Molecular insights into genetic diversity and population dynamics of five medicinal *Eulophia* species: a threatened orchid taxa of Africa

Paromik Bhattacharyya¹ · Johannes van Staden¹

Received: 11 May 2017 / Revised: 16 November 2017 / Accepted: 5 March 2018 / Published online: 16 May 2018
© Prof. H.S. Srivastava Foundation for Science and Society 2018

Abstract Genetic diversity existing amongst five *Eulophia* orchid species were assessed using start codon targeted polymorphism (SCoT) and inter-retrotransposon amplified polymorphism (IRAP) markers. A total of 12 SCoT and 5 IRAP markers revealed an average of 63% genetic variability [SCoT = 63.87; IRAP = 64.95%] amongst the five *Eulophia* species investigated. The genetic similarities were assessed using both UPGMA and Bayesian approaches which indicated identical clustering patterns at a genetic similarity level of 50%. Analysis of molecular variance (AMOVA) revealed the presence of a significant degree of genetic variability, mostly compartmentalized within the species level. Amongst the five assessed *Eulophia* species, *E. parviflora* was the most genetically diverse representative whereas *E. welwitschii* was found to be least diverse based on a comparative assessment of various population genetic parameters like Nei's gene diversity (*h*) and Shannon's information index (*I*) with an overall gene flow value greater than 1. In order to evaluate the comparative marker efficiency, SCoT and IRAP marker data were subjected to various benchmark analyses like marker index, resolving power, polymorphic index content, multiplex ratio and effective multiplex ratio which revealed the robustness of both the marker techniques in assessment

of genetic diversity. The present report provides the first molecular insights into the aspects of inter and intra specific genetic variability in medicinally as well as horticulturally important *Eulophia* species along with addressing their conservation concerns. In a nutshell, the present approach is simple, rapid and cost effective and can be extended for analysis of genetic diversity of other related plant species.

Keywords Gene targeted markers · African medicinal orchids · SCoT-PCR · IRAP-PCR · AMOVA

Abbreviations

PCR	Polymerase chain reaction
PIC	Polymorphic information content
AMOVA	Analysis of molecular variance
R _p	Resolving power

Introduction

Orchidaceae, is one of the largest and most diverse plant families comprising of approximately 800 genera and between 25,000 and 30,000 species spread all over the world and in almost all ecological environments (Chowdhery 2001). Amongst the various orchid species distributed worldwide, genus *Eulophia* deserves special mention. It is the largest genus in the subtribe Eulophiinae comprising of approximately 200 species of which 156 are distributed over sub-Saharan Africa and Western Indian Ocean Islands (WCSP 2016). This terrestrial genus of orchid exhibits tremendous morphological diversity and is distributed in a wide range of habitats which is also reflected in their phenotypic diversity (Martos et al. 2014). However, to assess the relationships amongst various species of orchid,

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s12298-018-0523-6>) contains supplementary material, which is available to authorized users.

✉ Johannes van Staden
Rcpgd@ukzn.ac.za

¹ Research Centre for Plant Growth and Development, School of Life Sciences, University of KwaZulu-Natal, Pietermaritzburg, Private Bag X01, Scottsville 3209, South Africa

the conventional method based on morphological traits of leaf and flower is used traditionally (Jin and Yao 2006). The classical approach is influenced by the plant growth conditions and individual biases which makes the task of novel genotype/cultivar identification difficult for botanists. Thus, systematic knowledge about the orchids genetic resources is important for formulation of conservation strategies and also the qualitative improvement of the genus *Eulophia* and related orchid species (Sharma et al. 2011).

Amongst the various *Eulophia* species distributed worldwide, *E. parviflora*, *E. streptopetala*, *E. welwitschii*, *E. speciosa* and *E. clavicornis* have special importance as they are being used in Traditional African Pharmacopeia (TAP) preparations primarily in the treatment of infertility (Chinsamy et al. 2011). The whole plant of *E. parviflora* is being used as love or protective charm by the traditional healers of Africa (Cunningham 1988). Similarly, *E. streptopetala* (tubers and roots) and *E. welwitschii* (tubers) is also traded in the treatment of infertility but in a lesser volume (Hulme 1954; Hutchings 1996). Likewise, *E. speciosa* finds an important role in Emetics. In African pharmacopeia, use of emetics in the treatment of diseases has played an important role. Emetics are agents which causes emesis by either inducing nausea or vomiting. The extracts of *E. speciosa* have been used particularly as an Emetic agent traditionally for both humans and animals (Gerstner 1941). In congruence with the other *Eulophia* species, *E. clavicornis* are used separately to create dolls that represent fertility (Hutchings 1996). The tubers are used as an infusion that is sprinkled around the home to ward off evil (Hulme 1954).

Advances in molecular marker technology in recent years have provided effective alternatives to classical morpho-phenotypic characterization techniques. As the molecular markers are not influenced by any physical or biochemical factors, they help in a vivid manner in cultivar identification in early developmental stages (Chen et al. 2009; Yu et al. 2009; Cabo et al. 2014). In the recent past, extensive research has been carried out on the aspects of genetic diversity in various orchid species at both inter-specific, as well as intra-specific, levels with special reference to the genus *Dendrobium* (Sharma et al. 2011; Manners et al. 2013; Bhattacharyya and Kumaria 2015; Bhattacharyya et al. 2013, 2015, 2017). However, till date there exists no scientific documentation of genetic diversity in any species of *Eulophia*. Amongst the various markers developed in the recent past, trait related markers namely start codon targeted polymorphism (SCoT) and inter-retrotransposons amplified polymorphisms (IRAP) deserves special mention.

