Single nucleotide polymorphism analysis in plastomes of eight *Catharanthus roseus* cultivars

Rania M. Makkia, Areej A. Saeedia, Thana K. Khan, Hani M. Ali and Ahmed M. Ramadan

**ABSTRACT**

*Catharanthus roseus*, an important medicinal plant, is known for the production of various pharmaceutical compounds including the anti-tumour drugs, vinblastine, and vincristine. Also, this ornamental plant is widely known for its flowers with different colours. Morphological characterization cannot be performed before the flowering stage, leading to a major obstacle for consumers, since most of the sales associated with its medicinal value occur at the vegetative stage. In the present study, we utilized high-throughput, next-generation sequencing to detect chloroplast single nucleotide polymorphism (SNP) that are unique to each variety for molecular characterization. The total genomic DNA of eight *C. roseus* varieties were sequenced using Illumina HiSeq 2000 platform. The alignment of resulting sequences to chloroplast reference genome showed the presence of specific SNPs for all eight varieties. Also, intravarietal SNPs were found that confirmed the applicability of heteroplasmy theory in the chloroplast genome of this species. Thus, this investigation provides valuable insights into molecular characterization of *C. roseus*, especially at the vegetative stage.

**ARTICLE HISTORY**

Received 19 October 2018
Revised 29 January 2019
Accepted 4 February 2019

**KEYWORDS**

*Catharanthus roseus*; chloroplast genome; hot spot SNPs regions; heteroplasmy

**Introduction**

The tropical plant *Catharanthus roseus*, also known as *Vinca rosea* or Madagascar periwinkle, is a flowering species that belongs to the family Apocynaceae. It originated in the Madagascar Island and has widely spread to warmer areas including Saudi Arabia [1]. This plant is extensively studied for its pharmacological content and medicinal importance. It has traditionally been used to treat diabetes, diarrhea, scurvy, hypertension, helminthes infection, ulcer, chronic wounds, and loss of memory [2]. The past decade has witnessed increased use of metabolites obtained from this plant as anti-cancer drugs with yearly revenue of 25 to 300 million U.S. dollars [1]. *C. roseus* is one of the major sources of terpenoid indole alkaloids (TIAs), which are secondary metabolites and synthesized by the plant as a means of defense against insects and pests [3]. Approximately, 130 TIAs with pharmaceutical properties [4] are produced by the plant, including the powerful species-specific anti-tumour drugs, vinblastine, and vincristine [5]. These drugs are utilized as a part of chemotherapy to treat a broad range of cancers, which is attributable to their distinctive mechanisms to block mitosis [6].

Apart from its anti-cancer properties, *C. roseus* is considered as a widespread ornamental plant owing to its various and attractive flower colours with an excellent ability to tolerate dry and nutrition-deficient conditions [7]. The implementation of breeding programs has resulted in a broad spectrum of flower petal colours, including white, pink, scarlet, reddish-orange, peach, and purple with flower eyes of various shades such as white, red, pink, deep pink, pale-yellow, and many more [4]. Many studies have concluded that the varieties with coloured flowers yield more alkaloids than those producing colourless flowers. For example, *C. roseus var rosea* (pink) has a higher alkaloid content than *C. roseus var alba* (white) [8–10]. The major obstacle is the inability to morphologically characterize the varieties before the flowering stage. This causes problems for consumers since most of the sales associated with its medicinal features take place at the
vegetative stage. The limitation has prompted investigators to utilize molecular markers for characterization of the varieties during the past decade. Although molecular markers based on fragment data such as RAPD, ISSR, SSR, EST-SSR and AFLP have many advantages, they also suffer from some drawbacks including limited repeatability and variation in fragment size accuracy [11]. Nonetheless, these limitations have been overcome with the recent revolution in molecular makers and their detection platforms. In this regard, the emergence of next-generation sequencing has facilitated the use of single nucleotide polymorphism (SNP) as a recently developed molecular marker [12].

Single nucleotide polymorphism involves a variation in a single nucleotide and at a specific position in the genome of individuals. It has received much attention because it is the most accurate, abundantly present in the genomes, and is a potential molecular marker for high-throughput detection platforms [13]. These advantages make SNP detection a powerful approach to discover relationships between varieties and identify the genetic basis of commercially important traits [14]. The publication of C. roseus chloroplast genome [15] provided a reference genome to examine chloroplast SNPs that are unique to each variety and determine phylogenetic relationships among varieties. SNPs in chloroplast genome could be inter- or intravarietal. The detection of intravarietal SNPs is referred to as heteroplasmy phenomenon and is commonly present in the chloroplast genome [16].

The present study investigated the (1) reliability of using chloroplast SNP variants to identify C. roseus varieties and (2) applicability of the heteroplasmy theory in chloroplast genome, and (3) investigated the phylogenetic distances among varieties.

