International Journal of Engineering, Science and Mathematics
Vol. 7 Issue 1, January 2018,
ISSN: 2320-0294 Impact Factor: 6.765
Journal Homepage: **http://www.ijesm.co.in, Email: ijesmj@gmail.com**
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gage as well as in Cabell's Directories of Publishing Opportunities, U.S.A

Identification, functional annotation and analysis of COS markers in *Zingiber* and its utility in DNA barcoding

Lesly Augustine*
Smini Varghese**
Geethu Elizabath Thomas***
George Thomas****

Abstract

Species identification is exceedingly difficult in Zingiberaceae due to similarities in morphological characteristics between species, the plants' phenotypic plasticity, and their short and seasonal flowering cycles. DNA barcoding has great potential in Zingiberaceae, both for the confirmation of raw materials in the pharmaceutical sector and as a tool in conservation biology and ecology. A suitable barcode target in the family Zingiberaceae is yet to be discovered. Conserved ortholog set (COS) markers are single-copy evolutionary conserved genes in two or more species that share common ancestry (orthologous). Due to the conserved nature of these markers across genera they may be possible barcode candidates and these conserved set may be used as low-variation universal loci across the monocots. In the present study identification, functional annotation and analysis of COS markers in Zingiberaceae is carried out. About 37 COS markers were identified, validated and tested.The functional annotation revealed house-keeping genes (85.7%), a defense gene, 3 mitochondrial proteins and one chloroplast protein. The utilization of a COS marker (ZE372342) as barcode target for *Zingiber* genera were tested and the results provide a promising target.

Keywords:

COS;
DNA barcoding;
Zingiberaceae;
EST;
Single-copy gene.

Author correspondence:

Geethu Elizabath Thomas
Assistant Professor, Department of Botany,
St. Thomas' College, Thrissur-680001, Kerala, India

1Introduction

Zingiberaceae (gingers) is a pantropic family consisting of approximately 53 genera and 1200 species [1], [2]. Most of the gingers are rhizomatous, and the rhizomes (underground stem) of several gingers are used as spices, vegetables, neutraceuticals, drugs or indispensable ingredients in traditional medicines in Southeast Asian countries [3], [4]. Species identification is exceedingly difficult in Zingiberaceae due to similarities in morphological characteristics between species, the plants' phenotypic plasticity, and their short and seasonal flowering cycles. DNA barcoding has great potential in Zingiberaceae, both for the confirmation of raw materials in the pharmaceutical sector [5], [6], [7] and as a tool in conservation biology

*Assistant Professor, Department of Botany, Sacred Heart College-682013, Thevara, Kerala, India
**Research Student, Plant Molecular Biology, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram-695014, Kerala, India
***Assistant Professor, Department of Botany, St. Thomas' College, Thrissur-680001, Kerala
****Scientist F, Plant Molecular Biology, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram-695014, Kerala, India

and ecology. Despite considerable debate over the last 8 years, there is little consensus regarding the choice of loci for barcoding plants. Each plant barcode locus has different strengths and weaknesses, and their resolvability and universality vary considerably between taxa [8], [9]. Shi et al. [4] evaluated the universality and resolvability of five chloroplast loci and nrDNA internal transcribed spacer 2 (ITS2) in Zingiberaceae members in China and found ITS2 be the most promising locus. Contrastingly study by Vinitha et al.[10] suggests that *matK* and *rbcL* loci are suitable for barcoding Zingiberaceae members and they highlight the poor utility of ITS and its intragenomic heterogeneity in the species tested. Another study points out that *Curucuma,* a Zingiberaceae member is a challenging group for DNA barcoding wherein the four barcode regions they used (*matK, rbcL, trnH-psbA* and *trnL-F*) yielded no barcode gaps [11]. Thus this highlights the lack of a suitable barcode target in the family Zingiberaceae.

