earliest evolutionarily and *S. declinata* is the most evolved. *S. asoca* and *S. indica* have strong molecular resemblances but their molecular diversity makes them separate species while *S. declinata* and *S. palembanica* have different molecular identities even though *S. palembanica* is defined as being the same as *S. declinata*.

**ACKNOWLEDGEMENTS**

The authors are thankful to the Director of the Botanical Survey of India, Scientist E, and the Head of the Office at the CNH, AJC Bose Indian Botanic Garden, Shibpur, Howrah, West Bengal, India for both their permission to use their samples and their consultation. They are also thankful to the Heads of the Department of Botany, Hooghly Muhammad Mohsin College, Chinsurah, Hooghly, West Bengal, India, for their cooperation and excellent laboratory environment and to the Heads of the Department of Botany, The University of Burdwan for their support.

**REFERENCES**

Androsov P, Jastrzębski JP, Paukszto L, Makowczenko K, Okorski A, Pszczolowska A, Chwedorzewskia JG, Gorecki R, Gielawianska I. 2020. Evolutionary dynamics of the plastid genome sequences of six Colobanthus species. Sci Rep 10: 11522. DOI: 10.1038/s41598-020-68563-5

Aparicio A, Martin-Hetnansz S, Parejo-Farnes C, Arroyo J, Laverge S, Yesilyurt EB, Zhenggang X, Meng D, Xiaomei Z, Wenzhen K, Zhenggang X. 2019. Detecting useful polymorphic sites: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 158: 1321-1327. DOI: 10.1093/genetics/158.3.1321

Kumar S, Stecher G, Li M, Knyaz C, Tamura K. 2018. MEGA X: Molecular evolutionary genetics analysis across computing platforms. Mol Bio Evol 35: 71-79. DOI: 10.1093/molbev/msy096

Lever MA, Torti A, Eickenbusch P, Michaud AB, Šantl-Temkiv T, Jørgensen BB. 2015. A modular method for the extraction of DNA and RNA, and the separation of DNA pools from diverse environmental sample types. Front Microbiol 6: 476. DOI: 10.3389/fmicb.2015.00476

Li Y, Gao LM, Poudel RC, Li DZ, Forrest A. 2011. High universality of *matK* primers for barcoding gymnosperms. J Syst Evol 49 (3): 169-175. DOI: 10.1111/j.1759-6831.2011.00128.x

Li Y, Zhou J, Chen X, Cui Y, Xu Z, Li Y, Song J, Duan B, Yao H. 2017. Gene losses and partial deletion of small single-copy regions of the plastid genomes of two hemiparasitic *Taxisia* species. Sci Rep 7: 12834. DOI: 10.1038/s41598-017-13404-1

LPWG. 2017. A new subfamily classification of the Leguminosae based on a taxonomically comprehensive phylogeny. Taxon 66: 44-77. DOI: 10.12705/6613.1

Li JM, Wen J, Lutz S, Wang YP, Li DZ. 2012. Phylogenetic relationships of Chinese Adiantum based on five plastid markers. J Plant Res 125: 237-249. DOI: 10.1007/s10265-011-0441-y

Meng D, Xiaozei M, Wennze K, Zhenggang X. 2019. Detecting useful genetic markers and reconstructing the phylogeny of an important medicinal resource plant, *Artemisia selengensis*, based on chloroplast genomics. PLoS One 14 (2): e0211340. DOI: 10.1371/journal.pone.0211340

Mishra SB, Vijaykumar M. 2014. Anti-hyperglycemic and anti-oxidant effect of *Saraca asoca* (Roxb.) De Wilde flowers in Streptozocin-Induced diabetic rat: A therapeutic study. J Biomed Biomed 2: 338-343. DOI: 10.4172/1948-593X.S12-003

Mishra P, Kumar A, Nagreddy A, Mani DN, Shukla AK, Tiwari R, Sundaresan V. 2015. DNA barcoding: An efficient tool to overcome authentication challenges in the herbal market. Plant Biotechnol J 13 (1): 8-21. DOI: 10.1111/pbi.12419

Mohan C, Reddy MS, Kumar SM, Manzelat SF, Cherku PD. 2017. RAPD studies of *Saraca asoca* by fluorescent-labeled primers and development of micropropogation protocol for its conservation. Int J Appl Agri Res 12 (2): 137-151

Monkheang P, Runglawan S, Tanee T, Nookot K. 2011. Species diversity, usages, molecular markers and barcode of medicinal *Senna* species
(Fabaceae, Caesalpinioideae) in Thailand. J Med Plant Res 5 (26): 6173-6181. DOI: 10.5897/JMPR11.1075

More RP, Chandrashekhkar M, Purohit HJ. 2016. matK-QR classifier: A patterns-based approach for plant species identification. BioData Mining 9: 39. DOI: 10.1186/s13040-016-0120-6

Nag D, Ghosh M, Mukherjee A. 2013. Antimutagenic and genoprotective effects of Saraca asoca bark extract. Toxicol Ind Health 1: 1-8. DOI: 10.1177/0748233713483200

Nei M, Kumar S. 2000. Molecular Evolution and Phylogenetics. Oxford University Press, New York.

Nguyen TP, Trang NM, Doo NM, Sinh NV, Triest L. 2015. Application of DNA barcoding markers to the identification of Hopea species. Genetics Mol Res 14 (3): 9181-9190. DOI: 10.4238/2015.August.7.28

Pal TK, Bhattacharya S, Dey A. 2014. Evaluation of antioxidant activities of flower extract (Fresh and dried) of Saraca indica grown in West Bengal. Int J Current Microb Appl Sci 3 (4): 251-259.

Preeti B, Bharti A, Sharma AN, Singh V. 2012. A review on Saraca indica plant. Int J Res Pharma 3: 80-84.