**Materials and methods**

**Plant material and sampling**

The present study included eight varieties of C. roseus obtained from a plant nursery in the Makkah region, Jeddah, Saudi Arabia. Each variety is characterized by a distinct colour of the petal, flower eye, and center (Table 1, Figure 1). Flower colours have been described in detail previously [17, 18]. Samples were kept in the greenhouse under normal conditions that included 14 h of light per day at 22°C and 80% humidity and irrigated with full-strength Hoagland solution for 2 weeks until the samples started blooming. Samples from a single plant of each eight variety in three replicates were harvested and flash-frozen in liquid nitrogen and stored at −80°C for genomic DNA extraction. Each sample was about 50 mg of fresh and young leaf tissue.

**Genomic DNA extraction and purification**

Total DNA was extracted from leaf tissues using DNeasy plant mini kit (no. 69104) purchased from Qiagen. To achieve maximum concentration, DNA from three samples of each variety was mixed and applied to DNA SpeedVac concentrator. The Qubit 2.0 fluorometer was used for evaluating DNA concentration. High-quality DNA with a concentration between 20 and 40 ng/μL and OD ratio (260/280) close to 1.8 was finally used for further analysis.

**Genome sequencing, reference assembly and SNP analysis**

Total genomic DNA samples were transported in dry ice to the Beijing Genomics Institute (BGI), Shenzhen, China for deep sequencing using Illumina HiSeq 2000 platform. The received raw data for the eight varieties consisted of about 70 million, 100 bp paired-end reads. Before analyzing the raw data, adapter sequences were trimmed and reads with low-quality bases were filtered. High quality reads (≥20) from each variety were then assembled separately to C. roseus chloroplast reference genome (NCBI accession no. KC561139) using the CLC Genomics Workbench 3.6.5 software with parameters of length fraction ≥ 50% and identity ≥ 80%. The same software was used to identify SNPs for each chloroplast genome. The SNP detection
parameters were average coverage = 500, central base quality ≥ 20, and SNP frequency ≥ 35%.

**Multiple sequence alignment and phylogenetic analysis**

Fragments from nine chloroplast genomes of *C. roseus* varieties (eight from this study and one as a reference from GenBank) were aligned using the CLC Genomics Workbench. The fragments were selected depending on the presence of SNPs. Maximum likelihood (ML) phylogenetic tree for chloroplast SNPs was constructed using the CLC Genomics Workbench to estimate the genetic distances among them.

**Results and discussion**

**Mapping of reads to reference genome**

The total reads of all eight *C. roseus* varieties ranged from 73.06 to 75.40 million reads. Mapping the reads to chloroplast reference genome (NCBI accession no. KC561139) covered approximately 99.95% to 100% of the genome (154,950 bp). The number of mapped reads to chloroplast genome varied from 9,309,744 to 12,998,306 that represented 12.58 to 17.39% of the total reads (Table 2, Supplemental Figure S1). The average coverage of chloroplast reads ranged from 5,387.75 to 7,491.05.

**Table 2. Summary of alignment results to the chloroplast reference genome (GenBank No. KC561139).**

| Variety | Total no. reads | No. mapped reads | Coverage | % mapped reads |
|---------|-----------------|------------------|----------|----------------|
| PW      | 73,925,432      | 12,156,126       | 7,003.41 | 16.44          |
| FKD     | 73,995,072      | 10,959,007       | 6,340.78 | 14.81          |
| FKP     | 74,686,038      | 11,831,482       | 6,847.73 | 15.84          |
| ER      | 73,981,268      | 9,309,744        | 5,387.75 | 12.58          |
| ED      | 75,019,248      | 11,661,747       | 6,745.48 | 15.54          |
| CO      | 74,722,670      | 12,998,306       | 7,491.05 | 17.39          |
| VR      | 75,406,360      | 10,571,820       | 6,115.36 | 14.01          |
| BP      | 73,061,630      | 12,235,439       | 7,040.31 | 16.74          |

**Variety-specific SNPs and phylogenetic relationship**

Although extensive research has been carried out on *C. roseus*, none has utilized SNPs to characterize varieties. All previous studies on *C. roseus* were focused on molecular markers based on fragment data, such as RAPD, ISSR, SSR, EST-SSR and AFLP [17–23]. The emerging technology of high-throughput, next-generation sequencing, such as Illumina HiSeq, has facilitated the discovery of SNPs in different genomes and eliminated the drawbacks of low-throughput approaches [12]. The use of organellar SNPs has seen a tremendous increase [24, 25] owing to the relatively small size of the chloroplast, uniparental inheritance, and repeatability within a single cell as compared to nuclear genome [26]. The literature reports few studies that have employed chloroplast genome to detect SNP variants [16, 27–31] however, these studies could...
detect only a low level of genetic polymorphism and were limited owing to relatively lower level of mutation rate in chloroplast genomes compared to nucleus and mitochondria. Thus, it acted as a major barrier to detection of chloroplast markers [32, 33].