SCoT was first reported by Collard and Mackill (2009) and is a very accurate, reliable and reproducible marker

system used to assess genomic diversity in a wide range of plant species including orchids (Xiong et al. 2011; Bhattacharyya et al. 2013; Mulpuri et al. 2013; Cabo et al. 2014; Feng et al. 2015). The primers are designed from the conserved region surrounding the translation initiation codon, ATG popularly known as translation start site (TSS). Unlike the conventional marker systems like RAPD, ISSR and AFLP, SCoT targets a specific gene and can generate molecular information which can be correlated with various biological traits. Like SCoT, the retrotransposon-derived marker system IRAP is another advanced marker system used in recent studies for assessment of genetic diversity (Kalendar et al. 1999; Kalendar and Schulman 2006). Retrotransposons (RTNs) are abundant and dispersed within the plant genomes which offer a tremendous basis for the improvement of molecular marker systems (Waugh et al. 1997).

Keeping in view the above discussed perspectives, in the present investigation SCoT and IRAP markers were used to evaluate genetic diversity at both inter- and intra-specific levels amongst 30 collections representing five *Eulophia* species (six collections of each species), viz. *Eulophia parviflora*, *E. streptopetala*, *E. welwitschii*, *E. speciosa* and *E. clavicornis*. All of these five plant species were sampled from South Africa and are mostly endangered and threatened in their natural habitats.

Materials and methods

Plant material

Five *Eulophia* species constituting of 6 individuals from each species viz. *E. parviflora* (EP1–EP6), *E. streptopetala* (ES1–ES6), *E. welwitschii* (EW1–EW6), *E. speciosa* (ESP1–ESP6) and *E. clavicornis* (EC1–EC6), respectively were collected from various parts of South Africa and were maintained in the Botanical Garden of the University of KwaZulu-Natal (UKZN). Leaf samples from 30 individual plants belonging to five plant species were used for genomic DNA extraction.

DNA extraction

Amplifiable quality genomic DNA was extracted from the leaf tissue samples using Qiagen DNAeasy plant DNA extraction kit (Germany) and the quality of the extracted DNA was checked by mupid gel electrophoresis with 1% (w/v) agarose in 1X Tris Acetate EDTA (TAE) buffer.

Molecular analysis using SCoT and IRAP

Forty-five SCoT primers were screened of which the 12 primers which proved to be most reproducible were chosen for the final amplification reaction. Similarly from 11 IRAP primers, the 5 best primers were screened. The PCR reactions for SCoT and IRAP were carried out in accordance to the protocol described by Bhattacharyya et al. (2016) with minor modifications. The end products of the PCR amplification reaction were resolved in 1.8% denaturing agarose gel using 1X TAE buffer under Syngene Gel DOC, Syngene, Synoptic Ltd., UK.

Data analysis

The visible, clear and unambiguous DNA bands generated by both SCoT and IRAP marker systems were recorded into a binary matrix on the basis of presence (1) and absence (0) of the band. Each polymorphic fragment generated by the molecular markers was considered as a different locus corresponding to a unique position in the genome. Using NTSYS version 2.02 k software package (Rohlf 1998), Jaccard's genetic similarity index was calculated and the dendrograms were generated through the SAHN clustering module using the unweighted pair group method (UPGMA). Also, the assignment of individuals to different groups and the existent relationships amongst the sampled genotypes were estimated with STRUCTURE 2.2 software (Pritchard et al. 2000; Falush et al. 2003, 2007) following the Bayesian clustering model. The number of genetically distinct clusters (K) was assumed to range between 1 and 6 and the model [Admixture model assuming "Allele frequency correlated amongst the populations"] was subjected to 100,000 independent runs for each K , with a subsequent burn-in length of 10,000 and run length of 50,000 iterations respectively. The most likely number of clusters present was estimated in accordance to the protocol given by Evanno et al. (2005) using the ΔK statistics based on the rate of change with respect to K in the log probability of data.

In order to determine the comparative marker efficiency, polymorphic information content (PIC), resolving power (R_p) value of the primers, multiplex ratio (MR), effective multiplex ratio (EMR) and marker index (MI) were calculated in accordance to the methods described by Powell et al. (1996) and Smith et al. (1997). Various population genetic parameters like Nei's gene diversity index (H), Shannon diversity index (I), percentage polymorphic bands (PPB), gene flow and other necessary genetic diversity parameters were calculated using POPGENE software; Ver. 1.31 (Yeh et al. 1999). At two hierarchical levels, analysis of molecular variance (AMOVA) using Arlequin version 3.01 (Excoffier et al. 2005) was performed in order

to determine the differences existent amongst and within population levels. F -statistics (F_{ST}) or fixation index was also calculated using Arlequin v. 3.01. The significance of this test was tested by 100,000 random permutations of sequences amongst populations (Miller 1998).

Results

SCoT and IRAP-PCR

The genetic diversity amongst the five species of *Eulophia* was measured using 12 SCoT primers producing a total of 105 fragments of which 83 were polymorphic (79.04%). The primers SCoT-(5, 7, 9, 13, 16 and 18) exhibited the highest level of polymorphism (Table 1; Fig. 1a). A maximum of 12 amplicons were produced by the primers SCoT-(7 and 9) and the minimum was 7 with primers SCoT-(2, 20 and 34) with an average of 6.9 fragments/primer. The PIC value for the SCoT primers was 0.40 whereas the R_p value of the primers ranged from 17.4 (SCoT-3) to 45.3 (SCoT-2); (Table 1).

Like SCoT, the IRAP primers LTR-2, 3 and 5 exhibited the highest levels of polymorphism (Table 1; Fig. 1b). On average, 6.1 bands were produced per primer with the maximum being 10 (LTR-3) and the minimum being 6 (LTR-6). The PIC value for the IRAP primers was 0.52 (Table 1). The R_p value of the primers ranged from 26.2 (LTR-10) to 38.7 (LTR-5).