Ruchiansakun S, Niet TVD, Triboun P, Techaprasan J, Jenjittikul T, Sukathan P. 2015. Phylogenetic analysis of molecular data and reconstruction of morphological character evolution in Asian Impatiens section Semeiocardiun (Balsaminaceae). Bioone Complete 40 (4): 1063-1074. DOI: 10.1600/036364415X690102

Saha J, Mitra T, Gupta K, Mukherjee S. 2012. Phytoconstituents and HPTLC analysis in Saraca asoca (Roxb.) Wilde. Int J Pharm Sci 4 (1): 96-99.

Saha J, Gupta K, Gupta B. 2013. Phylogenetic analyses and evolutionary relationships of Saraca asoca with their allied taxa (Tribe-Detarieae) based on the chloroplast matK gene. J Plant Biochem Biotechnol 24: 65-74. DOI: 10.1007/s13562-013-0237-3

Saini A, Hegde S, Hegde HV, Choklute SD, Roy S. 2018. Assessment of genetic diversity of Saraca asoca (Roxb.) De Wilde: A commercially important, but endangered, forest tree species in Western Ghats, India. New Zealand J For Sci 48 (17): 1-12. DOI: 10.1186/s40490-018-0122-x

Sherin DR, Manojkumar TK. 2017. Flavonoids from Saraca asoca- Ideal medication for breast cancer: A molecular simulation approach. 1 (6): 1-3. DOI: 10.26717/BISTR.2017.01.000533

Sil S, Mallick T, De KK, Pramanik A, Ghosh A. 2018. Comparative morphological study of three species of Saraca L. (Fabaceae) by the statistical approach to find out the logic of potent morphological markers. Beni-Suef Univ J Basic Appl Sci 7: 612-619. DOI: 10.1016/j.byas.2018.07.004

Sil S, Mallick T, Pal T, Mondal A, De KK, Ghosh A. 2019. Pollen morphology of Indian species of Saraca L. (Leguminosae)- A threatened and legendary medicinal tree. Phytochemistry 88: 295-315. DOI: 10.1016/j.phytochem.2019.06907

Tajima F, Nei M. 1984. Estimation of evolutionary distance between nucleotide sequences. Mol Biol Evol 1: 269-285. DOI: 10.1093/oxfordjournals.molbev.a040317

Tajima F. 1989. Statistical methods to test for nucleotide mutation hypothesis by DNA polymorphism. Genetics 123: 585-595.

Tamura K. 1992. Estimation of the number of nucleotide substitutions when there are strong transition-transversion and G + C-content biases. Mol Biol Evol 9: 678-687. DOI: 10.1093/oxfordjournals.molbev.a040752

Tamura K, Nei M. 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10: 512-526. DOI: 10.1093/oxfordjournals.molbev.a040823

Tasci N, Yuzbasioğlu S, Celen Z, Ekim T, Bilgin AN. 2013. Molecular phylogeny of Galanthus (Amaryllidaceae) of Anatolia inferred from multiple nuclear and chloroplast DNA regions. Turk J Bot 37: 993-1007. DOI: 10.3906/bot-1209-41

Thiers B. 2021. [continuously updated] Index Herbariorum: A global directory of public herbaria and associated staff. New York Botanical Garden’s Virtual Herbarium. http://sweetgum.nybg.org/science/db/ (accessed on 14th June 2021).

Thilagam E, Chidambaram K, Raviteja C, Vahan T, Vasudev P. 2021. Anti-hyperglycemic and hypolipidemic effects of Saraca asoca (Roxb.) Wild. flowers in alloxan-treated diabetic rats. J Pharma Pharmac Res 9 (1): 58-68.

Torres DC, Lima JPM, Fernandes AG, Nunes EP, Grangeiro TB. 2011. Phylogenetic relationships within Chamaecrista sect. Xerocalyx (Leguminosae, Caesalpinioideae) inferred from the cpDNA trnE-trnT intergenic spacer and nDNA ITS sequences. Genet Mol Biol 34: 244-251. DOI: 10.1590/S1514-47572011000200014

Wilde WJJ. 1967. A new combination and a new species in Saraca L. (Leguminosae, Caesalpinioideae) of Anatolia inferred from the cpDNA trnE-trnT intergenic spacer and nDNA ITS sequences. Genet Mol Biol 34: 244-251.

Yang Z, Zhao T, Ma Q, Liang L. Wang G. 2018. Comparative genomics and phylogenetic analysis revealed the chloroplast genome variation and interspecific relationships of Corylus (Betulaceae) Species. Front Plant Sci 9: 927. DOI: 10.3389/fpls.2018.00927

Zhai K, Li L, Quan H, Yang J, Zhang Z, Liao Z, Lan X. 2021. Comparative analysis of chloroplast genomes from 14 Zanthoxylum species: Identification of variable DNA markers and phylogenetic relationships within the genus. Front Plant Sci 11: 605793. DOI: 10.3389/fpls.2020.605793

Zuijderhoudt GFP. 1967. A revision of the genus Saraca L. (Leguminosae). Blumea 15 (2): 392-395.

A. 1967. A new combination and a new species in Saraca L. (Leguminosae). Blumea 15 (2): 392-395.
### SUPPLEMENTARY DATA

Analysed data for different sequences of *Saraca asoca* (Roxb.) De Wilde

Table S1. Composition distance among different samples of *S. asoca* in compatibility mode

| Sample ID         | Distance | Average |
|-------------------|----------|---------|
| Saraca_asoca_KU994830.1 | 0.001111111 | 0.000000000 |
| Saraca_asoca_KC592389.1 | 0.002222222 | 0.003333333 |
| Saraca_asoca_MG735764.1 | 0.0025974026 | 0.0038961039 | 0.000000000 |
| Saraca_asoca_KX162281.1 | 0.0050651230 | 0.0011111111 | 0.0055555556 | 0.0064935065 |
| Saraca_asoca_KY492334 | 0.0018348624 | 0.0000000000 | 0.0000000000 | 0.0000000000 | 0.0000000000 | 0.0000000000 |
| Average           |          | 0.0022146888 |        |