Conserved ortholog set (COS) markers [12] are single-copy evolutionary conserved genes in two or more species that share common ancestry (orthologous). Due to the conserved nature of these markers across genera they may be possible barcode candidates and these conserved set can be used as low-variation universal loci across the plant kingdom [13]. COS markers have been identified in several flowering plant (angiosperms) families, including the Euasterids-Solanaceae, Rubiaceae, and Asteraceae [14], the legume family (Leguminosae) and grass family, Poaceae [15], the composite family, Asteraceae [16], Rosaceae [17], gymnosperms, Pinaceae [18] and forest trees [19]. A comparative genomic approach is not yet attempted in Zingiberaceae family. In the present study identification, functional annotation and analysis of conserved orthologous sequences (COS) in genus *Zingiber* is carried out. Later on the possibility of a COS marker as putative barcode candidate is tested across *Zingiber* genus.

2Materials and methods

2.1 Ginger EST Database

Zingiber officinale EST collection was downloaded from NCBI (National Center for Biological Information) (http://www.ncbi.nlm.nih.gov/dbEST/). The EST collection include 38139 ESTs which were classified into three tissue libraries of leaves 13274 (DV544275-ES560515), rhizomes 12763 (DY350707-DY363469) and roots 12092 (DY363470-DY375561) [20].

2.2 Computational Screening of Conserved Ortholog Set Markers

Conserved orthologs between rice (*Oryza sativa*) and ginger were identified following the method of Fulton et al.[12] with modifications. Rice genome was downloaded from NCBI (http://www.ncbi.nlm.nih.gov/genome). A conservative computational strategy was followed to avoid paralogs. Ginger ESTs were scanned against Rice genome using Spidey, a computer program that aligns spliced sequences to genomic sequences, using local alignment algorithms and heuristics to put together a global spliced alignment (http://www.ncbi.nlm.nih.gov/spidey) [21]. Three major criteria were implied for selecting a conserved ortholog, (i) EST should provide hit against rice genome at an expect value $< E^{-20}$ (ii) Next best rice genome hit must be of lower significance, i. e. expect value should be less than E^{-1} (iii) Match should cover 100% coverage of EST. Those ESTs which obey the above mentioned criteria were validated against the ginger unigene set composed of 38139 contigs to ensure that all COS markers chosen represent unique ginger genes. They were blasted against rice genome also to validate its uniqueness. The selected COS regions were annotated by BLASTX analysis against the GenBank protein database maintained at the NCBI (http://www.ncbi.nlm.nih.gov/).

2.3 Primer designing

After rigorous screening, about 37 regions were selected as putative COS markers. These selected regions were blasted against the available monocot sequences and those which provide hits with an expected value $<E^{-20}$was identified and selected for primer designing. Thus the primers designed were supposed to be universal primers for orthologous genes across monocots. Primers were designed using Primer 3 software [22] and primers were custom synthesized by Metabion, Germany. A total of 51 primer pairs were synthesized and their annealing temperature ranged from 50-66°C.

2.4 Plant Materials

Preliminary screening of the primers was done in one accession each of *Z. officinale* and *Z. zerumbet*. Two accessions each of seven species of *Zingiber* were used for the barcoding studies (Table 1). GenElute Plant Genomic Kit (Sigma) was used to isolate total genomic DNA from the selected samples.

Table 1 List of plant materials used for the study

Species	Accession no.	Source
Z.wightianum	*Z.wightianum_2*	Anamalai, Kerala
	Z.wightianum_1	Anamalai, Kerala
Z. zerumbet	*Z. zerumbet_1*	Poovathode, Kerala
	*Z. zerumbet_2**	Poovathode, Kerala

Z.parishi	*Z.parishi_1*	Andaman
	Z.parishi_2	Andaman
Z.officinale	*Z.officinale_1**	Valparai, TamilNadu
	Z.officinale_2	Valparai, TamilNadu
Z.oderiferum	*Z.oderiferum_1*	Andaman
	Z.oderiferum_2	Andaman
Z.nimonii	*Z.nimonii_1*	Poovathode, Kerala
	Z.nimonii_2	Palakkad, Kerala
Z.neesanum	*Z.neesanum_2*	Ponmudi, Kerala
	Z.neesanum_1	Idukki,Kerala