The cumulative dataset of SCoT and IRAP marker matrices revealed that a total of 148 amplicons were produced; out of which 120 fragments were polymorphic (81.08%). With an average of 7.05 polymorphic bands per primer and the genetic distance calculated using Jaccard's similarity coefficient ranging from 0.34 to 0.85 (Table 1).

Cluster analysis

Based on the UPGMA clustering algorithm generated from the cumulative SCoT and IRAP marker dataset, the populations were grouped into two major clusters (Fig. 2a). The representatives from *E. parviflora*, *E. streptopetala*, and *E. welwitschii* were grouped into one major cluster, whereas those from *E. speciosa* and *E. clavicornis* were grouped into a second major cluster at a genetic similarity level of 50%. Mantel test for the combined SCoT and IRAP dataset was performed which revealed that the correlation coefficient between SCoT and IRAP markers was significant ($r = 0.91$) and the value of determination coefficient was also high ($r^2 = 0.82$; Table 1). The Bayesian modeling of the marker data using STRUCTURE programme revealed that the maximum data probability was achieved when the samples were clustered in 2 or 4 groups ($K = 2$

Table 1 Data of the SCoT and IRAP markers used in the present study and extent of polymorphism within five *Eulophia* species

SL.	Primer name	Primer sequence	TB	PB	PPB	Rp value	PIC	MI	R	R ²
<i>SCoT</i> ^a									0.91	0.82
1.	SCoT-2	CAACAATGGCTACCACGA	7	5	71.4	45.3	0.65	46.4		
2.	SCoT-3	CAACAATGGCTACCACGC	9	6	66.6	17.4	0.34	22.6		
3.	SCoT-5	CAACAATGGCTACCACGT	10	8	80.0	23.4	0.33	26.4		
4.	SCoT-7	CAACAATGGCTACCAGCC	12	10	83.3	19.3	0.28	23.3		
5.	SCoT-9	ACCATGGCTACCACCGAG	12	11	91.6	19.8	0.31	26.2		
6.	SCoT-11	ACCATGGCTACCACCGTC	10	7	70.0	27.2	0.53	37.1		
7.	SCoT-13	CCATGGCTACCACCGCAG	9	8	88.8	38.7	0.25	22.7		
8.	SCoT-16	GCAACAATGGCTACCACC	8	7	87.5	26.2	0.56	49.0		
9.	SCoT-18	ACCATGGCTACCACCGCC	6	5	83.3	29.1	0.27	22.5		
10.	SCoT-20	ACCATGGCTACCACCGGC	7	5	71.4	38.7	0.43	30.7		
11.	SCoT-31	CCATGGCTACCACCGCG	8	5	62.5	37.2	0.41	29.2		
12.	SCoT-34	CCATGGCTACCACCGCAG	7	6	85.7	39.4	0.46	31.2		
<i>IRAP</i> ^b										
13.	LTR-2	CTTGCTGGAAAGTGTGTGAGAGG	10	8	80.0	26.7	0.98	78.4		
14.	LTR-3	TGTTAATCGCGCGCTCGGGTGGGAGCA	10	9	90.0	30.5	0.54	48.6		
15.	LTR-5	CTGGCATTTCATTGTGTCGTCGATGC	9	8	88.8	38.7	0.25	22.7		
16.	LTR-6	GCATCAGCCTGGACCAGTCCTCGTCC	6	5	83.3	29.1	0.27	22.5		
17.	LTR-10	TGAGTTGCAAGGTCCAGGCATCA	8	7	87.5	26.2	0.56	49.0		

TB, total no of bands; PB, total number of polymorphic bands; PPB, percentage polymorphic bands; Rp, resolving power; PIC, polymorphic information content; MI, marker index; R, correlation coefficient; R², determination coefficient

^aSCoT primers developed by Collard and Mackill (2009)

^bIRAP primers developed by Kalendar and Schulman (2006)

and 4; Fig. 2b–d; Table S2). These groupings were consistent with those derived from the UPGMA clustering.

Analysis of molecular variance (AMOVA)

The analysis of molecular variance (AMOVA) of the SCoT data revealed that out of the total recorded variation, 65.58% was recorded within the species level whereas 34.42% was compartmentalized amongst the species. Similar in distribution to that of SCoT, IRAP marker data accounted for 68.61% variability at inter-specific level and 31.39% at intra-specific level. AMOVA of the pooled matrix revealed 66.21% variability at inter-specific level and the remaining 32.29% at intra-specific level (Table 2). The fixation index value (F_{ST}) of the SCoT marker (0.325) was found higher than that of the IRAP marker (0.292) with an observed (F_{ST}) value of 0.315 for the pooled SCoT and IRAP matrix (Table 2).

Genetic diversity and population structure of five *Eulophia* species

The genetic diversity parameters like percentage polymorphism (PP), Nei's gene diversity (h), Shannon's

information index (I), observed number of alleles (N_a) and effective number of alleles were determined (Table 3). In the analysis of both SCoT and IRAP markers, the highest PP was estimated in *E. parviflora* [SCoT = 70.23%; IRAP = 73.23%] followed by *E. clavicornis* (67.21 and 68.31%); *E. speciosa* (63.39 and 62.43%); *E. streptopetala* (61.23 and 62.35%) and *E. welwitschii* (57.32 and 58.45%) but comparatively IRAP was found to produce higher PP value in all five *Eulophia* species in comparison to SCoT. However, Nei's gene diversity (h) amongst the five *Eulophia* species estimated by the SCoT and IRAP markers was found to be almost similar [Table 3; (SCoT = 0.227; IRAP = 0.267)]. Likewise, the values of Shannon's information index (I), observed number of alleles (N_a) and effective number of alleles (N_e) of the SCoT and IRAP markers were also identical (Table 3). The values of the total species diversity amongst the population level (H_T), within the population diversity (H_s) and gene flow (N_m) were also almost identical for both the markers (Table 4).

