ABSTRACT

The present study was carried out to re-affirm the identity of 196 plant species belonging to 67 families, at molecular level using DNA Barcoding technique. It resulted not only in the re-affirmation of the plant species but also laid down the foundation of gene library in Pakistan. The present effort also helped in revising the Lahore district flora compiled by Kashyap and Joshi in 1936 on the basis of the morphological features of the plant species. The plants were collected from different sites of Lahore district during 2013-2016 and the GPS data thus obtained was recorded. All the plant specimens collected were mounted on herbarium sheets and deposited as voucher specimens at Dr. Sultan Ahmed Herbarium GC University Lahore. A small piece of the leaves of all these plant species was processed for DNA extraction and its sequencing using standard protocols, at Canadian Centre of DNA Barcoding, University of Guelph Canada. The results thus obtained after the evaluation of barcode regions for species identification based on best match and best close-match through *rbcL, trnH-psbA* and *matK* plants core barcode regions, were interpreted and discussed for the determination of intra-specific and inter specific divergence, assessment of barcoding gap, reconstruction of phylogenetic trees and finally re-affirming the species identity. Gene sequencing of all plant species was a big success towards the cataloguing of molecular taxonomic record. The results showed that the sequence success was high for *rbcL* and *trnH-psbA* as compared to *matK*. According to the DNA Barcode data cladogram of order Rosales indicated that family Moraceae formed an individual clade in all Neighbour Joining trees formed on the basis of *rbcL*, *matK*, *trnH-psbA* while in previous studies it was sister to Urticaceae and Cannabaceae. Another change was also observed in the placement of Family Rosaceae that it was not forming a separate clade as it did in previous studies but it was sister to Cannabaceae, Rhamnaceae and Urticaceae and embedded among these families. Our study does not support the previous taxonomic position of order Rosales based on morphological characters and some DNA sequences which are not standard barcodes now. In Monocot group three families Poaceae, Cyperaceae and Typhaceae were the member of order Poales. In this group the bio marker *matK* performed well by placing all these families properly under their respective order poales in NJ tree of *matK*. While in the NJ tree of *rbcL*, two families' Cyperaceae and Poaceae were well-placed but third one Typhaceae showed deviation. The results revealed that *rbcL* and *trnH-psbA* showed comparatively less overlapping for the distribution of inter specific and intra-specific divergence. In addition the highest discriminating ability for correct species identification was also observed in this region.

Therefore, *rbcL*, *trnH-psbA* and *matK* were found to be a significant barcode region for the identification of most of the plant species of the study area.

The data analysis was accomplished on the basis of three criteria; Amplification and Sequencing Success of *rbcL*, *trnH-psbA* and *matK*; Barcode Gap Analysis as Nearest Neighbor (NN) Analysis and Neighbor Joining (N/J) Cluster Analysis. PCR and sequencing success were found very high for the *rbcL* region, i.e. 94.73% (630 / 665 specimens), for *trnH-psbA* 73% (490/665 specimens) while *matK* had the lowest overall rate of recovery, i.e. 57.9% (385 / 665 specimens). The *matK* showed the highest sequence quality for almost all the recovered samples. Barcode Gap Analysis revealed that 95.22% sequences of *matK*, 92% of *trnH-psbA* and 96.63 % of *rbcL* exhibited no intra-specific variation. The inter-specific divergence varied from 0.0% to 19.06% for *rbcL* , 0.0% to 11.2% for *trnH-psbA* while 0.0% to 4.12% for *matK*. Out of 105 congeneric species with *rbcL* sequences, 69/105 (65.71%) were identified while in congeneric species with *matK* sequences 54/71 species (76.05%) were successfully differentiated. In pair wise divergence across all the species (non-congeneric), both *matK* and *rbcL* sequences showed clear boundaries between the 84.21% (96/114) and 69.92% (93/133) of the species, respectively. In Barcode Gap Analysis, *matK* showed more discriminatory power than *rbcL* and *trnH-psbA*. Among taxa, patterns of sequence divergence was visualized by means of Neighbor-Joining (NJ) cluster on MEGA6. In all three trees of rbcL, matK and trnH-psbA, assignment of families within their respective orders was 83.61% (51/61), 79.59% (39/49) and 83.67% (41/49), respectively. In all three trees of rbcL, matK and trnH-psbA, species were identified on the basis of "Bootstrap Threshold Value" and "Monophyly". In neighbor Joining trees (N/J) of *rbcL*, *matK* and *trnH-psbA*, the number of monophyletic families were 49/61(80.33%), 40/49 (81.63%) and 40/47 (85.11%), respectively. In neighbor joining (N/J) trees of *rbcL*, *matK* and *trnH-psbA*, the number of monophyletic families were 49/61(80.33%), 40/49 (81.63%) and 40/47 (85.11%), respectively. Therefore, on the basis of monophyletic recovery and node support, all three neighbor joining trees were found best resolved monophyletic trees having more than 80% monophyletic families.