Comparative de novo flower transcriptome analysis of polygamodioecious tree *Garcinia indica*

Reshma V. Patil¹ · Kiran D. Pawar¹

Received: 28 September 2018 / Accepted: 31 January 2019 / Published online: 9 February 2019
© King Abdulaziz City for Science and Technology 2019

Abstract

To extend our understanding of molecular mechanism of sex determination in agro-economically important, polygamodioecious tree *Garcinia indica* (Kokum), high-throughput, next-generation flower transcriptome sequencing (NGS), and comparative analyses were performed to investigate differentially expressed gene in bisexual, female, and male flowers. A total of 49414 unigenes in BS, 45944 unigenes in FL, and 49028 unigenes in ML flowers were annotated. KO annotations revealed that 25 functional categories were large number of genes which were annotated to 'signal transduction'. We identified 33 genes for 'auxin response factor' and 50 for 'ethylene-responsive factor' whose expression changed significantly in all the three paired library combinations. Furthermore, key regulators of floral development such as FLC, SVP, AP1, AP2, AP3, AG, AGL2, AGL4, AGL9, and PI were identified. A total of 327 differentially expressed MADS-box genes were identified in *G. indica* transcriptome. Analysis of MADS-box genes identified five genes such as MADS AGL11, CRS2-associated factor chloroplastic, conserved hypothetical protein, uncharacterized protein LOC104422218, and MADS-box JOINTLESS-like isoform X3 significantly expressed in only FL flower. In addition, number of DEGs like dynamin 2A, auxin response factor, and spermidine synthase involved in sex expression and reproduction were discovered. The expression patterns of selected genes matched well with the expression levels of unigenes by transcriptome sequencing. Our large-scale comparative analyses may provide valuable hints for the next insights into the molecular mechanism of sex determination in *G. indica*.

Keywords *Garcinia indica* · Polygamodioecious · Transcriptome · MADS-box gene · Sex determination · Floral development

Abbreviations

BS	Bisexual flowers
FL	Female flowers
ML	Male flowers
SVP	Short vegetative phase
AP	Apetela
PI	Pistilata
AGL	Agamous-like
FLC	Flowering locus C
DEGs	Differentially expressed genes

Introduction

Garcinia indica Choisy (family Clusiaceae), commonly known as Kokum or Kokum butter tree, is an endemic plant to the Western Ghats of India (Shetty and Kaveriappan 2001) which also grows in tropical Asia, Africa, and Polynesia (Nayak et al. 2010; Kaur et al. 2012). This phytochemically and agro-economically important tree species is commercially cultivated in western coastal states of India for its fruits which are used in many culinary preparations in these parts of India. The plant has been reported with many remarkable medicinal properties, the most important of which is anti-obesity and anti-cholesterol (Berkhout et al. 1990) properties. The fruit rinds contain hydroxycitric acid (–)-(HCA) and garcinic acid which strongly inhibit the activity of citrate cleavage enzyme (Watson et al. 1969; Jena et al. 2002; Shara et al. 2004; Preuss et al. 2004), and, therefore, can be used in the formulations of anti-obesity drugs worldwide. In addition to these properties, fruits are also reported with

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s13205-019-1601-3>) contains supplementary material, which is available to authorized users.

✉ Kiran D. Pawar
kdp.snst@unishivaji.ac.in; pawarkiran1912@gmail.com

¹ School of Nanoscience and Biotechnology, Shivaji University, Kolhapur, Maharashtra, India

activities such as anti-helminthic and anti-tumor, as cardi-
 tonic and in the treatment of dysentery and piles (Reddy
 and Prabhakar 1994). As Kokum is slow growing, cross-
 pollinated, polygamodioecious tree species, the differen-
 tiation between female, male, and bisexual plants is pos-
 sible only after flowering stage that takes almost 7–8 years
 of age (George et al. 2002). The sexual variation in seven
 populations of *G. indica* of Western Ghats of India was
 estimated to ascertain the extent of floral trait variation,
 and it was found to be trioecious in nature (Joseph and
 Murthy 2015). The seeds of Kokum are recalcitrant due to
 high sensitivity to desiccation and freezing with shelf life
 of only 4 weeks. Thus, it becomes difficult to raise seed-
 lings throughout the year (Malik et al. 2005). Moreover,
 it was reported that the progeny of *G. indica* raised from
 seeds yields male, female, and bisexual plants in the ratio
 of 37:55:8 (Thatte and Deodhar 2012). Determination of
 gender of a seedling at the early stage of development
 is expected to enable the cultivators to maintain required
 ratio of male-to-female plants for commercial production.
 This necessitates the methods to identify the gender of the
 plant at an early stage for maintaining the proper ratio of
 male-to-female plants.