Fig. 1 Banding profile of five *Eulophia* species using SCoT primer **a** SCoT-13 and IRAP primer, **b** LTR-3 [Lane L- Gene Ruler marker; Lanes (1–6)—*E. parviflora*; Lanes (7–12)—*E. streptopetala*; Lanes

(13–18)—*E. welwitschii*; Lanes (19–24)—*E. speciosa*; Lanes (25–30)—*E. clavicornis*]

Comparison of SCoT and IRAP in evaluating genetic diversity of five *Eulophia* species

Both SCoT and IRAP marker systems were compared on the basis of various parameters. Fraction of polymorphic loci (β) and average *PIC* values were found to be 0.79 and 0.86 and 0.40 and 0.52, respectively. The multiplex ratio (*n*) and effective multiplex ratio (EMR) parameters varied between 3.57 and 3.31 for SCoT and IRAP with marker index (MI) values of 1.43 and 1.72 subsequently for SCoT and IRAP (Table 5).

Discussions

Analysis of the extent of polymorphism amongst five *Eulophia* species

Firsthand knowledge of genetic diversity of a plant species provides important molecular insight into the aspects of population dynamics and fitness which makes it an important prerequisite for designing conservation strategies. In case of species which are already categorized as rare, endangered and threatened (RET), information about genetic diversity is of primary importance as loss of genetic diversity reduces the species adaptability, accelerating the

species extinction rate (Swarts and Dixon 2009a, b; Swarts et al. 2009; Muñoz et al. 2010). In general, the genetic characteristics of a species are affected primarily by various ecological and biological traits, modes of reproduction and breeding and also by the various human mediated anthropogenic activities (Hamrick and Godt 1996; Nybom 2004).

Thus, in attempting to study the genetic diversity within orchids, the classical approaches is primarily based upon morphological traits (Wang et al. 2009) which has got major limitations (Sharma et al. 2011). In comparison to these classical approaches, a molecular marker based approach is much more accurate, precise and simple and has efficiently characterized genetic variability both at inter-specific, as well as intra-specific levels (Gepts 2002), and provide important directives for development of species specific markers (Tong-Jian et al. 1991; Nekrutenko et al. 2000). Thus, in order to analyze the genome diversity of RET category plants including orchids, molecular markers such as RAPD, ISSR, AFLP have been used routinely (Sharma et al. 2011; Manners et al. 2013; Bhattacharyya et al. 2013, 2015, 2017; Bhattacharyya and Van Staden 2016). These approaches are pragmatic and useful for designing conservation of plant genetic resources which are facing the risk of habitat fragmentation and extinction. It also helps in cultivar identification and selection of

Fig. 2 **a** UPGMA dendrogram of five *Eulophia* species (6 representatives from each species) constructed from pooled SCoT and IRAP marker data. **b** The highest probability of the pooled SCoT and IRAP marker data was attained when samples were clustered into two and four groups ($K = 2$ and 4) as shown by the peak value of ΔK (Evanno

et al. 2005). Assignment of groups estimated with the STRUCTURE software based on individuals belonging to five species of *Eulophia* sp. using pooled SCoT and IRAP marker data set for **c** $K = 2$ and **d** $K = 4$ (colour figure online)

Table 2 Summary of analysis of molecular variance (AMOVA) based on SCoT and IRAP analysis of *Eulophia* species

Marker	Source of variance	df	Sum of squares	Variance components	Percentage of variation	F _{ST}
SCoT	Amongst the species	2	260.41	8.53	34.32	0.3253
	Within the species	27	694.00	15.23	65.68	
IRAP	Amongst the species	2	191.47	5.87	31.39	0.2921
	Within the species	27	628.01	15.22	68.61	
SCoT +IRAP	Amongst the species	2	245.32	6.62	32.29	0.3155
	Within the species	27	663.23	14.32	66.21	

parents for various hybridization programs (Graner et al. 2004).

Along with the various conventional molecular markers, a new generation of trait-related markers namely SCoT and IRAP have been developed. The selection of trait-specific genetic markers is important and critically depends on the purpose of use (Gupta et al. 2002). These new techniques are unique in itself as the primers have been designed from the conserved region. The marker data reveals greater genetic information than the conventional markers like RAPD and ISSR which targets non-coding regions (Collard and Mackill 2009; Bhattacharyya et al. 2013). In orchids, reports on their genetic diversity both at intra- and inter-

specific levels are quite limited. Reports are particularly fewer in case of terrestrial orchids. In case of *Eulophia* sp. there exists no published report on its genetic diversity both at intra-specific as well as inter-specific level. Keeping the present perspectives into consideration, in the present study we have attempted using two advanced marker systems namely SCoT and IRAP to obtain massive and comprehensive data to analyze genetic variability both at intra-specific as well as inter-specific levels within five species of *Eulophia* collected from South Africa in an authentic and scientific manner. The results indicated that both the SCoT and IRAP markers are equally efficient in the detection of polymorphism, however SCoT proved to have

Table 3 Mean genetic parameters based on SCoT and IRAP analysis of the five species of *Eulophia*

Marker	Species	Sample size	PP	Mean n_a	Mean n_e	Mean h	Mean I
SCoT	<i>E. parviflora</i>	6	70.23	1.702	1.532	0.243	0.365
	<i>E. streptopetala</i>	6	61.23	1.612	1.413	0.223	0.323
	<i>E. welwitschii</i>	6	57.32	1.573	1.378	0.216	0.315
	<i>E. speciosa</i>	6	63.39	1.721	1.523	0.219	0.332
	<i>E. clavicornis</i>	6	67.21	1.672	1.487	0.235	0.346
	Among all genotypes	30	63.87	1.656	1.466	0.227	0.336
IRAP	<i>E. parviflora</i>	6	73.23	1.732	1.533	0.312	0.362
	<i>E. streptopetala</i>	6	62.35	1.623	1.487	0.275	0.325
	<i>E. welwitschii</i>	6	58.45	1.584	1.334	0.299	0.313
	<i>E. speciosa</i>	6	62.43	1.701	1.543	0.223	0.329
	<i>E. clavicornis</i>	6	68.31	1.669	1.478	0.227	0.339
	Among all genotypes	30	64.95	1.661	1.475	0.267	0.333