In this study, the number of chloroplast SNPs ranged from 24 to 84 per variety with total 429 SNPs among all eight varieties (Supplemental Table S1). Unique SNPs varied from 1 to 36 per variety with a total of 82 unique SNPs across the varieties (Figure 2, Table 3). The highest number of total (84) and unique (36) SNPs was recorded for PW variety, whereas the lowest number of total (24) and unique (1) SNPs was recorded for ED variety (Figure 2). Approximately 40% of the SNPs were heteroplasmic owing to the presence of both reference and alternate nucleotides in the reads of a given variety. The number of shared SNPs was 347 among varieties, 51 SNPs shared by two varieties, 31 by three varieties, 7 by four varieties, 13 by five varieties, 5 by six varieties, 3 by seven varieties, and 1 by all varieties. SNPs were located in 193 distinct locations within the chloroplast genome with 30 in intragenic regions, 39 in introns, 89 in intergenic spacers (IGS), 14 in rRNA-related sequences and 21 in tRNA-related sequences. The intragenic SNPs resulted in 10 synonymous (S) and 20 non-synonymous (NS) substitutions (Figure 3).

In general, the observations of the present study are consistent with previous data in terms of small number of SNPs detected in chloroplast genomes. However, our results revealed a higher number of SNPs as compared to those detected by the most recent analogous study [16]. The latter group compared nine date palm cultivars for organellar SNPs but could reveal only 30 chloroplast SNPs including 15 unique SNPs in seven out of nine cultivars [16]. In the current study, we could successfully discover molecular signature by variety-specific chloroplast SNPs for all eight C. roseus varieties (Figure 2). A possible explanation might be that C. roseus varieties are wild germplasms with a high rate of genetic variation as compared to the cultivated crops that have been subjected to domestication and homology processes [34]. Another possible explanation is that SNPs were retrieved under stringent methodology used by [16], which led to the reduction in the number of SNP detected.

**Location of organellar single nucleotide polymorphism**

The number of chloroplast SNPs ranged from 24 to 84 per variety with a total of 429 SNPs among all eight varieties (Supplemental Table S1). Unique SNPs varied from 1 to 36 per variety with a total of 82 unique SNPs across the varieties (Figure 2, Table 3). The highest number of total (84) and unique (36) SNPs was recorded for PW variety, whereas the lowest number of total (24) and unique (1) SNPs was recorded for ED variety (Figure 2). Approximately 40% of the SNPs were heteroplasmic owing to the presence of both reference and alternate nucleotides in the reads of a given variety. The number of shared SNPs was 347 among varieties, 51 SNPs shared by two varieties, 31 by three varieties, 7 by four varieties, 13 by five varieties, 5 by six varieties, 3 by seven varieties, and 1 by all varieties. SNPs were located in 193 distinct locations within the chloroplast genome with 30 in intragenic regions, 39 in introns, 89 in intergenic spacers (IGS), 14 in rRNA-related sequences and 21 in tRNA-related sequences. The intragenic SNPs resulted in 10 synonymous (S) and 20 non-synonymous (NS) substitutions (Figure 3).

![Figure 2](image-url)  
**Figure 2.** Number of total and unique SNPs detected for each C. roseus variety. Note: More information about variety abbreviations is provided in Table 1.
in 10 synonymous (S) and 20 non-synonymous (NS) substitutions (Figure 3).

These results agree with the phenomenon suggesting that substitutions evolve rapidly in non-coding sequences due to the lack of direct functional constraints. Compared with non-coding regions, coding regions are characterized by conservation and low rate of substitutions [35]. A noteworthy observation in the present study was that non-synonymous SNPs (nsSNPs) were located in six different genes, namely ycf2, ycf15, rpl2, rps12, rpl23 and ndhB. The first two genes are hypothetical chloroplast protein-coding regions with unknown functions. The genes ycf1 and ycf15 consist of 12 and 4 nsSNPs, respectively. This is in accordance with the studies that have suggested that ycf1 and ycf15 genes strongly display high rates of mutation that can be attributed to them being targets for positive natural selection [36, 37]. Only one nsSNP encountered in rpl2, rps12 or rpl23 gene encoding ribosomal proteins has been shown to be involved in the assembly of ribosomal subunits. The ndhB gene also contained one nsSNP in PW variety. The gene encodes for NADH dehydrogenase subunit 2, which has a role in photosynthesis. A previous study by [38] in tobacco has demonstrated a moderate decline in photosynthesis, leading to growth retardation in ndhB-inactivated plants. It is interesting to mention that PW variety has a lower rate of photosynthesis and growth compared to other varieties [8]. It can thus be hypothesized that the reduction of PW variety in terms of photosynthesis and growth may correspond to deleterious substitution in the ndhB gene.