Table S2. Pair wise distance among different samples of *S. asoca* in compatibility mode

| Sample ID         | Distance | Average |
|-------------------|----------|---------|
| Saraca_asoca_KU994830.1 | 0.0033420427 |          |
| Saraca_asoca_KC592389.1 | 0.0100747704 | 0.0067067104 |
| Saraca_asoca_MG735764.1 | 0.0117904345 | 0.0078468726 | 0.0000000000 |
| Saraca_asoca_KX162281.1 | 0.0058195165 | 0.0011120909 | 0.0078209115 | 0.0091615768 |
| Saraca_asoca_KY492334 | 0.0036817658 | 0.0000000000 | 0.0000000000 | 0.0000000000 | 0.0000000000 |
| Average           |          | 0.004910195 |        |

Table S3. Maximum Likelihood Estimate of Transition/Transversion Bias

| From/To | A       | T       | C       | G       |
|---------|---------|---------|---------|---------|
| A       | -       | 8.4953  | 3.9373  | 7.9601  |
| T       | 8.4953  | -       | 7.9601  | 3.9373  |
| C       | 8.4953  | 17.147  | -       | 3.9373  |
| G       | 17.147  | 8.4953  | 3.9373  | -       |

Table S4. Nucleotide composition data

| Sample ID         | T(U)    | C       | A       | G       | Total |
|-------------------|---------|---------|---------|---------|-------|
| Saraca_asoca_KU994830.1 | 37.62662808 | 16.8596237 | 30.6078148 | 14.9059334 | 1382 |
| Saraca_asoca_KC592389.1 | 38      | 17.7777778 | 30      | 14.2222222 | 900   |
| Saraca_asoca_MG735764.1 | 38.11111111 | 17.8888889 | 29.7777778 | 14.2222222 | 900   |
| Saraca_asoca_KX162281.1 | 38.18181818 | 18.1818182 | 28.961039 | 14.6753247 | 770   |
| Saraca_asoca_KY492334 | 37.55442063 | 13.6094675 | 31.7159763 | 17.1203156 | 2535  |
| Average           | 38.3482385 | 20      | 28.8073394 | 12.8440367 | 545   |

| Sample ID         | Pos #1   | Pos #2   | Pos #3   |
|-------------------|----------|----------|----------|
| Saraca_asoca_KU994830.1 | 34.273189 | 34.1442595 | 34.273189 |
| Saraca_asoca_KC592389.1 | 34.273333 | 18.2212581 | 31.2364425 |
| Saraca_asoca_MG735764.1 | 34.333333 | 20.6666667 | 32.333333 |
| Saraca_asoca_KX162281.1 | 33.984375 | 20.703125  | 29.296875  |
| Saraca_asoca_KY492334 | 33.3727811| 15.147929  | 31.0051972 |
| Average           | 37.0165476| 23.2044199 | 26.519337  |

| Sample ID         | Pos #1   | Pos #2   | Pos #3   |
|-------------------|----------|----------|----------|
| Saraca_asoca_KU994830.1 | 37.0932755 | 37.0932755 | 37.0932755 |
| Saraca_asoca_KC592389.1 | 37.333333 | 17.6666667 | 33.333333 |
| Saraca_asoca_MG735764.1 | 38.5214008 | 18.2879377 | 31.5175097 |
| Saraca_asoca_KX162281.1 | 40.1183432 | 13.6094675 | 34.4378698 |
| Saraca_asoca_KY492334 | 36.2637363 | 18.1318681 | 34.6153846 |
| Average           | 38.336887 | 16.4605544 | 33.497548  |

| Sample ID         | Pos #1   | Pos #2   | Pos #3   |
|-------------------|----------|----------|----------|
| Saraca_asoca_KU994830.1 | 41.5217391 | 14.1304348 | 28.6956522 |
| Saraca_asoca_KC592389.1 | 42.333333 | 15      | 27.333333 |
| Saraca_asoca_MG735764.1 | 42.6666667 | 15      | 26.6666667 |
| Saraca_asoca_KX162281.1 | 42.0323463 | 15.5642023 | 26.0700389 |
| Saraca_asoca_KY492334 | 39.1715976 | 12.0710059 | 29.704142 |
| Average           | 41.7582418 | 18.6813187 | 25.2747253 | 14.2857143 | 182   |

| Sample ID         | Pos #1   | Pos #2   | Pos #3   |
|-------------------|----------|----------|----------|
| Saraca_asoca_KU994830.1 | 40.9982935 | 14.1211604 | 28.0716724 |
| Saraca_asoca_KC592389.1 | 42.333333 | 15      | 27.333333 |
| Saraca_asoca_MG735764.1 | 42.6666667 | 15      | 26.6666667 |
| Saraca_asoca_KX162281.1 | 42.0323463 | 15.5642023 | 26.0700389 |
| Saraca_asoca_KY492334 | 39.1715976 | 12.0710059 | 29.704142 |
| Average           | 40.9982935 | 14.1211604 | 28.0716724 | 16.8088737 | 390.666667 |
Table S5. Maximum Likelihood Estimate of Gamma Parameter for Site Rates

| From\To | A     | T     | C       | G       |
|---------|-------|-------|---------|---------|
| A       | -     | 9.4643| 4.0846 | 7.8618  |
| T       | 7.6319| -     | 7.9798 | 3.8391  |
| C       | 7.6319| 18.4898| -       | 3.8391  |
| G       | 15.6289| 9.4643| 4.0846 | -       |

Table S6. Codon usage pattern of sample sequences of *S. asoca*

| Codon | Count | RSCU |
|-------|-------|------|
| UCU(F) | 23.7  | 1.37 |
| UUC(F) | 11    | 0.63 |
| UUA(L) | 11.3  | 1.44 |
| UUG(L) | 10.2  | 1.29 |
| CCA(P) | 12.2  | 1.55 |
| CUC(L) | 2.2   | 0.28 |
| CUA(A) | 6.5   | 0.83 |
| CUG(L) | 4.8   | 0.61 |
| AUU(V) | 21    | 1.65 |
| AUC(I) | 7.8   | 0.62 |
| AAA(I) | 9.3   | 0.73 |
| ACG(M) | 10.3  | 1    |
| GCU(V) | 4.8   | 0.99 |
| GUC(V) | 4     | 0.82 |
| GUA(V) | 5.5   | 1.13 |
| GUG(V) | 5.2   | 1.06 |