PP, percentage polymorphism; n_a , observed number of alleles; n_e , effective number of alleles; h , Nei's gene diversity index; and I , Shannon's information index

Table 4 Population genetic structure and estimate of gene flow within the five species of the genus *Eulophia*

Marker	H_T	H_s	G_{ST}	N_m
SCoT	0.4342	0.3412	0.2932	1.323
IRAP	0.4143	0.3286	0.2712	1.299

H_T , total diversity; H_s , diversity within population; G_{ST} , coefficient of gene differentiation and N_m , gene flow based on G_{ST}

Table 5 Comparison of SCoT and IRAP markers in evaluating genetic diversity within five *Eulophia* species

Parameter	SCoT	IRAP
Number of assay units	12	5
Number of polymorphic band	83	37
Number of loci	105	43
Fraction of polymorphic marker (β)	0.79	0.86
Average PIC	0.401	0.52
Multiplex ratio (n)	4.53	3.85
Effective multiplex ratio ($E = \beta \times n$)	3.57	3.31
Marker index ($MI = E \times PIC$)	1.43	1.72

slightly higher detection capacity i.e. 91.6% in comparison to IRAP (90%; Table 1). The mean number of amplified bands were 6.91 and 6.16 respectively in case of SCoT and IRAP respectively. It is evident from the above findings that both marker systems (SCoT and IRAP) are equally potent to differentiate the closely related species within the same genus. In general, marker efficiency depends upon its degree of polymorphism which could be detected amongst the sampled group of plants. The consistently higher

reproducibility rate exhibited by the SCoT and IRAP markers are primarily due to the fact that both of them have longer primers with a higher annealing temperature (Kalender and Schulman 2006; Collard and Mackill 2009).

Analysis of molecular variance (AMOVA) within five *Eulophia* species

Along with other important parameters, analysis of various criteria of genetic diversity reveals the robustness of the present marker approach. Both the SCoT and IRAP markers were found to be equally efficient in decoding the aspects of genetic diversity amongst the five chosen *Eulophia* species. However, a comparative evaluation of the specific benchmark parameters expected heterozygosity, Nei's gene diversity and Shannon's index for these five *Eulophia* species, as well as populations on the whole; SCoT is found to be more efficient than IRAP. The AMOVA results based on the SCoT marker data had revealed about 34% variation amongst the *Eulophia* species while 65% variation was within the species. Similarly, the AMOVA results obtained on the basis of the IRAP marker data accounted for 31% variability amongst the species whereas 68% of genetic variability was compartmentalized at interspecific level. Combined SCoT and IRAP marker data reflected the pattern of genetic variability compartmentalization at intra- and inter-specific levels showing 32% variability amongst the species and 66% variability within the species. Both the marker techniques i.e. SCoT and IRAP suggested a significant differentiation with significant F_{ST} values (0.32 and 0.29) with an overall value of 0.31. The present findings are closely supported by the findings of other co-workers in related plant species including orchids (Bhattacharyya et al. 2013; Bhattacharyya and Kumaria 2015).

Genetic diversity and population structure of five *Eulophia* species

The population genetic structure of the *Eulophia* species was also determined using parameters like percentage polymorphism (PP), observed number of alleles (n_a), effective number of alleles (n_e), Nei's genetic diversity index (h) and Shannon's information index (I). A comparative study amongst the various parameters reveals the efficacy of both markers with SCoT being more precise (Table 3). According to the popular concept of population genetics, gene flow value ($N_m < 1$) signifying less than one migrant per generation into a population. Along with the gene flow index, value of genetic differentiation (G_{ST}) > 0.25 is generally regarded as the threshold quantity, beyond which significant population differentiation occurs (Slatkin 1987). Taken into consideration the above benchmarks, the analyzed *Eulophia* population is just above the threshold level and SCoT was superior to IRAP in determining gene flow and gene differentiation in comparison to IRAP [(SCoT- $N_m = 1.32$; $G_{ST} = 0.29$); (IRAP- $N_m = 1.29$; $G_{ST} = 0.27$)]. Differences amongst populations are commonly estimated by the use of several statistical parameters like Wright's inbreeding coefficient or Fixation index (F_{ST}) and Nei's coefficient of gene differentiation (G_{ST}). Based on the obtained F_{ST} and G_{ST} values for the SCoT and IRAP markers, the former was found significantly better in evaluating population differentiation and structure (Table 4).

With a recorded gene flow value greater than one, the possibility of substantial differentiation due to the effect of genetic drift is significantly nullified (Slatkin and Barton 1989). In outcrossing species, with gene flow (N_m) values greater than 1, might be due to their well evolved and developed pollen dispersal mechanism. Thus, the gene flow estimated by both SCoT and IRAP markers was sufficient to counteract the differentiation due to genetic drift in *Eulophia*. However, the present study also reveals a high degree of gene differentiation which might be due to the occurrence of fragmented populations of small size subjected to active human anthropogenic influences accounting for drastic reduction in the number of wild populations. The above discussed phenomenon leads to population isolation and consequently an extensive and recurrent gene flow which is further justified by the findings of Ellstrand and Elam (1993) who had reported that population isolation may lead to stochastic differentiation by genetic drift.