* Plants used for initial primer screening

2.5 PCR amplification

PCR was performed on a GeneAmp PCR System 9700 (Applied Biosystems) in a 10-µL volume containing 10 ng total DNA, 2.5 mM $MgCl_2$, 0.2 mM of each dNTP, 1picomole each of forward and reverse primers, 0.1mg/ml BSA, 4% DMSO, 0.5 U AmpliTaq Gold polymerase (Applied Biosystems) and 1X AmpliTaq Gold PCR buffer. The PCR profile included an initial denaturation at 95°C for 5 min followed by 40 cycles of denaturation at 95°C for 30 s, primer annealing at 50 – 66°C for 40 s and DNA strand extension at 72°C for 1 min with a final extension of 7 min at 72°C.The amplified products were visualised in agarose gel electrophoresis. Only the primers which provided single intense band were selected for further sequencing. In the present study only one region (ZE372342) was used for barcoding*Zingiber* genus.

2.6 ExoSAP IT clean-up and DNA Sequencing

The amplified PCR product after visualisation in agarose gel electrophoresis, were cleaned up using ExoSAP-IT PCR product clean up reagent, Thermo Fisher Scientific. About 1 µl of ExoSAP-IT is added to 5 µl PCR product and kept for 45 minutes incubation at 37°C. After treatment, ExoSAP-IT is inactivated by incubating at 80°C for 15 minutes. The cleaned up amplicon were sequenced using ABI® prism Big Dye Terminator v3.1 Cycle Sequencing Kit (USA). The PCR profile consisted of an initial denaturation at 96^0C for 2 min followed by 25 cycles at 96^0C for 30 sec, 50^0C for 40 sec and 60^0C for 4 min. in a DNA Thermal cycler (ABI). After post sequencing clean up capillary electrophoresis were carried in ABI 3730 DNA Analyzer.

2.7 Sequence analysis

The sequences obtained from the 14 samples were aligned using the program Geneious Pro 5.0.4 [23] and refined manually as needed. Phylogenetic tree was constructed using distance-based methods, unweighted pair group method with arithmetic mean, UPGMA. In the tree-based approach, a species is considered resolved if it is recovered monophyletically, *i.e.* when all accessions of the species are clustered under one node. Bootstrap analysis (500 replicates) was performed to assess the robustness of clustering in the trees.

3Results andDiscussion

3.1 Selection of COS markers

The 37 COS markers described here were identified by screening *Zingiber* EST sequences against rice genome (http://www.ncbi.nlm.nih.gov/genome) using the software Spidey. The purpose of this screen was to identify single-copy tomato genes that have a single best match to one region of the rice genome and hence would qualify as potential orthologs.

To obtain the COS markers described here, 38139 *Zingiber* ESTs were screened as described above. To standardize all results, the COS marker set was rescreened against both rice genome and the *Zingiber* EST/unigene set. Only those regions which fulfilled the earlier mentioned three criteria were selected as putative COS markers which were only 37. The number of COS markers were less compared to the earlier studies [12],[18], [19] which may be due to our stringent modifications to obtain single copy genes with 100% mRNA coverage.

3.2 Annotation of COS markers

Out of 37 selected COS markers 35 were annotated by BLASTX analysis against the GenBank protein database maintained at the NCBI (http://www.ncbi.nlm.nih.gov/). Annotation for the remaining two was not obtained. The identified putative COS genes belonged to diverse functional groups (Table 2). Out of 35 annotated regions, majority were house-keeping genes such as chlorophyll a/b binding protein (ZE377718), transporters (ABC transporters-ZE350073; vesicular transport SNAPs-ZE351074), kinases, proteases, ubiquitins, enzymes involved in metabolism of purines (phosphoribosylaminoimidazolecarboxamide