Table S7. Quantification of amino acids evaluated from the sequences of *S. asoca*

| Alphabetic Code | Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu |
|----------------|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| MT535510       | 1.94384492 | 1.51187905 | 2.80777538 | 4.75161987 | 8.63930886 | 2.37580994 | 2.59179266 | 10.799136 | 7.77537797 | 12.0950324 |
| KU994830.1     | 2.0 | 1.66666667 | 3 | 3.66666667 | 9.66666667 | 2 | 3.33333333 | 9.33333333 | 8.33333333 | 12 |
| K592386.1      | 2.0 | 1.66666667 | 3 | 3.66666667 | 9.66666667 | 2 | 3.33333333 | 9.33333333 | 8.33333333 | 12 |
| MG735764.1     | 2.34375 | 1.953125 | 2.73475 | 3.515625 | 7.965625 | 1.95125 | 3.125 | 8.984375 | 7.421875 | 12.890625 |
| KX162281.1     | 2.16925247 | 1.38067061 | 3.33333333 | 5.25443479 | 8.48126233 | 2.36863923 | 2.76134212 | 10.2564103 | 7.10059172 | 12.4260355 |
| KY492334.1     | 1.65745856 | 1.65745856 | 1.65745856 | 4.41989895 | 9.94475138 | 1.65745856 | 4.41989895 | 8.28729228 | 7.18232044 | 13.2596685 |
| Avg.           | 2.042850025 | 1.94491533 | 2.8989854 | 4.38465371 | 9.16791231 | 2.14520125 | 2.02891870 | 9.76581963 | 7.62331839 | 12.5567514 |

Table S8. Composition distance among different samples of *S. declinata* in compatibility mode

| Genus           | Weighted composition distance |
|-----------------|-------------------------------|
| Saraca_declinata_MG816814.1 | 0.0000000000 |
| Saraca_declinata_MG816793.1 | 0.0000000000 |
| Saraca_declinata_KX538519.1 | 0.0017513135 |
| Saraca_declinata_EU362033.1 | 0.00052356021 |
| Saraca_palembanica_MT535512 | 0.0000000000 |
| Saraca_palembanica_EU362035.1 | 0.0000000000 |
| Average         | 0.0026726890 |

Analysed data for different sequences of *Saraca declinata* (Roxb.) De Wilde
Table S9. Pair wise distance among different samples of *S. declinata* in compatibility mode

| Species                        | Distance         | Genetic Distance | Protein Distance | Total    |
|--------------------------------|------------------|------------------|------------------|----------|
| *Saraca_declinata_MG816814.1*  | 0.0000000000     | 0.0017592359     | 0.0001820622     | 0.00018103684 |
| *Saraca_declinata_KX538519.1*  | 0.00035041829     | 0.00035041829    | 0.00035041829    | 0.00035041829 |
| *Saraca_palembanica_MT535512*  | 0.0000000000     | 0.0000000000     | 0.0000000000     | 0.0000000000 |
| *Saraca_palembanica_MT535512*  | 0.0000000000     | 0.0000000000     | 0.0000000000     | 0.0000000000 |

Average: 0.0016635407

Table S10. Maximum Likelihood Estimate of Transition/Transversion Bias

| From| To  | A   | T   | C   | G   | Total |
|-----|-----|-----|-----|-----|-----|-------|
| A   |     | 10.0612 | 4.7012 | 6.5205 | 4.7012 | |
| T   | 10.0612 | 13.9547 |       | 4.7012 |       | |
| C   | 13.9547 | 10.0612 |       |       |       | |
| G   |     |       |       |       |       | |

Table S11. Nucleotide composition data

| Species                        | T(U) | A   | C   | G   | Total |
|--------------------------------|------|-----|-----|-----|-------|
| *Saraca_declinata_MG816814.1*  | 37.07482993 | 18.1972789 | 29.7619048 | 14.9659864 | 588 |
| *Saraca_declinata_MG816793.1*  | 37.07482993 | 18.1972789 | 29.7619048 | 14.9659864 | 588 |
| *Saraca_declinata_KX538519.1*  | 37.33413752 | 15.8021713 | 31.0615199 | 15.8021713 | 1658 |
| *Saraca_declinata_MT535512*    | 37.66937669 | 17.6151762 | 30.6233062 | 14.0921409 | 1107 |
| *Saraca_declinata_MT535512*    | 37.26169844 | 15.5979203 | 31.2536106 | 15.8687707 | 1731 |
| Avg.                          | 37.3541384 | 16.4179104 | 30.7869742 | 15.4409769 | 1228.3333 |

| From| To  | A   | T   | C   | G   | Total |
|-----|-----|-----|-----|-----|-----|-------|
| T-1 | C-1 |     |     | A-1 | G-1 |       |
| T-2 | C-2 | A-2 | G-2 |     |     |       |

Table S12. Maximum Likelihood Estimate of Gamma Parameter for Site Rates

| From| To  | A   | T   | C   | G   | Total |
|-----|-----|-----|-----|-----|-----|-------|
| A   |     | 9.1770 | 4.0335 | 10.7742 | |
| T   | 7.5636 | 5.6816 | 3.7935 |       | |
| C   | 7.5636 | 12.9267 |       | 3.7935 | |
| G   | 21.4822 | 9.1770 | 4.0335 |       | |