In order to further analyze the genomic data generated by the SCoT and IRAP marker systems, the relationship between the genome coverage of the SCoT and IRAP marker systems, Mantel test statistics was performed. Correlation between the two marker systems was found to be high ($r = 0.91$) with a determination coefficient value of

0.82 (r^2). The above findings signify that the markers share a minimum of 43% genome for evaluation in common which might be due to the large variability in the genome size arising due to the difference in the chromosome numbers amongst the five species of *Eulophia* (Mehra and Khosla 1973). Furthermore, SCoT and IRAP's target specific regions of the genome and obtains the polymorphic data encoded in these nuclear regions of the genome. Thus, from the present experimental findings, it can be further inferred that the DNA regions amplified by the SCoT primers could have been segregated by IRAP. Also, an identical degree of recorded population differentiation amongst the sampled genotypes of *Eulophia* sp. indicates that the genetic relatedness information revealed by SCoT can be directly correlated with that of IRAP marker data. Thus, a comparison of effective multiplex ratio (EMR) and marker index (MI) of both the SCoT and IRAP markers on the basis of the efficiency parameters like multiplex ratio (MR), effective multiplex ratio (EMR) and marker index (MI) were evaluated which revealed the utility and usefulness of both the markers (Table 5). Our observation is in close agreement with the findings of various workers on different plant species where they have reported a close degree of genetic relatedness between the molecular data (Landry et al. 1994; Shimada et al. 1999; Soorni et al. 2013).

Cluster analysis

Combining different marker data set provides a comprehensive taxonomic view representing distinct levels of taxonomic differentiation, an aspect similar to that of Sneller et al. (1997). Keeping into consideration the advantages of cumulative dataset analysis, the molecular marker data obtained from SCoT and IRAP markers were pooled into a single integrated matrix. Based on the UPGMA method, the sampled genotypes of the *Eulophia* species were clustered into two broad clusters which were further sub-clustered into three more sub-clusters. Bayesian cluster analysis revealed that the five *Eulophia* species collected could be represented by two ($K = 2$) or four ($K = 4$) large genotypic groups. However, subdivision into a larger number of clusters ($K = 4$) revealed that *E. parviflora*, *E. streptopetala* and *E. welwitschii* were grouped into one cluster (shaded red) whereas the remaining representatives from *E. speciosa* and *E. clavicornis* formed a separate cluster (shaded green; Fig. 2c). A systematic analysis of standard deviation of probabilities recorded in different simulations helps to define the number of groups and population structure. In the present study, the organization of the sampled genotypes into four groups ($K = 4$) coincided with one of the lowest recorded standard deviation values of probability (Table S1). Thus the population

genetic structure of the analyzed individuals is better represented by groups. The obtained Bayesian matrix clustering is closely matched with the UPGMA dendrogram pattern which also affirms that the sampled individuals can be clustered into two major groups. The above experimental finding is in close congruence with the results obtained by AMOVA which reveal that a greater degree of variation exists amongst the population and a lesser degree of variation is compartmentalized between the populations which is supported by the findings of Li and Ge (2006).

Considerations for conservation

Maintenance of the evolutionary potential is the principal objective of natural conservation by maintaining higher levels of genetic diversity. Thus, a firsthand knowledge of the genetic variability both at intra-specific as well as inter-specific levels helps to formulate effective conservation strategies (Milligan et al. 1994). Conservation and management of wild species are always influenced by both selective (gene flow and genetic drift) and non-selective (natural selection) factors in population subdivision. Thus, a reduction in the size of natural wild populations will lead to an increased influence of various selective forces which will ultimately lead to the disintegration of the natural population structure ultimately resulting in a cumulative negative effect in the population dynamics of the species by putting in contact with two differentiated gene pools through outbreeding depression or hybridization between locally adapted populations (Holsinger et al. 1999; Lande 1999). As genetic variations are fundamentally involved in the survival and evolution of species, and even more critical for RET category plants like orchids, the results of this study have direct implications for conservation and management of *Eulophia*. Our results reveal that the genetic diversity of *E. welwitschii* is the least amongst five assessed *Eulophine* species of orchids. This may be due to habitat destruction, over collection and exploitation of *E. welwitschii*. Being a terrestrial orchid genus, the entire *Eulophine* orchid genus subjected to lower seed vigour, internal seed dormancy and above all an extremely lower rate of germination in nature (i.e. less than 3%) as fungal association is needed. Like most of the orchid species, *Eulophia* has got strong habitat preference and pollinator dependence. Thus, habitat protection will ensure species co-existence with other organisms like fungi and pollinators on which the orchid life cycle is influenced and dependant (Li and Ge 2006). The gemplasms sampled from diverse ecological niches could be maintained in seed banks for future conservation endeavors (Maunder et al. 2000) which will in turn help to improve the conservation status of this endangered orchid taxa using in situ and ex situ conservation protocols.

Conclusions

The five sampled *Eulophia* species depicted a high level of genetic diversity with *E. parviflora* and *E. clavicornis* found to possess a higher level of genetic diversity in comparison to the other three species. Being reliable and authentic markers, we suggest the use of the above protocol for the assessment of genetic variability and DNA fingerprinting of other related plant species as the protocol is low cost, reproducible and does not require any prior sequence information for designing the primers. Furthermore, the SCoT and IRAP can be further utilized in the development of the Sequence Characterized Amplified Region (SCAR) markers which will in turn assist in the molecular validation and DNA barcoding of this endangered, medicinally important orchid taxa.

Acknowledgements Paromik Bhattacharyya thanks the University of KwaZulu-Natal (UKZN), South Africa for support in the form of a Postdoctoral Fellowship.