formyltransferase-ZE352864), chlorophyll (glutamate-1-semialdehyde 2,1-aminomutase-ZE35589; 3,8-divinyl protochlorophyllide a 8-vinyl reductase-ZE350566; Mg-protoporphyrin IX monomethyl ester cyclase-ZE380805), sterols (sterol methyl transferase2-ZE351047) and lipopolysaccharides (bifunctional polymyxin resistance arnA protein-ZE361721), transcription factors, translation factors etc. One gene region identified coded for disease resistance, complement proteins (ZE356530). Mitochondrial proteins were also identified such as carbamoylphosphate synthase large subunit-like gene (ZE355175), formate dehydrogenase 1 (ZE355804) and succinate dehydrogenase flavoprotein subunit (ZE372342). Oxygen evolving enhancing protein of photosystem II (ZE381543) was the chloroplast protein identified. The annotation results are in concordance with previous studies in the presence of majorly house-keeping genes and less number of evolving genes such as disease resistance genes which may be due to their conserved nature across genera and families [18], [19].

Table 2 Functional annotation for the putative conserved ortholog set identified for *Zingiber*

Sl. No.	**COS markers**	**Sequence annotation**
1	ZE381543	oxygen evolving enhancing protein of photosystem II for *Zea mays*
2	ZE355175	carbamoylphosphate synthase large subunit-like gene for *Triticum monococcum*
3	ZE356530	MAC/Perforin domain containing protein for *Z. mays*
4	ZE361721	bifunctional polymyxin resistance arnA protein for *Z. mays*
5	ZE350073	ATP BINDING CASSETTE PROTEIN 1 for *Arabidopsis thaliana*
6	ZE380805	Mg-protoporphyrin IX monomethyl ester cyclase for *Hordeum vulgare*
7	ZE355842	Protein Kinases *A. thaliana*
8	ZE353441	nucleic acid binding protein for *A. lyrata*
9	ZE349334	chloroplast translational elongation factor Tu (tufA)
10	ZE376518	zinc metalloprotease *Ricinus communis*
11	ZE350166	N-acetylglucosaminyltransferase *Musa acuminata*
12	ZE378062	ATP-dependent peptidase/ ATPase/ metallopeptidase (ftsh9) *A. thaliana*
13	ZE351047	sterol methyl transferase2 (smt2) for *Z. mays*
14	ZE362589	serine hydroxymethyltransferase *Z. mays*
15	ZE366271	lysyl-tRNA synthetase for *Z. mays*
16	ZE350566	3,8-divinyl protochlorophyllide a 8-vinyl reductase *A. thaliana*
17	ZE345717	fructose-1,6-bisphosphatase
18	ZE367620	myb2 for *Z. mays*
19	ZE354692	alpha-amylase for *H. vulgare*
20	ZE382569	AMP-dependent synthetase and ligase family protein for *A. thaliana*
21	ZE353562	heat shock protein 70 (HSP70)-interacting protein *R. communis*
22	ZE345478	eukaryotic translation initiation factor 3 for *Z. mays*
23	ZE355804	formate dehydrogenase 1 for *Z. mays*
24	ZE371351	gcip-interacting family protein-like mRNA
25	ZE377718	chlorophyll a/b binding protein cab-PhE7 for *P. edulis*
26	ZE372342	succinate dehydrogenase flavoprotein subunit for *Z. mays*
27	ZE352864	phosphoribosylaminoimidazolecarboxamide formyltransferase *A. thaliana*
28	ZE351074	alpha-soluble NSF attachment protein for *Elaeis guineensis*
29	ZE355895	glutamate-1-semialdehyde 2,1-aminomutase for *Z. mays*
30	ZE381127	formyltetrahydrofolate synthetase for *Populus*
31	ZE361874	putative cell elongation protein for *E. guineensis*
32	ZE351405	malate dehydrogenase
33	ZE355959	Eukaryotic translation initiation factor 3 subunit for *Ricinus, Arabidopsis*
34	ZE353425	S-adenosyl-L-homocysteine hydrolase (SH6.2) for *T. aestivum*
35	ZE364870	E3 ubiquitin protein ligase upl2 for *R. communis*

3.3 Primer designing and PCR amplification

Primers were designed for the selected 37 regions and their PCR amplification was tested. Altogether 51 primer pairs were designed, custom synthesized and validated. Out of the 51 primer pairs of 37 regions only 17 primer pair belonging to 14 regions yielded single band. All the others yielded multiple non-specific bands and hence were removed from further studies. Out of these 17 primer pair only 11 successfully completed sequencing belonging to 8 regions. Eventually only 8 COS markers were commended for further use belonging to 8 regions (Table 3). These 8 targets can be used across monocots as rate of transferability across species–genera is greater and their probability of orthology is high [18].