BIODIVERSITAS
Table S13. Codon usage bias of sample sequences of *S. declinate*

| Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU |
|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|------|
| UUC(F) | 24.2 | 1.36 | UCU(S) | 18.2 | 2.66 | UAU(Y) | 20.2 | 1.53 | UGU(C) | 4.7 | 1.6 |
| UUC(F) | 11.5 | 0.64 | UCC(S) | 5.2 | 0.76 | UAC(Y) | 6.2 | 0.47 | UGC(C) | 1.2 | 0.4 |
| UUA(L) | 12 | 1.56 | UCA(S) | 10.5 | 1.54 | UAA(*) | 2.5 | 2.25 | UGA(*) | 0.5 | 0.45 |
| UUG(L) | 9.8 | 1.28 | UCG(S) | 2.2 | 0.32 | UAG(*) | 0.3 | 0.3 | UGG(W) | 6.7 | 1 |
| CUA(L) | 11.3 | 1.47 | CCC(P) | 7 | 2.02 | CAU(H) | 9.3 | 1.78 | CGU(R) | 3.3 | 0.91 |
| CUC(L) | 2.7 | 0.35 | CCC(P) | 1 | 0.29 | CAC(H) | 1.2 | 0.22 | CGC(C) | 0.5 | 0.14 |
| CAU(L) | 6.3 | 0.82 | CCA(P) | 4.8 | 1.4 | CAA(Q) | 8.8 | 1.68 | CGA(R) | 6.8 | 1.86 |
| CUG(L) | 4 | 0.52 | CGG(P) | 1 | 0.29 | CAG(Q) | 1.7 | 0.32 | CGG(R) | 4 | 1.09 |
| AUU(L) | 22 | 1.58 | ACC(G) | 3.7 | 1.49 | AGU(N) | 17.8 | 1.56 | AGU(S) | 4.5 | 0.66 |
| AUC(L) | 8.5 | 0.61 | ACC(G) | 3.7 | 1.49 | AAC(N) | 5 | 0.44 | AGC(S) | 0.5 | 0.07 |
| AUU(L) | 11.2 | 0.8 | ACA(T) | 2.5 | 1.02 | AAA(K) | 20.8 | 1.45 | AGA(R) | 4.5 | 1.23 |
| AUG(M) | 9 | 1 | AGC(T) | 0 | 0 | AAG(K) | 7.8 | 0.55 | AGG(R) | 2.8 | 0.77 |
| GUU(V) | 5.3 | 1.03 | GCU(A) | 4 | 1.96 | GAU(D) | 9 | 1.46 | GGU(G) | 3 | 1.2 |
| GUC(V) | 4.5 | 0.87 | GCC(A) | 2 | 0.98 | GAC(D) | 3.3 | 0.54 | GGC(G) | 0.5 | 0.2 |
| GUU(V) | 7.2 | 1.39 | GCA(A) | 1.2 | 0.57 | GAA(E) | 14 | 1.35 | GGA(G) | 2.7 | 1.07 |
| GUG(V) | 3.7 | 0.71 | GCG(A) | 1 | 0.49 | GAG(D) | 6.7 | 0.65 | GGG(G) | 3.8 | 1.53 |

Table S14. Quantification of amino acids estimated from the sequences of *S. declinate*

| Ala | Cys | Tyr | Glu | Phe | Gly | His | Thr | Val | Ile | Lys | Leu | Met |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| MG168141.1 | 2.07253886 | 1.55440415 | 0.01507475 | 10.36296358 | 2.59067358 | 2.95067358 | 8.0829016 | 7.77202073 | 10.36296358 |
| MG167793.1 | 2.07253886 | 1.55440415 | 0.01507475 | 10.36296358 | 2.59067358 | 2.95067358 | 8.0829016 | 7.77202073 | 10.36296358 |
| KX338519.1 | 2.19123506 | 1.39422321 | 3.3845418 | 8.56737057 | 2.7884462 | 2.58964143 | 10.5577689 | 6.97211555 | 11.5537849 |
| EU362031.1 | 2.17821782 | 1.58415842 | 3.36636636 | 5.34653465 | 8.17287129 | 2.37623726 | 5.27425743 | 10.49049549 | 6.93093079 | 11.4851485 |
| MT353512 | 2.16801821 | 1.62601626 | 2.7100271 | 4.06504065 | 8.94308943 | 2.16801821 | 2.98102981 | 10.8401084 | 7.58807588 | 10.8401084 |
| EU362031.1 | 2.17821782 | 1.58415842 | 3.36636636 | 5.34653465 | 8.14581458 | 2.37623726 | 5.27425743 | 10.69306939 | 6.93093079 | 11.6831683 |
| Avg. | 2.16446846 | 1.54388696 | 2.95454773 | 4.89633877 | 8.95456551 | 2.47022497 | 2.46666666 | 10.32601155 | 7.19019109 | 11.2483458 |

Table S15. Composition distance among different samples of *S. thaipingensis* in compatibility mode

| Saraca_thaipingensis_MT153511 | 0.0175925926 | 0.0000000000 | 0.011728395 |
| Saraca_thaipingensis_MT153511 | 0.0175925926 | 0.0000000000 | 0.011728395 |
| Average | 0.0175925926 | 0.0000000000 | 0.011728395 |

Table S16. Pair wise distance among different samples of *S. thaipingensis*

| Saraca_thaipingensis_MT153511 | 0.0226121299 | 0.0000000000 | 0.015074753 |
| Saraca_thaipingensis_MT153511 | 0.0226121299 | 0.0000000000 | 0.015074753 |
| Average | 0.0226121299 | 0.0000000000 | 0.015074753 |