Sex determination in plants is fundamental develop-
 ment process, particularly in economic crops, because the
 sexual phenotype decides how they are bred and cultivated
 (Tanurdzic and Banks 2004). Differences between female
 and male plants are primarily detected in reproductive
 organs, which occur through differential growth, repres-
 sion, or abortion of sex organs in unisexual flowers (Mat-
 sunaga 2006; Chuck 2010). Genes such as floral meristem
 identity genes, flowering genes, and floral organ identity
 genes are known to play a major role in sex determination
 of unisexual flowers (Miko 2008; Iovene et al. 2014). As
 proposed by the developmental and molecular studies in
Arabidopsis thaliana and *Antirrhinum majus* (Coen and
 Meyerowitz 1991), floral organ development is determined
 mainly by MADS-box gene. In wide range of angiosperms,
 MADS-domain transcription factors are encoded by floral
 organ identity genes which are much enlarged in plants
 compared to the other eukaryotes (Gramzow and Theissen
 2010). The possible strategy to identify the genes in the
 developmental process is to screen mRNA transcripts that
 are present in one and absent in the other sample (Perl-
 Treves 1999). The transcriptome sequencing, especially
 with high-throughput next-generation sequencing (NGS)
 technology, is an attractive alternative to whole-genome
 sequencing. The mRNA transcripts' sequencing generates
 high functional information content that can be used to
 identify new and novel genes with the known or predicted
 functions (Andersen and Lubberstedt 2003; Bouck and
 Vision 2007). Many studies have been conducted in candi-
 date genes for sex determination such as *Ricinus communis*

(Tan et al. 2016), *Carica papaya* (Zeng et al. 2014), and
Eucalyptus grandis (Vining et al. 2015) and others.

In this study, we performed high-throughput next-gener-
 ation sequencing of transcriptome of bisexual (BS), female
 (FL), and male (ML) flowers of *G. indica* using NGS Illu-
 mina TruSeq, and analyzed the data to discover the differen-
 tially expressed genes (DEGs) in three tested flower types.
 Based on transcriptome sequencing and analysis, we first
 identified the putative functions of unigenes by subject-
 ing to the blastx search and Gene Ontology was assigned
 using Blast2GO pro. To calculate differential gene expres-
 sion, gene abundance was mapped to unigenes. The uni-
 genes homologous to MADS-box genes were extracted to
 analyze differentially expressed MADS-box genes. This
 study will provide useful information on genes differentially
 expressed in three studied flower types and reveal the key
 genes responsible for sex determination in *G. indica*.

Materials and methods

Flower collection and RNA extraction

Six-to-eight fully grown BS, FL, and ML flowers (Fig. 1)
 were collected and sampled from individual trees under nat-
 ural field conditions from Sawantwadi (16°60'N, 73°19'E)
 and Vengurla (16°80'N, 73°06'E) in Kokan region of
 Maharashtra during flowering season in the month of Janu-
 ary–February. Sampled flower were stored in RNALater
 (Sigma–Aldrich, Waltham, Massachusetts, US) at 4 °C until
 further use for RNA extraction. Total RNAs were isolated

Fig. 1 Tree habit of *Garcinia indica* and three different types of flowers used in the study

from 6 to 8 mature, fully grown flowers and pooled together prior to RNA isolation using manual protocol. The quality of the total RNA was checked on 1% denatured agarose gel and quantified using Qubit fluorometer (ThermoFisher Scientific, USA).

NGS library construction and sequencing

Synthesis of cDNAs and 2×150 paired-end library preparations for each of the BS, FL, and ML flowers were carried out starting with 2.5 µg of total RNA using the TruSeq RNA sample Prep Kit V2 (Illumina, California US) following the manufacturer's instructions. Quality of sequencing libraries was checked using the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA) technology using High-Sensitivity (HS) DNA chip as per manufacturer's instructions. After obtaining the Qubit concentration for the library and estimating the mean peak size from the Bioanalyzer profile, libraries were loaded into Illumina platform for cluster generation and sequencing. The template fragments were sequenced in both forward and reverse directions using paired-end sequencing. The library molecules were bound to complementary adaptor oligos on paired-end flow cell. The adaptors were designed to allow selective cleavage of the forward strands after re-synthesis of the reverse strand during sequencing. The copied reverse strands were then used to sequence from the opposite end of the fragment.

Sequence assembly and gene annotation

All the raw sequences data were processed to remove low-quality reads and adaptor sequences. The high-quality (HQ) reads from all three libraries were assembled using Trinity, a De Bruijn Graph-based assembler. Following transcript construction, unigenes were generated from the transcript using TGICL for three flower-type libraries. The unigene size histogram was constructed using an in-house perl script. The primary method of annotation of unigenes from sequence and domain-based databases was through BLASTX. The unigenes were annotated with known gene names and function by homology searching against NR and Uniport databases using BLASTX.