In the present study the potential of one COS marker (ZE372342) as barcode target for *Zingiber* species was validated. Two accessions each of seven *Zingiber* species were used for barcoding.

Table 3 List of primer combinations successfully sequenced and validated in *Zingiber* genus

Sl. No.	OligoName	5'<-----Sequence----->3'	Amplicon size	Annealing temp.
1	ZE356530F	TGGGCCCAAAGCTCTTCATTAGCAC	1kb	55°C
	ZE356530R	GTGCACCCAGGAACCCTGGARAA		
2	ZE376518F	AAGCTGTGGAGGCATCCATGAACAC	500bp	55°C
	ZE376518R	ACGTGCCAACTCTGCAAGATCYTCT		
3	ZE367620F	AACCTGCACGTCCTGATCCTGTTGA	900bp	58°C
	ZE367620R	AATTCTTGGAGTAGCACCAGGSCCA		
4	ZE345478F	GGCTTCGCTGGATCTTTCACAAGCA	1kb	55°C
	ZE345478R	CCCGCAACTTCTTCAAGGTCACCAC		
5	ZE372342F*	CCACTGTGCACTATAATATGGGTGGAA	500bp	60°C
	ZE372342R	ACACGTTGCATGTTGAGACGTATTT		
6	ZE381127F	TTAAAATGCATGGAGGGGGCCCTGA	1kb	55°C
	ZE381127R	CAACGCCACTAGCACCATAGAACTTRGC		
7	ZE355959F	AGATCATTGCCTTTGCATGGGAGCC	700bp	55°C
	ZE355959R	CCTGGTCCTCTTGTTCGTACTTCTTRC		
8	ZE353425F	CTCCCTCCACATGACCATCCAGACC	500bp	55°C
	ZE353425R	TCTTGACACCGGTGGTGGTCTCCTC		

* Primer pair used for DNA barcoding studies

3.4 Sequencing and analysis

For all the 14 samples studied bidirectional sequencing was performed. The sequences were multiple aligned and sequence characteristics were studied. Out of the 467 sites, 439 were conserved sites (94%) and 28 were variable sites (6%). The variable sites included 24 parsimony informative sites (5.13%) and 4 singleton sites (0.87%).

Phylogenetic tree was constructed using distance-based methods and the resultant UPGMA dendrogram resolved all the seven *Zingiber* species studied proving its potential in DNA barcoding (Fig. 1). All the nodes showed a high bootstrap value (> 50) highlighting the robustness of the dendrogram. Sequence databases are comparatively poor in the case of Zingiberaceae compared to other angiospermic plant families with high economic importance. A comparative genomic approach is preferable because of the lack of genetic resources for primer designing. The success of COS markers as a tool for phylogenetic analyses have been promising in crop species [14], [16].

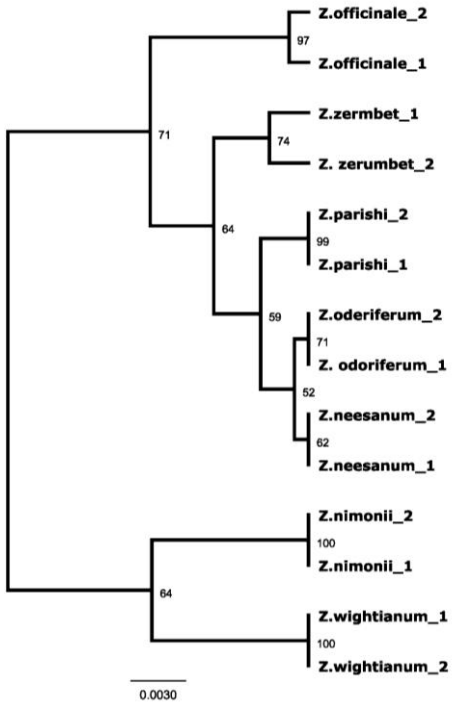

Figure 1 UPGMA dendrogram of 14 individuals belonging to sevenspecies of *Zingiber* obtained using distance based methods computed from the data obtained from sequencing