Table S17. Maximum likelihood transition/transversion bias

| From/To | A | T | C | G |
|---------|---|---|---|---|
| A | 7.1828 | 3.3097 | 9.1523 |
| T | 7.1828 | 9.1523 | 3.3097 |
| C | 7.1828 | 9.1523 | 3.3097 |
| G | 19.8626 | 7.1828 | - | 3.3097 |
### Table S18. Nucleotide composition data

| Species                  | U (%) | C (%) | A (%) | G (%) | Total (%) |
|--------------------------|-------|-------|-------|-------|-----------|
| Saraca_thaipingensis_MT535511 | 37.59259259 | 18.0555556 | 30.5555556 | 13.7962963 | 1080 |
| Saraca_thaipingensis_KX162285.1 | 37.58949881 | 15.8114558 | 30.9665871 | 15.6324582 | 1676 |
| Saraca_thaipingensis_KX162286.1 | 37.58949881 | 15.8114558 | 30.9665871 | 15.6324582 | 1676 |
| Avg.                      | 37.59025271 | 16.3583032 | 30.8664264 | 15.1850181 | 1477.3333 |
| T-1                      | C-1     | A-1    | G-1   | Pos #1 |
| Saraca_thaipingensis_MT535511 | 34.1666667 | 19.4444444 | 29.4444444 | 16.9444444 | 360 |
| Saraca_thaipingensis_KX162285.1 | 33.6314848 | 17.352415  | 30.7692308 | 18.2468964 | 559 |
| Saraca_thaipingensis_KX162286.1 | 33.6314848 | 17.352415  | 30.7692308 | 18.2468964 | 559 |
| Avg.                      | 33.7618403 | 17.8619756 | 30.4465494 | 17.9296346 | 492.666667 |
| T-2                      | C-2     | A-2    | G-2   | Pos #2 |
| Saraca_thaipingensis_MT535511 | 37.22222222 | 19.4444444 | 33.0555556 | 10.2777778 | 360 |
| Saraca_thaipingensis_KX162285.1 | 37.745975  | 16.2790698 | 33.614848  | 12.3437405 | 559 |
| Saraca_thaipingensis_KX162286.1 | 37.745975  | 16.2790698 | 33.614848  | 12.3437405 | 559 |
| Avg.                      | 37.6184032 | 17.0500677 | 33.4912043 | 11.8403248 | 492.666667 |
| T-3                      | C-3     | A-3    | G-3   |
| Saraca_thaipingensis_MT535511 | 41.3888889 | 15.2777778 | 29.1666667 | 14.1666667 | 360 |
| Saraca_thaipingensis_KX162285.1 | 41.3978495 | 13.7928328 | 28.4946237 | 16.3082437 | 558 |
| Saraca_thaipingensis_KX162286.1 | 41.3978495 | 13.7928328 | 28.4946237 | 16.3082437 | 558 |
| Avg.                      | 41.395664  | 14.1598916 | 28.6585366 | 15.7850709 | 492 |

### Table S19. Maximum likelihood estimate of Gamma parameter for each site

| From/To | A     | T     | C     | G     |
|---------|-------|-------|-------|-------|
| A       | 7.5962 | 3.057 | 11.7766 |
| T       | 6.2375 | 7.2376 | 3.0868 |
| C       | 6.2375 | 16.6316 | - | 3.0868 |
| G       | 23.9383 | 7.5962 | 3.057 | - |

### Table S20. Codon usage pattern of sample sequences of *S. thaipingensis*

| Codon | Count | RSCU |
|-------|-------|------|
| UUU(F) | 29.7 | 1.38 |
| UUC(S) | 22.3 | 2.66 |
| UAU(Y) | 25 | 1.56 |
| UAG(*) | 3.3 | 2.5 |
| UGC(R) | 11.3 | 1.91 |
| CCA(Q) | 11 | 1.65 |
| CAG(Q) | 2.3 | 0.35 |
| AAU(N) | 21.7 | 1.57 |
| AAC(N) | 6 | 0.43 |
| AAA(K) | 25 | 1.5 |
| AAG(K) | 8.3 | 0.5 |
| GCU(A) | 12 | 1.5 |
| GCC(P) | 0.5 | 0.15 |
| GGC(P) | 5 | 0.92 |
| GCA(A) | 1.6 | 0.7 |
| GAA(A) | 0.43 |
| GAC(A) | 0.7 |

### Table S21. Quantification of amino acid estimate of *S. thaipingensis*

| Ala | Cys | Asp | Glu | Phe | Gly | His | Ile | Lys | Leu |
|-----|-----|-----|-----|-----|-----|-----|-----|-----|-----|
| MT535511 | 2.22222222 | 1.66666667 | 2.77777778 | 4.16666667 | 9.16666667 | 2.22222222 | 3.05555556 | 10.2777778 | 7.22222222 | 10.83333333 |
| KX162285.1 | 1.94444444 | 1.11111111 | 2.5 | 5.55555556 | 9.72222222 | 2.22222222 | 3.05555556 | 9.44444444 | 6.11111111 | 11.11111111 |
| KX162286.1 | 1.94444444 | 1.11111111 | 2.5 | 5.55555556 | 10 | 2.22222222 | 3.05555556 | 10 | 6.11111111 | 10.55555556 |
| Avg. | 2.037037037 | 1.296296296 | 2.592925925 | 5.092925925 | 9.629269629 | 2.22222222 | 3.05555556 | 9.9074074074 | 6.4814814814 | 10.833333333 |
| Trp | 1.388888889 | 3.688888889 | 4.444444444 | 2.77777778 | 4.16666667 | 10.833333333 | 2.5 | 5.55555556 | 1.66666667 | 6.66666667 |
| Ser | 1.944444444 | 6.666666667 | 3.611111111 | 4.72222222 | 10.55555556 | 9.444444444 | 5.27777778 | 1.388888889 | 7.5 | 360 |
| Thr | 1.944444444 | 6.666666667 | 3.611111111 | 4.72222222 | 10.55555556 | 9.444444444 | 5.27777778 | 1.5740740741 | 7.22222222 | 360 |
| Val | 1.75925926 | 6.666666667 | 3.888888889 | 4.357037037 | 10.55555556 | 2.037037037 | 5.27777778 | 1.5740740741 | 7.22222222 | 360 |
Analysed data of different sample sequences of six species of *Saraca*