**Single nucleotide polymorphism hotspots**

Another obvious finding was the uniform distribution pattern of SNPs in the SNP hotspots of the chloroplast genome. Hotspots were found in 10 different positions within chloroplast genomes with size ranging from 40 to 520 bp. All hotspots occupied non-coding regions including introns, IGS, pseudogenes and tRNA- and rRNA-related sequences. Eight of them were shared among varieties, and only two hotspots were present in single varieties. The number of shared SNPs in

---

**Table 3. Unique SNPs and their positions in the chloroplast genome.**

| Variety | Number of unique SNPs | Positions |
|---------|-----------------------|-----------|
| PW      | 36                    | 286, 293, 4797, 32316, 32657, 32660, 32673, 32675, 32683, 32714, 37702, 37703, 43547, 54156, 77058, 77059, 86366, 86388, 86492, 97250, 97254, 97257, 97293, 97298, 111009, 111012, 111335, 111334, 111359,111378,115013, 121734, 121736, 144530, 145946. |
| FKD     | 4                     | 86639, 86669, 101836, 131548 |
| FKP     | 1                     | 145811 |
| ER      | 2                     | 52027, 131530 |
| ED      | 1                     | 145952 |
| CO      | 18                    | 6362, 32324, 32447, 32519, 32534, 32571, 32589, 52142, 69044, 95377, 95393, 95396, 99952, 99970, 121738, 130572, 131771, 154031 |
| VR      | 13                    | 109364, 109367, 109374, 109390, 109413, 109421, 109431, 109436, 143709, 145320, 145339, 145368, 145385 |
| BP      | 7                     | 33205, 52132, 60684, 93301, 93308, 93319, 133426 |

**Figure 3.** Locations of SNPs in the chloroplast genome. IGS, intergenic spacer; NS, non-synonymous; S, synonymous; I, intron; tRNA, tRNA-related sequence; rRNA, rRNA-related sequence. Note: More information about variety abbreviations is provided in Table 1.
Table 4. Summary of SNPs hotspots analysis results.

| No. | Hotspot position (size) | Location | Variety | No. SNPs |
|-----|-------------------------|----------|---------|----------|
| 1   | ≈ 6,340 .. 6,400 (60 nt)| Between rps16 and trnQ-UGU genes | CO      | 7        |
| 2   | ≈ 32,260 .. 32,720 (460 nt)| Downstream, within and upstream of trnT-GGU gene | PW, FKD, FKP, ER, ED, CO and VR | 27–37 |
| 3   | ≈ 51,960 .. 52,160 (200 nt)| Between ndhC and trnV-UAC genes | PW, FKP, ER, ED, CO and BP | 5–7 |
| 4   | ≈ 86, 340 .. 86,680 (340 nt)| In intron of rpl2 gene | PW, FKD, ER and ED | 5–9 |
| 5   | ≈ 99,920 .. 100,000 (80 nt)| In CDS and upstream of rps12 gene | FKP and ER | 5 |
| 6   | ≈ 109,360 .. 109,440 (80 nt)| Within trnR-ACG gene | CO and VR | 5–11 |
| 7   | ≈ 111,000 .. 111,380 (380 nt)| Within and upstream of ycf1p gene | PW | 9 |
| 8   | ≈ 131,520 .. 131,840 (320 nt)| Downstream of rns5 gene and between rns5 and rns4.5 genes | FKD, FKP, ER, CO and VR | 6–9 |
| 9   | ≈ 132,920 .. 133,440 (520 nt)| Within rns23 gene | FKP and BP | 7-9 |
| 10  | ≈ 154,020 .. 154,060 (40 nt)| In intron of rpl2 gene | ER, VR and BP | 5 |

hotspots was different from one variety to the other (Table 4, Supplemental Figure S2.1–S2.33). Generally, even smaller clusters of four SNPs were considered not to be randomly distributed [39]. The reason behind the formation of SNP hotspots has remained elusive with many hypotheses proposed. First, natural selection can create regions with non-random increased variability by balancing selection [40]. Second, some regions of the genome feature relatively high conservation causing other regions to appear with variability above average [39].

**Chloroplast INDELS**

Several deletion/insertion polymorphisms (DIPs) were generated during the assembly of each of the eight C. roseus varieties. They ranged from 3 to 16 DIPs for each variety with a total of 81 DIPs across varieties. The deletions and insertions represented 63 and 37% of the total DIPs, respectively. All but Deep Pink variety had at least one unique DIP with a total of 26 DIPs (Table 5). DIPs were present in 44 different locations, mostly in non-coding regions and only nine were in coding regions that resulted in frame-shift mutations.

This is attributed to the fact that the chances of substitution are higher close to Indels in eukaryotes as they enhance mutation in surrounding sequences [41, 42]. However, the existence of DIPs in coding regions results in deleterious effects on the organism [42]. DIP density in the coding regions in this study was 20.4% relative to non-coding regions.

**Phylogenetic analysis**

Maximum likelihood (ML) phylogenetic tree was generated for the eight C. roseus varieties based on chloroplast SNPs (Figure 4). The results indicated that the Patricia White variety was less similar to other varieties. Also, no relationship was detected between petal colours and the ML tree topology. On the other hand, the moderate TIA accumulation varieties (First Kiss Peach and First Kiss Polka Dot) were grouped together in the tree. High TIA varieties in the tree were sub-grouped depending on other traits (CO and BP), (VR and ER), and (ED and REF). The most related varieties were Cooler Orchid and Blue Pearl, and the most distinct were Victory Red and Patricia White.