To assign the functions to unambiguously mapped genes identified in all the libraries, functional annotation was performed to gene ontology (GO) terms (<http://www.geneontology.org/>) that consisted of three categories: molecular function, cellular component, and biological process. Based on the NR annotation, the Blast2GO software (Conesa et al. 2005) was used to obtain GO annotations for the unigenes. The domain-specific annotation was done using KOG database. Common numbers of unigenes were annotated between sequence-based and domain-based databases for each sample and the plots were generated in Venny ([\[cnb.csic.es/tools/venny/\]\(http://cnb.csic.es/tools/venny/\)\). Furthermore, the WEGO plot for GO annotation was plotted for three samples. The distribution of unigenes from each sample across different pathway categories in KEGG \(<http://www.genome.jp/kegg>\) \(Kanehisa et al. 2008\) and the pathway annotation was done using KAAS server.](http://bioinfopg.</p>
</div>
<div data-bbox=)

Analysis of differential gene expression

The HQ reads were mapped on to the corresponding unigenes using bowtie2 (Ben et al. 2009) (<http://bowtie-bio.sourceforge.net/index.shtml>) to generate BAM file which, in turn, was used to estimate unigene abundance (i.e., generate raw read count for each unigene) for which reads were mapped to unigenes. Common hit accessions based on BLAST against NCBI-NR database considering *E* value $1e-6$ were considered for differential gene expression analysis and difference in the level of expression was represented in log2 ratio of unigene abundances. Finally, these files were used as input to DESeq2, an R package implementing Generalized Linear Model, to calculate differential gene expression (DEG) in three paired combinations as BS vs FL, BS vs ML, and FL vs ML. The heat map was generated considering Log2Fold change (L2FC) using heatmap.2, an R package.

Identification of orthologous pairs and analysis of positive selection

The gene sequences were predicted using Trans Decoder (<http://transdecoder.sf.net>) by predicting coding sequences (genes) from the unigenes followed by extracting orthologous pair of genes between two samples in combinations. Orthologous pair was calculated by implementing the Reciprocal Blast Hit (RBH) where each of the samples in combination was considered query and database turn-wise. We then used BLASTN to find homologous genes between BS and FL, BS and ML, and FL and ML libraries. The hits were sorted by bit score (also alignment length) so as to get best hits for each gene. The shortlisted pairs were then further filtered by choosing only those pairs whose unigene length was same. Finally, the Ka/Ks ratio was calculated using KaKs calculator, a stand-alone C++ tool implementing sliding window technique for positively selected sites, and various models for calculating substitution rate. Model averaging: a maximum-likelihood method was then used to calculate the Ka/Ks ratio.

Identification and annotation of MADS-box genes

MADS-box gene-specific expression analysis was done by identifying the MADS-box genes in the three flower-type libraries. This was performed by discovering the unigenes representing the genes from MADS-box family. A set of

327 well-known MADS-box genes from six different plants, namely, *A. thaliana*, *Jatropha curcas*, *Populus trichocarpa*, *R. communis*, *Vitis vinifera*, and *Citrus sinensis*, were short-listed and reciprocal blast hits (RBH) was implemented to extract homologous MADS-box gene using TBLASTX with E value $< 1e-3$. The mutual best hits between MADS-box genes and each of the flower-type library unigenes were considered as its corresponding MADS-box genes expression values in that respective library.

Real-time RT-PCR validation

First-strand cDNA was generated from 1 µg total RNA isolated from each of the flower and whorls using ReadyScript® cDNA Synthesis Mix (Sigma–Aldrich, Saint Louis, MO 63103, USA). Primers for quantitative reverse transcription PCR (qRT-PCR) were designed using Primer-BLAST NCBI (Primer 3 software) and synthesized by Sigma-Aldrich Co., Ltd. qRT-PCR was performed on Bio-Rad CFX96 Real-time system 1000 touch thermal cycler (Bio-Rad, Richmond, CA, USA) using a SYBR Green-based PCR assay. Each reaction mixture was 20 µl containing 6 µL of diluted first-strand cDNAs, 10 mM of each primer, and 10 µL of SYBR Green PCR mix (Genei, Bangalore, India). The qPCRs were run as follows: 95 °C for 2 min, followed by 40 cycles of 95 °C for 15 s, 61–67 °C for 30 s, and 65–95 °C dissociation for 10 s. The *ubiquitin* gene was used as an internal standard to calculate relative fold differences based on comparative cycle threshold ($2^{-\Delta\Delta C_t}$) values.