4Conclusions

In the present study we have identified COS markers in *Zingiber* which can be utilized across the monocots. These conserved orthologous markers were annotated and majority of the genes were house keeping. All the 37 regions were used for primer designing and all of them were tested in one accession each of *Z. officinale* and *Z. zerumbet*. The utility of one of the COS markers in DNA barcoding was analysed and it proved to be a putative barcode target in *Zingiber* genera. As plant barcoding is still not a complete success, identification of new putative targets and their utility across species are essential. Our study paves a new path in exploiting the utility of comparative genomic approach in less sequenced non-model plant genera.

References

[1] Kress, W. J., Prince, L. M., & Williams, K. J. (2002). The phylogeny and a new classification of the gingers (Zingiberaceae): evidence from molecular data. *American Journal of Botany*, *89*(10), 1682-1696.

[2] Sabu, M. (2003). Revision of the genus *Zingiber* in South India. *Folia Malaysiana*, *4*(1), 25-52.

[3] Tushar, Basak, S., Sarma, G. C., & Rangan, L. (2010). Ethnomedical uses of Zingiberaceous plants of Northeast India. *Journal of Ethnopharmacology*, *132*(1), 286-296. doi: https://doi.org/10.1016/j.jep.2010.08.032

[4] Shi, L.C., Zhang, J., Han, J. P., Song, J. Y., Yao, H., Zhu, Y. J., Li, J. C., Wang, Z. Z., Xiao, W. & Lin, Y. L. (2011). Testing the potential of proposed DNA barcodes for species identification of Zingiberaceae. *J Syst Evol*, *49*, 261-266.

[5] Srirama, R., Senthilkumar, U., Sreejayan, N., Ravikanth, G., Gurumurthy, B. R., Shivanna, M. B., Sanjappa, M., Ganeshaiah, K. N. & Uma Shaanker, R. (2010). Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepato-protective plant using molecular tools. *Journal of Ethnopharmacology*, *130*(2), 208-215. doi: https://doi.org/10.1016/j.jep.2010.04.042

[6] Xue, C.Y. & Li, D.Z. (2011). Use of DNA barcode sensu lato to identify traditional Tibetan medicinal plant *Gentianopsis paludosa* (Gentianaceae). *Journal of Systematics and Evolution*, *49*(3), 267-270. doi: 10.1111/j.1759-6831.2011.00127.x

[7] Li, M., Cao, H., But, P.P.H. & Shaw, P.C. (2011). Identification of herbal medicinal materials using DNA barcodes. *Journal of Systematics and Evolution*, *49*(3), 271-283. doi: 10.1111/j.1759-6831.2011.00132.x

[8] CBOL Plant Working Group. (2009). A DNA barcode for land plants. *Proceedings of the National Academy of Sciences*, *106*(31), 12794-12797. doi: 10.1073/pnas.0905845106

[9] Hollingsworth, M. L., Andra Clark, A., Forrest, L. L., Richardson, J., Pennington, R. T., Long, D. G., Cowan, R., Chase, M. W., Gaudeul, M. & Hollingsworth, P. M. (2009). Selecting barcoding loci for plants: evaluation of seven candidate loci with species-level sampling in three divergent groups of land plants. *Molecular ecology resources*, *9*(2), 439-457. doi: 10.1111/j.1755-0998.2008.02439.x

[10] Vinitha, M. R., Kumar, U. S., Aishwarya, K., Sabu, M., & Thomas, G. (2014). Prospects for discriminating Zingiberaceae species in India using DNA barcodes. *Journal of Integrative Plant Biology*, *56*(8), 760-773. doi: 10.1111/jipb.12189