The phylogenetic ML tree of chloroplast SNPs was inadequate in terms of colours of a flower petal since it showed a distinct genetic relationship among varieties with similar flower colours. However, the ML tree provided some information toward partial separation of varieties with different center colours.

**Intravarietal single nucleotide polymorphism or heteroplasmy**

Heteroplasmy was present in the chloroplast genomes of many C. roseus varieties in the present study (Supplemental Table S1). The chloroplast heteroplasmy is a phenomenon that has been detected in many flowering plants, for example, *Pelargonium* [43], *Gossypium* [44], *Oenothera* [45], *Medicago* [46], *Actinidia* [47], *Cynomorium* [48], *Passiflora* [49] and *Phoenix dactylifera* [16]. However, a concern has been raised about the commonness of this phenomenon in chloroplast genomes. As more molecular studies on organelle genomes were conducted, more evidence was gathered to support and/or prove chloroplast heteroplasmy to be a common and regular phenomenon. Heteroplasmy occurs in two circumstances, either biparental or uniparental chloroplast inheritance. In the case of biparental inheritance, both parents transmit their chloroplasts to the zygote resulting in heteroplasmy. Heteroplasmy, in uniparental inheritance, is caused by either mutation in chloroplast genomes after zygote formation or variation in parental chloroplast owing to incomplete vegetative segregation [50]. In C. roseus, heteroplasmy in chloroplast genome is a result of uniparental inheritance [15].

**Conclusions**

Overall, this study detected chloroplast SNPs and found DIPs to be a reliable approach to identify C. roseus varieties. The study also confirmed the
Table 5. The total of chloroplast DIPs ordered by their position in the genome.