Table S22. Disparity index test of the sequences of different species of *Saraca*

| MT535510       | 0.0021691974 | 0.0000000000 | 0.0021598272 | 0.0636363636 | 0.0000000000 |
|----------------|--------------|--------------|--------------|--------------|--------------|
| EU362033.1     | 0.3820000000 | 0.0000000000 | 0.0178217822 | 0.0292792793 | 0.0000000000 |
| MT535512       | 1.0000000000 | 1.0000000000 | 0.0000000000 | 0.0710382514 | 0.0027777778 |
| KX162281.1     | 0.3940000000 | 0.0660000000 | 1.0000000000 | 0.0695067265 | 0.0000000000 |
| MT526218       | 0.0000000000 | 0.0780000000 | 0.0020000000 | 0.0000000000 | 0.0807799443 |
| MT535511       | 1.0000000000 | 1.0000000000 | 0.2980000000 | 1.0000000000 | 0.0040000000 |

Table S23. Composite distance of different species of *Saraca*

| Saraca_asoca_MT535510 | Saraca_declinata_EU362033.1 | 0.01672 |
|-----------------------|-------------------------------|--------|
| Saraca_palembanica_MT535512 | 0.10547 | 0.13438 |
| Saraca_dives_KX162282.1 | 0.01520 | 0.01493 | 0.07422 |
| Saraca_indica_MT526218 | 0.03960 | 0.03267 | 0.16758 | 0.03622 |
| Saraca_thaipingensis_MT535511 | 0.00648 | 0.01759 | 0.00741 | 0.01111 | 0.13092 |
| Average               | 0.05403 |       |        |        |

Table S24. ML Pair wise distance of different species of *Saraca*

| Saraca_asoca_MT535510 | Saraca_declinata_EU362033.1 | 0.0162 |
|-----------------------|-------------------------------|--------|
| Saraca_palembanica_MT535512 | 0.0955 | 0.0830 |
| Saraca_dives_KX162282.1 | 0.0131 | 0.0231 | 0.0919 |
| Saraca_indica_MT526218 | 0.0249 | 0.0320 | 0.1062 | 0.0287 |
| Saraca_thaipingensis_MT535511 | 0.0332 | 0.0226 | 0.0207 | 0.0303 | 0.0452 |
| Average               | 0.0444 |       |        |        |

Table S25. Maximum likelihood transition/transversion bias

| From/To | A      | T      | C      | G      |
|---------|--------|--------|--------|--------|
| A       | -      | 8.3938 | 3.8640 | 8.0334 |
| T       | 8.3938 | -      | 8.0334 | 3.8640 |
| C       | 8.3938 | 17.4510 | -      | 3.8640 |
| G       | 17.4510 | 8.3938 | 3.8640 | -      |
### Table S26. Nucleotide composition data

| Codon | T(U) | C | A | G | Total |
|-------|------|---|---|---|-------|
| Saraca asoca MT35510 | 37.62662808 | 16.8596237 | 30.6078 | 14.9059334 | 1382 |
| Saraca declinata EU362033.1 | 37.16291097 | 14.5179165 | 31.2523088 | 17.0668637 | 2707 |
| Saraca palembanica MT35512 | 39.10375 | 17.265625 | 31.640625 | 13.984375 | 1280 |
| Saraca dives KX162282.1 | 37.44047619 | 15.5952381 | 31.3095238 | 15.6547619 | 1680 |
| Saraca indica MT526218 | 38.13747228 | 16.1862528 | 30.3769401 | 15.2993348 | 1353 |
| Saraca thaipingensis MT35511 | 37.59295295 | 18.0555556 | 30.5555556 | 13.7962963 | 1080 |
| Avg. | 37.46045138 | 16.0620122 | 31.01666623 | 15.4608732 | 1580.3333 |

| Codon | T-1 | C-1 | A-1 | G-1 | Pos # |
|-------|-----|-----|-----|-----|-------|
| Saraca asoca MT35510 | 34.2733189 | 18.2212581 | 31.2364425 | 16.2689805 | 461 |
| Saraca declinata EU362033.1 | 34.1842397 | 15.8712542 | 31.0765816 | 18.8679245 | 901 |
| Saraca palembanica MT35512 | 33.0210773 | 18.969555 | 31.8501171 | 16.1592506 | 427 |
| Saraca dives KX162282.1 | 33.8680927 | 17.4688057 | 30.3030303 | 18.3600713 | 561 |
| Saraca indica MT526218 | 35.0332594 | 17.7383592 | 30.3769401 | 16.8514412 | 451 |
| Saraca thaipingensis MT35511 | 34.1666667 | 19.4444444 | 29.4444444 | 16.9444444 | 360 |
| Avg. | 34.1031319 | 17.5893705 | 30.7813983 | 17.5260993 | 526.8333 |

### Table S27. Maximum likelihood estimate of Gamma parameter of each site

| From To | A | T | C | G |
|---------|---|---|---|---|
| A       | - | 9.1453 | 3.9212 | 8.7199 |
| T       | 7.5721 | - | 7.4906 | 3.7745 |
| C       | 7.5721 | 17.4699 | - | 3.7745 |
| G       | 17.4933 | 9.1453 | 3.9212 | - |