References

- Bhattacharyya P, Kumaria S (2015) Molecular characterization of *Dendrobium nobile* Lindl., an endangered medicinal orchid, based on randomly amplified polymorphic DNA. *Plant Syst Evol* 301:201–210
- Bhattacharyya P, Van Staden J (2016) *Ansellia africana* (Leopard orchid): a medicinal orchid species with untapped reserves of important biomolecules—a mini review. *S Afr J Bot* 106:181–185
- Bhattacharyya P, Kumaria S, Kumar S, Tandon P (2013) Start Codon Targeted (SCoT) marker reveals genetic diversity of *Dendrobium nobile* Lindl., an endangered medicinal orchid species. *Gene* 529:21–26
- Bhattacharyya P, Kumaria S, Tandon P (2015) Applicability of ISSR and DAMD markers for phyto-molecular characterization and association with some important biochemical traits of *Dendrobium nobile*, an endangered medicinal orchid. *Phytochemistry* 117:306–316
- Bhattacharyya P, Kumaria S, Tandon P (2016) High frequency regeneration protocol for *Dendrobium nobile*: a model tissue culture approach for propagation of medicinally important orchid species. *S Afr J Bot* 104:232–243
- Bhattacharyya P, Ghosh S, Sen Mandi S, Kumaria S, Tandon P (2017) Genetic variability and association of AFLP markers with some important biochemical traits in *Dendrobium thyrsiflorum*, a threatened medicinal orchid. *S Afr J Bot* 109:214–222
- Cabo S, Ferreira L, Carvalho A, Martins-Lopes P, Martín A, Lima-Brito JE (2014) Potential of Start Codon Targeted (SCoT) markers for DNA fingerprinting of newly synthesized tritordeums and their respective parents. *J Appl Genet* 55:307–312
- Chen L, Zhao L, Bai Y, Hu R, Si J (2009) Genetic relationship analysis of different provenances of *Leonurus japonicus* by ISSR marker. *China J Chin Mater Med* 34:1343–1345
- Chinsamy M, Finnie JF, Van Staden J (2011) The ethnobotany of South African medicinal orchids. *S Afr J Bot* 77:2–9
- Chowdhery HJ (2001) Orchid diversity in North-east India. *J Orchid Soc India* 15:1–17

- Collard BCY, Mackill DJ (2009) Start codon targeted (SCoT) polymorphism: a simple, novel DNA marker technique for generating gene-targeted markers in plants. *Plant Mol Biol Rep* 27:86–93
- Cunningham AB (1988) An investigation of the herbal medicine trade in Natal/KwaZulu. Institute of Natural Resources, University of Natal, Pietermaritzburg
- Ellstrand NC, Elam DR (1993) Population genetic consequences of small population size: implications for plant conservation. *Annu Rev Ecol Syst* 24:217–242
- Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Mol Ecol* 14:2611–2620
- Excoffier L, Laval G, Schneider S (2005) Arlequin (version 3.0): an integrated software package for population genetics data analysis. *Evol Bioinform Online* 1:47
- Falush D, Stephens M, Pritchard JK (2003) Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics* 164:1567–1587
- Falush D, Stephens M, Pritchard JK (2007) Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Mol Ecol Notes* 7:574–578
- Feng S, He R, Yang S, Chen Z, Jiang M, Lu J, Wang H (2015) Start codon targeted (SCoT) and target region amplification polymorphism (TRAP) for evaluating the genetic relationship of *Dendrobium species*. *Gene* 567:182–188
- Gepts P (2002) A comparison between crop domestication, classical plant breeding, and genetic engineering. *Crop Sci* 42:1780–1790
- Gerstner J (1941) A preliminary check list of Zulu names of plants: with short notes. *Bantu Stud* 15:277–301
- Graner A, Dehmer KJ, Thiel T, Börner A (2004) IV. Plant genetic resources: benefits and implications of using molecular markers. *Evol Role Genebanks Fast-developing F Mol Genet Genet Resour No* 11, 26 Aug 2004
- Gupta PK, Varshney RK, Prasad M (2002) Molecular markers: principles and methodology. In: *Molecular techniques in crop improvement*. Springer, pp 9–54
- Hamrick JL, Godt MJW (1996) Effects of life history traits on genetic diversity in plant species. *Philos Trans R Soc Lond Ser B Biol Sci* 351:1291–1298
- Holsinger KE, Mason-Gamer RJ, Whitton J (1999) Genes, denies, and plant conservation. In: Landweber LF, Dobson AP (eds) *Genetics and the extinction of species*. Princeton University Press, Oxford, pp 23–46
- Hulme M (1954) *Wild flowers of Natal*. Shuter & Shooter, Pietermaritzburg
- Hutchings A (1996) *Zulu medicinal plants: an inventory*. University of Natal Press, Pietermaritzburg
- Jin WT, Yao SP (2006) *Cultivation and appreciation of noble Spring Orchid cultivars*. Guangdong Science and Technology Press, Guangzhou, China
- Kalendar R, Schulman AH (2006) IRAP and REMAP for retrotransposon-based genotyping and fingerprinting. *Nat Protoc* 1:2478–2484
- Kalendar R, Grob T, Regina M, Suoniemi A, Schulman A (1999) IRAP and REMAP: two new retrotransposon-based DNA fingerprinting techniques. *Theor Appl Genet* 98:704–711
- Lande R (1999) Extinction risks from anthropogenic, ecological and genetic factors. In: Landweber LF, Dobson AP (eds) *Genetics and the extinction of species*. Princeton University Press, pp 1–22
- Landry BS, Li RQ, Cheung WY, Granger RL (1994) Phylogeny analysis of 25 apple rootstocks using RAPD markers and tactical gene tagging. *Theor Appl Genet* 89:847–852
- Li A, Ge S (2006) Genetic variation and conservation of *Changnienia amoena*, an endangered orchid endemic to China. *Plant Syst Evol* 258:251–260
- Manners V, Kumaria S, Tandon P (2013) SPAR methods revealed high genetic diversity within populations and high gene flow of *Vanda coerulea* Griff ex Lindl (Blue Vanda), an endangered orchid species. *Gene* 519:91–97
- Martos F, Johnson SD, Peter CI, Bytebier B (2014) A molecular phylogeny reveals paraphyly of the large genus *Eulophia* (Orchidaceae): a case for the reinstatement of *Orthochilus*. *Taxon* 63:9–23
- Maunder M, Culham A, Alden B, Zizka G, Orliac C, Lobin W, Bordeu A, Ramirez JM, Glissmann-Gough S (2000) Conservation of the Toromiro tree: case study in the management of a plant extinct in the wild. *Conserv Biol* 14:1341–1350
- Mehra PN, Khosla PK (1973) Cytological studies of Himalayan Rutaceae. *Silvae Genet* 22:182–188
- Miller MP (1998) AMOVA-PREP 1.01: a program for the preparation of AMOVA input files from dominant-markers raw data. *Comput Softw Distrib* 22:927–934
- Milligan BG, Leebens-Mack J, Strand AE (1994) Conservation genetics: beyond the maintenance of marker diversity. *Mol Ecol* 3:423–435
- Mulpuri S, Muddanuru T, Francis G, Sujatha M, Tarakeswari M, Francis G (2013) Start codon targeted (SCoT) polymorphism in toxic and non-toxic accessions of *Jatropha curcas* L and development of a codominant SCAR marker. *Plant Sci* 207:117–127
- Muñoz M, Warner J, Albertazzi FJ (2010) Genetic diversity analysis of the endangered slipper orchid *Phragmipedium longifolium* in Costa Rica. *Plant Syst Evol* 290:217–223
- Nekrutenko A, Makova KD, Baker RJ (2000) Isolation of binary species-specific PCR-based markers and their value for diagnostic applications. *Gene* 249:47–51
- Nybom H (2004) Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Mol Ecol* 13:1143–1155
- Powell W, Morgante M, Andre C, Hanafey M, Vogel J, Tingey S, Rafalski A (1996) The comparison of RFLP, RAPD, AFLP and SSR (microsatellite) markers for germplasm analysis. *Mol Breed* 2:225–238
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* 155:945–959
- Rohlf JF (1998) NTSYS pc, numerical taxonomy and multivariate analysis system version 2.0 user guide. Setauket
- Sharma SK, Kumaria S, Tandon P, Rao SR (2011) Single primer amplification reaction (SPAR) reveals inter-and intra-specific natural genetic variation in five species of *Cymbidium* (Orchidaceae). *Gene* 483:54–62
- Shimada T, Hayama H, Haji T, Yamaguchi M, Yoshida M (1999) Genetic diversity of plums characterized by random amplified polymorphic DNA (RAPD) analysis. *Euphytica* 109:143–147
- Slatkin M (1987) Gene flow and the geographic structure of natural. *Science* 236:787–792
- Slatkin M, Barton NH (1989) A comparison of three indirect methods for estimating average levels of gene flow. *Evolution (NY)* 43:1349–1368
- Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegler J (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPs and pedigree. *Theor Appl Genet* 95:163–173
- Sneller CH, Miles JW, Hoyt JM (1997) Agronomic performance of soybean plant introductions and their genetic similarity to elite lines. *Crop Sci* 37:1595–1600