| Variety | Reference position | Reference | Allele variations | Frequencies | Counts | Coverage | Location |
|---------|--------------------|-----------|------------------|-------------|--------|----------|----------|
| CO 1714 | T                  | -         |                  | 75          | 3      | 4        | IGS      |
| FKD 4558 | -                  | -/G       |                  | 50.0/50.0   | 3/3    | 6        | IGS      |
| FKD 4560 | -                  | AAAT      |                  | 50          | 3      | 6        | IGS      |
| FKD 4662 | A                  | -         |                  | 100         | 5      | 5        | IGS      |
| CO 4778 | AA                 | -/T       |                  | 83.3        | 5      | 6        | IGS      |
| CO 4799 | T                  | -         |                  | 100         | 5      | 5        | IGS      |
| BP 5830 | T                  | -         |                  | 100         | 5      | 5        | IGS      |
| PW 7160 | -                  | T         |                  | 100         | 4      | 4        | IGS      |
| CO 7160 | T                  | -         |                  | 100         | 4      | 4        | IGS      |
| FKD 8126 | T                  | T/T/-    |                  | 60.0/40.0   | 3/2    | 5        | IGS      |
| BP 8126 | T                  | T/T/-    |                  | 75          | 3      | 4        | IGS      |
| PW 9926 | T                  | TATT     |                  | 75          | 3      | 4        | IGS      |
| PW 12946 | A                  | A/-/T    |                  | 50.0/50.0   | 4/4    | 8        | IGS      |
| FKD 12946 | A                  | A/-/T    |                  | 50.0/50.0   | 2/2    | 4        | IGS      |
| ED 12946 | A                  | -         |                  | 83.3        | 5      | 6        | IGS      |
| CO 12946 | A                  | A/-/T    |                  | 60.0/40.0   | 3/2    | 5        | IGS      |
| PW 12963 | T                  | -         |                  | 100         | 4      | 4        | IGS      |
| BP 12988 | TATTGAA             | -/TATTGAA|                  | 60.0/40.0   | 3/2    | 5        | IGS      |
| PW 21069 | -                  | -/A      |                  | 60.0/40.0   | 3/2    | 5        | IGS      |
| PW 28339 | -                  | TGTA     |                  | 80          | 4      | 5        | IGS      |
| CO 30975 | CC                 | -         |                  | 60.0/40.0   | 3/2    | 5        | IGS      |
| PW 31078 | TGTAC              | -         |                  | 100         | 4      | 4        | IGS      |
| CO 33347 | A                  | -         |                  | 100         | 4      | 4        | IGS      |
| CO 37707 | -                  | G         |                  | 100         | 5      | 5        | IGS      |
| BP 37707 | -                  | G         |                  | 100         | 4      | 4        | IGS      |
| FKD 44208 | T                  | T/T/-    |                  | 62.5/37.5   | 5/3    | 8        | IGS      |
| BP 44482 | T                  | -         |                  | 100         | 4      | 4        | IGS      |
| CO 45972 | -                  | -/AATCGAT|                  | 50.0/50.0   | 2/2    | 4        | IGS      |
| BP 45972 | -                  | TAATCGAT |                  | 80          | 4      | 5        | IGS      |
| CO 47635 | -                  | -/ATAG   |                  | 75          | 3      | 4        | IGS      |
| BP 47635 | -                  | -/ATAG   |                  | 50.0/50.0   | 2/2    | 4        | IGS      |
| PW 52934 | T                  | -         |                  | 100         | 4      | 4        | IGS      |
| CO 52934 | T                  | -         |                  | 75          | 3      | 4        | IGS      |
| BP 54032 | T                  | -         |                  | 75          | 3      | 4        | IGS      |
| CO 55424 | T                  | -/T      |                  | 85.7        | 6      | 7        | NS       |
| BK 55424 | -                  | T         |                  | 85.7        | 6      | 7        | NS       |
| FKD 55429 | -                  | T         |                  | 66.7        | 4      | 6        | NS       |
| BK 55429 | -                  | -/T      |                  | 60.0/40.0   | 3/2    | 5        | NS       |
| VR 55429 | -                  | -/T      |                  | 62.5/37.5   | 5/3    | 8        | NS       |
| FKD 55469 | CC                 | -         |                  | 77.8        | 7      | 9        | NS       |
| ER 55469 | CC                 | -         |                  | 75          | 3      | 4        | NS       |
| ED 55469 | CC                 | -         |                  | 66.7        | 4      | 6        | NS       |
| PW 61304 | -                  | A         |                  | 80          | 4      | 5        | IGS      |
| PW 61545 | T                  | -/T      |                  | 50.0/50.0   | 2/2    | 4        | IGS      |
| CO 67625 | A                  | -         |                  | 100         | 4      | 4        | IGS      |
| BP 69044 | C                  | -         |                  | 100         | 4      | 4        | IGS      |
| VR 72136 | -                  | -/T      |                  | 60.0/40.0   | 3/2    | 5        | IGS      |
| PW 79096 | T                  | -         |                  | 85.7        | 6      | 7        | NS       |
| FKD 79096 | T                  | -         |                  | 100         | 6      | 6        | NS       |
| FKP 79096 | T                  | T/T/-    |                  | 64.3/35.7   | 9/5    | 14       | NS       |
| ER 79096 | T                  | -         |                  | 76.9        | 10     | 13       | NS       |
| ED 79096 | T                  | -         |                  | 90          | 9      | 10       | NS       |
| CO 79096 | T                  | -         |                  | 90.9        | 10     | 11       | NS       |
| PW 79193 | T                  | -         |                  | 100         | 7      | 7        | NS       |
| FKD 79193 | T                  | -         |                  | 86.7        | 13     | 15       | NS       |
| FKP 79193 | T                  | T/T/-    |                  | 60.9/39.1   | 14/9   | 23       | NS       |
| ER 79193 | T                  | -         |                  | 80          | 4      | 5        | NS       |
| ED 79193 | T                  | -         |                  | 90          | 9      | 10       | NS       |
| CO 79193 | T                  | -         |                  | 85.7        | 6      | 7        | NS       |
| VR 79193 | T                  | T/T/-    |                  | 54.5/45.5   | 6/5    | 11       | NS       |
| BP 79193 | T                  | T/T/-    |                  | 60.0/40.0   | 3/2    | 5        | NS       |
| PW 82536 | G                  | G/-/T    |                  | 60.0/40.0   | 3/2    | 5        | NS       |
| FKP 82536 | G                  | G/-/T    |                  | 57.1/42.9   | 4/3    | 7        | NS       |

(continued)
applicability of the heteroplasmy theory in this species, thereby providing additional evidence. The study demonstrated a limited phylogenetic relationship among varieties with low support values. We believe that the study of the mitochondrial genome of C. roseus will add to our understanding of the utility of SNPs and DIPs in distinguishing closely related varieties. It will also throw light on the mechanism of heteroplasmy. However, the lack of information regarding C. roseus mitochondrial genome limits further examination of the genome for SNPs and DIPs. In addition, research could be conducted on nuclear genome owing to a higher prevalence of SNPs, which, in turn, could be suitable for constructing a high-resolution phylogenetic ML tree and therefore a better picture of relationships among C. roseus varieties.

Disclosure statement
The authors have no conflict of interest to declare

Funding
This project was funded by the Deanship of scientific Research (DSR) at King Abdulaziz University, Jeddah, under grant no. G-318-247-38. The authors, therefore, acknowledge with thanks DSR technical and financial support.