[11] Chen, J., Zhao, J., Erickson, D. L., Xia, N., & Kress, W. J. (2015). Testing DNA barcodes in closely related species of *Curcuma* (Zingiberaceae) from Myanmar and China. *Molecular ecology resources*, *15*(2), 337-348. doi: 10.1111/1755-0998.12319

[12] Fulton, T. M., Van der Hoeven, R., Eannetta, N. T., & Tanksley, S. D. (2002). Identification, Analysis, and Utilization of Conserved Ortholog Set markers for comparative genomics in higher plants. *The Plant Cell*, *14*(7), 1457-1467. doi: 10.1105/tpc.010479

[13] Hollingsworth, P. M., Li, D. Z., van der Bank, M., & Twyford, A. D. (2016). Telling plant species apart with DNA: from barcodes to genomes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *371*(1702). doi: 10.1098/rstb.2015.0338

[14] Wu, F., Mueller, L. A., Crouzillat, D., Petiard, V. & Tanksley, S. D. (2006). Combining bioinformatics and phylogenetics to identify largesets of single-copy orthologous genes (COSII) for comparative,evolutionary and systematic studies: a test case in the euasteridplant clade. *Genetics*, *174*, 1407-1420.

[15] Fredslund, J., Madsen, L. H., Hougaard, B. K., Nielsen, A. M., Bertioli, D., Sandal, N., Stougaard, J. & Schauser, L. (2006). A general pipeline for the development of anchor markers for comparative genomics in plants. *BMC genomics*, *7*(1), 207. doi: 10.1186/1471-2164-7-207

[16] Chapman, M. A., Chang, J., Weisman, D., Kesseli, R. V., & Burke, J. M. (2007). Universal markers for comparative mapping and phylogenetic analysis in the Asteraceae (Compositae).*Theoretical and Applied Genetics*, *115*(6), 747-755. doi: 10.1007/s00122-007-0605-2

[17] Cabrera, A., Rosyara, U. R., De Franceschi, P., Sebolt, A., Sooriyapathirana, S. S., Dirlewanger, E., Quero-Garcia, J., Schuster, M., Lezzoni, A. F. & van der Knaap, E. (2012). Rosaceae conserved orthologous sequences marker polymorphism in sweet cherry germplasm and construction of a SNP-based map. *Tree Genetics & Genomes*, *8*(2), 237-247. doi: 10.1007/s11295-011-0436-9

[18] Liewlaksaneeyanawin, C., Zhuang, J., Tang, M., Farzaneh, N., Lueng, G., Cullis, C., Findlay, S., Ritland, C. E., Bohlmann, J. & Ritland, K. (2009). Identification of COS markers in the Pinaceae. *Tree Genetics & Genomes*, *5*(1), 247-255. doi: 10.1007/s11295-008-0189-2

[19] Krutovsky, K. V., Elsik, C. G., Matvienko, M., Kozik, A., & Neale, D. B. (2006). Conserved ortholog sets in forest trees. *Tree Genetics & Genomes*, *3*(1), 61-70. doi: 10.1007/s11295-006-0052-2

[20] Chandrasekar, A., Riju, A., Sithara, K., Anoop, S., & Eapen, S. J. (2009). Identification of single nucleotide polymorphism in ginger using expressed sequence tags. *Bioinformation*, *4*(3), 119-122.

[21] Wheelan, S. J., Church, D. M. & Ostell JM. (2001). Spidey: a tool for mRNA-togenomicalignments. *Genome Res 11*, 1952-1957.

[22] Rozen S &Skaletsky,H. (2000). Primer3 on the WWW for generalusers and for biologist programmers. *Methods Mol Biol*, *132*, 365-386.

[23] Drummond, A. J., Kearse, M., Heled, J., Moir, R., Thierer, T., Ashton, B. & Wilson, A. (2010). Geneious v5.0.4, Available from http://www.geneious.com/