- Soorni A, Nazeri V, Fattahi R, Khadivi-Khub A (2013) DNA fingerprinting of *Leonurus cardiaca* L. germplasm in Iran using amplified fragment length polymorphism and inter-retrotransposon amplified polymorphism. *Biochem Syst Ecol* 50:438–447
- Swarts ND, Dixon KW (2009a) Terrestrial orchid conservation in the age of extinction. *Ann Bot* 104:543–556
- Swarts ND, Dixon KW (2009b) Perspectives on orchid conservation in botanic gardens. *Trends Plant Sci* 14:590–598
- Swarts ND, Sinclair EA, Krauss SL, Dixon KW (2009) Genetic diversity in fragmented populations of the critically endangered spider orchid *Caladenia huegelii*: implications for conservation. *Conserv Genet* 10:1199–1208
- Tong-Jian S, Liu-Qin Z, Xin S (1991) A marker-coupled method for site-directed mutagenesis. *Gene* 103:73–77
- Wang HZ, Wu ZX, Lu JJ, Shi NN, Zhao Y, Zhang ZT, Liu JJ, Zhang ZT, Liu JJ (2009) Molecular diversity and relationships among *Cymbidium goeringii* cultivars based on inter-simple sequence repeat (ISSR) markers. *Genetica* 136:391–399
- Waugh R, McLean K, Flavell AJ, Pearce SR, Kumar A, Thomas BBT, Powell W (1997) Genetic distribution of Bare-1-like retrotransposable elements in the barley genome revealed by sequence-specific amplification polymorphisms (S-SAP). *Mol Gen Genet* 253:687–694
- WCSP (2016) World Checklist of Selected Plant Families. Facilitated by the Royal Botanic Gardens, Kew. <http://apps.kew.org/wcsp/>. Accessed 20 June 2016
- Xiong F, Zhong R, Han Z, Jiang J, He L, Zhuang W, Tang R (2011) Start codon targeted polymorphism for evaluation of functional genetic variation and relationships in cultivated peanut (*Arachis hypogaea* L.) genotypes. *Mol Biol Rep* 38:3487–3494
- Yeh FC, Yang R-C, Boyle T (1999) PopGene version 1.31: Microsoft window-based freeware for population genetic analysis. University of Alberta, Edmonton
- Yu Q, Sheng XX, Sheng YF et al (2009) AFLP Analysis of genetic diversity of *Leonurus japonicus* germplasm resources. *Chin Tradit Herb Drugs* 40:1296–1299