References

[1] Schmelzer G, Gurib-Fakim A, Arroo R, et al. Plant Resources of Tropical Africa 11(1): Medicinal plants 1. Wageningen(NL): PROTA Foundation; 2008.
[2] Balaji H. Versatile therapeutic effects of Vinca rosea Linn. Int J Pharm Sci Health Care. 2014;1:59–76.
[3] Ziegler J, Facchini PJ. Alkaloid biosynthesis: metabolism and trafficking. Annu Rev Plant Biol. 2008;59:735–769.
[4] Heijden R, Jacobs D, Snoeijer W, et al. The Catharanthus roseus alkaloids: pharmacognosy and biotechnology. Curr Med Chem. 2004;11:607–628.
[5] Guirimand G, Courdavault V, Lanoue A, et al. Strictosidine activation in Apocynaceae: towards a "nuclear time bomb"?. BMC Plant Biol. 2010;10:182–201.
[6] Pasquier E, Kavallaris M. Microtubules: a dynamic target in cancer therapy. JUBMB Life. 2008;60:165–170.
[7] Nejat N, Valdiani A, Cahill D, et al. Ornamental exterior versus therapeutic interior of Madagascar Periwinkle (Catharanthus roseus): the two faces of a versatile herb. Sci World J. 2015;2015:1. [cited 2018 Nov 15]; Article ID 982412[19 pages].
[8] Idrees M, Naeem M, Khan MM. The superiority of cv ‘rosea’over cv ‘alba’of periwinkle (Catharanthus roseus L.) in alkaloid production and other physiological attributes. Turk J Biol. 2010;34:81–88.
[9] Bhutkar MA, Bhise SB. Comparative studies on antioxidant properties of Catharanthus rosea and Catharanthus alba. Int J Pharm Tech Res. 2011;3:1551–1556.
[10] Sharma V, Chaudhary S, Srivastava S, et al. Characterization of variation and quantitative trait loci related to terpenoidindole alkaloid yield in a recombinant inbred line mapping population of Catharanthus roseus. J Genet. 2012;91:1–21.
[11] Sharma R, Joshi A, Maloo SR, et al. Assessment of genetic finger printing using molecular marker in plants: a review. Sci Res Impact. 2012;1:29–36.
[12] Raza S, Shoabi MW, Mubeen H. Genetic Markers: importance, uses and applications. Int J Sci Res. 2016;6:2250–3153.
[13] Mammadov J, Aggarwal R, Buyyaramu R, et al. SNP markers and their impact on plant breeding. Int J Plant Genomics. 2012;2012:1.

[14] Henry RJ. Plant genotyping II: SNP technology. Wallingford (UK): CABI Publishing; 2008.

[15] Ku C, Chung WC, Chen LL, et al. The complete chloroplast genome sequence of Madagascar periwinkle Catharanthus roseus (L.) G. Don: chloroplast genome evolution, molecular marker identification, and phylogenetic implications in asterids. PLoS One. 2013;8: e68518.

[16] Sabir JS, Arabappan D, Bahieldin A, et al. Whole mitochondrial and chloroplast genome SNP analysis of nine date palm cultivars reveals chloroplast heteroplasmy and close phylogenetic relationships among cultivars. PLoS One. 2014;9:e94158.

[17] Shaw RK, Acharya L, Mukherjee AK. Assessment of genetic diversity in a highly valuable medicinal plant Catharanthus roseus using molecular markers. Crop Breed Appl Biotechnol. 2009;9:52–59.

[18] El-Domyati FM, Ramadan AM, Gadalla NO, et al. Identification of molecular markers for flower characteristics in Catharanthus roseus producing anticancer compounds. Life Sci J. 2012;9:5949–5960.

[19] Gupta K, Pandey-Rai S, Srivastava S, et al. Construction of genetic linkage map of the medicinal and ornamental plant Catharanthus roseus. J Genet. 2007;86:259–268.

[20] Shokeen B, Choudhary S, Sethy NK, et al. Development of SSR and gene-targeted markers for construction of a framework linkage map of Catharanthus roseus. Ann Bot. 2011;108:321–336.

[21] Lal S, Mistry KN, Shah SD, et al. Genetic diversity assessment in nine cultivars of Catharanthus roseus from Central Gujarat (India) through RAPD, ISSR and SSR markers. J Biol Res. 2011;1:667–675.

[22] Mishra RK, Gangadhar BH, Yu JW, et al. Development and characterization of EST based SSR markers in Madagascar periwinkle (Catharanthus roseus) and their transferability in other medicinal plants. Plant Omics. 2011;4:154.

[23] Salama IM, Ali GM. Genetic variant detected by RAPD-PCR and ISSR in Catharanthus roseus (L.) cells exposed to low doses of gamma rays. Egypt J Radiat Sci Appl. 2016;29:85–101.

[24] Agarwal M, Shrivastava N, Padh H. Advances in molecular marker techniques and their applications in plant sciences. Plant Cell Rep. 2008;27:617–631.

[25] Pérez-Zamorano B, Vallebueno-Estrada M, Martinez González J, et al. Organelar genomes from a ~5,000-year-old archaeological maize sample are closely related to NB genotype. Genome Biol Evol. 2017;9:904–915.

[26] Petit RJ, Duminil J, Fineschi S, et al. Comparative organization of chloroplast, mitochondrial and nuclear diversity in plant populations. Mol Ecol. 2004;14:689–701.