### Table S28. Codon usage bias of six species of Saraca

| Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU | Codon | Count | RSCU |
|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|------|-------|-------|------|
| UUU(F) | 33.8 | 1.43 | UCU(S) | 22.3 | 2.54 | UAU(Y) | 24.3 | 1.52 | UGU(C) | 6 | 1.53 |
| UUC(F) | 13.5 | 0.57 | UCC(S) | 7.2 | 0.82 | UC(Y) | 7.7 | 0.48 | UGC(C) | 1.8 | 0.47 |
| UUA(L) | 16.7 | 1.65 | UCA(S) | 13.5 | 1.54 | UAA(*) | 2.7 | 1.45 | UGA(*) | 2 | 1.09 |
| UUG(L) | 14.2 | 1.4 | UCG(S) | 3.5 | 0.4 | UAG(*) | 0.8 | 0.45 | UGG(W) | 8 | 1 |
| CUU(L) | 13 | 1.29 | CCA(P) | 8.8 | 2.21 | CA(U) | 11 | 1.73 | CGU(R) | 4.3 | 0.91 |
| CUC(L) | 4.2 | 0.41 | CCP(C) | 1.3 | 0.33 | CAC(H) | 1.8 | 0.27 | CGC(R) | 0.7 | 0.14 |
| CUA(L) | 7.7 | 0.76 | CCA(P) | 4.5 | 1.13 | CA(A) | 11.7 | 1.69 | CGA(R) | 9.5 | 1.99 |
| CGU(L) | 5 | 0.49 | CGG(P) | 1.3 | 0.33 | CAG(Q) | 2.2 | 0.31 | CGG(R) | 3.8 | 0.8 |
| AUA(L) | 28.7 | 1.6 | ACC(T) | 5.3 | 1.6 | AAA(N) | 22.3 | 1.53 | AGU(S) | 5.7 | 0.65 |
| AUC(L) | 10.8 | 0.61 | ACC(T) | 4.5 | 1.35 | AAC(N) | 6.8 | 0.47 | AGS(S) | 0.5 | 0.06 |
| AUA(L) | 14 | 1.79 | ACA(T) | 3.2 | 0.95 | AAA(K) | 28.3 | 1.48 | AGA(R) | 6.5 | 1.36 |
| AUC(L) | 10.5 | 1.0 | AGC(T) | 0.3 | 0.1 | AAG(K) | 10 | 0.52 | AGG(R) | 3.8 | 0.8 |
| GUU(V) | 6.5 | 1.03 | GCU(A) | 4.3 | 1.68 | GAU(D) | 12.2 | 1.54 | GGU(G) | 3.3 | 0.93 |
| GUC(V) | 5.2 | 0.82 | GCC(A) | 3 | 1.16 | GAC(D) | 3.7 | 0.46 | GGC(G) | 1.2 | 0.33 |
| GUA(V) | 8.8 | 1.39 | GCA(A) | 2 | 0.77 | GAA(E) | 15.7 | 1.25 | GGA(G) | 5.5 | 1.53 |
| GUG(V) | 4.8 | 0.76 | GCG(A) | 1 | 0.39 | GAG(E) | 9.3 | 0.75 | GGG(G) | 4.3 | 1.21 |
Table S29. Quantification of amino acid estimate of six species of Saraca

|       | Ala  | Cys   | Asp  | Glu  | Phe | Gly  | His  | Ile  | Lys  | Leu  |
|-------|------|-------|------|------|-----|------|------|------|------|------|
| MT535510 | 1.943844492 | 1.5118 | 2.8077538 | 4.75161987 | 8.6390886 | 2.37580994 | 2.59179266 | 10.7991361 | 7.77537797 | 12.0950324 |
| EU362033.1 | 2.178217822 | 1.58415842 | 3.3633663 | 5.34653465 | 8.71287129 | 2.37623762 | 2.57425743 | 10.4950495 | 6.93069307 | 11.4851485 |
| MT535512 | 2.16802168 | 1.62601626 | 2.7100271 | 4.06504065 | 8.94308943 | 2.16802168 | 2.98102981 | 10.8401084 | 7.58807588 | 10.8401084 |
| KK162281.1 | 2.169625247 | 3.3530572 | 5.12820513 | 8.0678501 | 2.56410256 | 2.76134122 | 10.2564103 | 7.29783037 | 12.6232742 |
| MT526218 | 2.222222222 | 1.66666667 | 2.77777778 | 4.16666667 | 9.16666667 | 2.22222222 | 3.05555556 | 10.2777778 | 7.22222222 | 10.8333333 |

MT535511 | 2.110022607 | 1.58251669 | 3.01431801 | 4.74755087 | 8.96759608 | 2.41145441 | 2.75056518 | 10.4747551 | 7.4227581 | 11.7181613 |

|       | Met  | Asn  | Pro  | Gln  | Arg  | Ser  | Thr  | Val  | Trp  | Tyr  | Total |
|-------|------|------|------|------|------|------|------|------|------|------|-------|
| 1.2958633 | 5.83153348 | 3.45572354 | 2.37580994 | 5.61555076 | 11.2311015 | 2.8077538 | 4.31965443 | 1.94384449 | 5.83153348 | 463 |
| 1.58415842 | 5.94059406 | 3.3663663 | 2.77227273 | 5.74257426 | 10.0990999 | 2.57425743 | 4.55455545 | 1.58415842 | 6.73267327 | 505 |
| 1.35501355 | 6.50406504 | 4.33604336 | 2.7100271 | 4.60704607 | 10.5691057 | 2.16802168 | 5.14905149 | 1.89701897 | 6.77506775 | 369 |
| 1.57790927 | 5.91715976 | 3.15581854 | 2.76134122 | 5.52268245 | 9.86193294 | 2.56410256 | 4.53648915 | 1.57790927 | 6.7061144 | 507 |
| 1.33333333 | 5.55555556 | 3.55555556 | 2.44444444 | 5.55555556 | 9.33333333 | 2.88888889 | 4.66666667 | 1.77777778 | 6 | 450 |
| 1.38888889 | 6.38888889 | 4.44444444 | 2.77777777 | 4.16666667 | 10.83333333 | 2.5 | 5.55555556 | 1.66666667 | 6.66666667 | 360 |
| 1.43180106 | 5.99095705 | 3.65486059 | 2.63752826 | 5.27505652 | 10.2863602 | 2.59984928 | 4.74755087 | 1.73323268 | 6.44310475 | 442.333333 |

SUPPLEMENTARY FIGURES

Figure S1. matK sequence alignment of different samples of S. asoca

Figure S2. Alignment of amino acid sequences of matK gene of different specimens of S. asoca
Figure S3. *matK* sequence alignment of different samples of *S. declinata* and *S. palembanica*

Figure S4. Alignment of amino acid sequences of *matK* gene of different specimens of *S. declinata* and *S. palembanica*

Figure S5. *matK* sequence alignment of different samples of *S. thaipingensis*

Figure S6. Alignment of amino acid sequences of *matK* gene of different specimens of *S. thaipingensis*