[27] Whittall JB, Syring J, Parks M, et al. Finding a (pine) needle in a haystack: chloroplast genome sequence divergence in rare and widespread pines. Mol Ecol. 2010;19:100–114.

[28] Doorduin L, Gravendeel B, Lammers Y, et al. The complete chloroplast genome of 17 individuals of pest species Jacobaea vulgaris: SNPs, microsatellites and barcoding markers for population and phylogenetic studies. DNA Res. 2011;18:93–105.

[29] McPherson H, Van der Merwe M, Delaney SK, et al. Capturing chloroplast variation for molecular ecology studies: a simple next generation sequencing approach applied to a rainforest tree. BMC Ecol. 2013;13:8.

[30] Melodelima C, Lobréaux S. Complete Arabidopsis thaliana chloroplast genome sequence and insight into its polymorphism. Meta Gene. 2013;1:65–75.

[31] Zhu Q, Gao P, Liu S, et al. Comparative analysis of single nucleotide polymorphisms in the nuclear, chloroplast, and mitochondrial genomes in identification of phylogenetic association among seven melon (Cucumis melo L.) cultivars. Breed Sci. 2016;66:711–719.

[32] McCauley DE. The use of chloroplast DNA polymorphism in studies of gene flow in plants. Trends Ecol Evol (Amst). 1995;10:198–202.

[33] Yan M, Xiong Y, Liu R, et al. The application and limitation of universal chloroplast markers in discriminating East Asian Evergreen Oaks. Front Plant Sci. 2018;9:569.

[34] Zhang H, Mittal N, Leamy LJ, et al. Back into the wild: Apply untapped genetic diversity of wild relatives for crop improvement. Evol Appl. 2017;10:5–24.

[35] Skuzu L, Filip E, Szucko I. Chapter 26. Use of organelle markers to study genetic diversity in soybean. In: Board J, editor. A comprehensive survey of international soybean research. Genetics, Physiology, Agronomy and Nitrogen Relationships. London (UK): Intech Open; 2013. p. 553–581.

[36] Parks MB. Plastome phylogenomics in the genus Pinus using massively parallel sequencing technology. [dissertation]. Corvallis (OR): Oregon State University; 2011.

[37] Zhang H, Li C, Miao H, et al. Insights from the complete chloroplast genome into the evolution of Sesamum indicum L. PLoS One. 2013;8:e80508.

[38] Horváth EM, Peter SO, Joët T, et al. Targeted inactivation of the plastid ndhF gene in tobacco results in an enhanced sensitivity of photosynthesis to moderate stomatal closure. Plant Physiol. 2000;123:1337–1350.

[39] Amos W. Even small SNP clusters are non-randomly distributed: is this evidence of mutational non-independence?. Proc Biol Sci. 2010;277:1443–1449.

[40] Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection, balanced polymorphism and background selection on equilibrium patterns of genetic diversity in subdivided populations. Genet Res. 1997;70:155–174.

[41] Tian D, Wang Q, Zhang P, et al. Single-nucleotide mutation rate increases close to insertions/deletions in eukaryotes. Nature. 2008;455:105–108.

[42] Chen QJ, Wu Y, Yang H, et al. Variation in the ratio of nucleotide substitution and indel rates across genomes in mammals and bacteria. Mol Biol Evol. 2009;26:1523–1531.
[43] Tilney-Bassett RAE, Birky CW. The mechanism of the mixed inheritance of chloroplast genes in Pelargonium: evidence from gene frequency distributions among the progeny of crosses. Theor Appl Genet. 1981;60:43–53.

[44] Lax AR, Vaughn KC, Duke SO, et al. Structural and physiological studies of a plastome cotton mutant with slow sorting out. J Hered. 1987;78:147–152. 1987

[45] Chiu WL, Stubbe W, Sears BB. plastid inheritance in Oenothera: organelle genome modifies the extent of biparental plastid transmission. Curr Genet. 1988;13:181–189.

[46] Johnson LB, Palmer JD. Heteroplasmy of chloroplast DNA in Medicago. Plant Mol Biol. 1989;12:3–11.

[47] Chat J, Decrooq S, Decrooq V, et al. A case of chloroplast heteroplasmy in kiwifruit (Actinidialdelicosa) that is not transmitted during sexual reproduction. J Hered. 2002;93:293–300.

[48] García MA, Nicholson EH, Nickrent DL. Extensive intraindividual variation in plastid rDNA sequences from the holoparasite Cynomorium coccineum (Cynomoriiaceae). J Mol Evol. 2004;58:322–332.

[49] Hansen AK, Escobar LK, Gilbert LE, et al. Paternal, maternal, and biparental inheritance of the chloroplast genome in Passiflora (Passifloraceae): implications for phylogenetic studies. Am J Bot. 2007;94:42–46.

[50] Chesser RK. Heteroplasmy and organelle gene dynamics. Genetics. 1998;150:1309–1327.