Results

Transcriptome sequencing, sequence assembly, and unigene statistics

For the study of flower transcriptome and its comparative analyses in BS, FL, and ML flowers, three cDNA libraries were generated and sequenced on Illumina HiSeq 2000 next-generation sequencing platform to generate > 10 GB data per sample [data deposition: NCBI PRJNA360660 (SRP096940), Male buds (SRS1926202), female buds (SRS1926203), and bisexual buds (SRS1926204)]. Table S1 shows the statistics of high-quality data, transcripts generated by Trinity assembly, and unigenes obtained from master assembly. After stringent quality checks and data cleaning such as filtering adaptors and low-quality sequences, 70 million HQ reads with an average GC percentage of 52.6 per samples were generated. To improve the assembly and annotation, the HQ reads from all three libraries were assembled using Trinity. As a result, a total of 141,517, 63,145, and 84,531 transcripts with an average transcript sizes of 453, 397, and 421 bp were obtained from BS, FL, and ML

libraries, respectively. The transcript length in analyzed libraries varied in the range 201–16.5 kb in BS, 201–14.6 kb in FL, and 201–6.1 kb in ML library. The N50 of transcripts was approximately 439 bp for three libraries.

Following transcript construction, unigenes were generated from transcript using TGICL for three samples. In total, 49,414 unigenes with 735 bp mean length in BS, 45,944 unigenes with 415 bp mean length in FL, and 49,028 unigenes with 499 bp mean length in ML library were identified. Furthermore, the unigene size histogram was constructed by obtaining the length of each sequence using an in-house perl script (Fig. S1). Among 82,547 unigenes identified, 9193 (11.1%) unigenes from BS, 31,445 (38.09%) unigenes from FL, and 26,641 (32.27%) unigenes from ML were in 200–400 bp size length, whereas 20,311, 8664 and 12471 unigenes from BS, FL, and ML library, respectively, were in the size range 400–600 bp. Similarly, 8445, 2793, and 4913 unigenes from BS, FL, and ML library, respectively, were in the size range of 600–800 bp (Fig. S1).

Functional annotations

In the present study, the annotations of the unigenes were done through BLAST search against the sequence-based databases NCBI-NR, UniProtKB, GO, and KEGG, and the domain-based databases Pfam and KOG with an E value cut-off of $1e-6$. Gene ontology (GO) for annotated unigenes was obtained using Blast2GO and associate ontology term. Based on BLAST search, 33,318 (67.4%) unigenes from BS, 27,706 (60.3%) unigenes from FL, and 32,726 (66.7%) unigenes from ML libraries were annotated in at least one database (Table S2). In BS library, a total of 17,414 unigenes were annotated in domain-based database and 33,310 unigenes were annotated in sequence-based database alignment. Among these, 1673 (3.3%) unigenes had hits in all six public databases. Similarly in FL library, 12,579 unigenes could be annotated in domain-based database and 27,692 unigenes could be annotated in sequence-based database alignment. Of these, 989 (2.1%) unigenes had hits in all six public databases. Annotation of unigenes from ML library resulted in annotations of 15,593 and 32,170 unigenes in domain-based database and sequence-based databases, respectively. Among these, 1533 (3.1%) unigenes had hits in all six databases (Table S2).

Figure S2 shows the overlap of sequence-based and domain-based database-annotated unigenes between the pairwise combinations of studied flower libraries. These analyses indicated that, in BS library, a total of 8253 (12.4%) domain-based unigenes were shared between Pfam and KOG (Fig. S2A), whereas 3294 (9.9%) sequence-based unigenes were common between Uniport, NR, and KEGG (Fig. S2B). In total, 23,574 unigenes were shared between Uniport and NR, 3294 unigenes were shared between Uniport and

KEGG, and 3290 unigenes were shared between NR and KEGG. In FL library, a total of 5806 (46.2%) domain-based unigenes were shared between Pfam and KOG (Fig. S2C) and 2563 (9.3%) sequence-based unigenes were common among Uniport, NR, and KEGG (Fig. S2D). Among these shared unigenes, 18,008 unigenes were shared between Uniport and NR, 2563 unigenes between Uniport and KEGG, and 2678 unigenes between NR and KEGG. Likewise, in ML library, a total of 7676 (46.5%) domain-based unigenes were shared between Pfam and KOG (Fig. S2E) and 3204 (9.8%) sequence-based unigenes were common among Uniport, NR, and KEGG (Fig. S2F).

To classify the functions of the predicted *G. indica* unigenes, we used GO analysis and assignments. Figure 2 represents the WEGO output for three GO categories among three libraries. Out of 82,547 annotated unigenes, 20,237 (24.51%) sequences were categorized into three main categories such as biological process, cellular component, and molecular function with 52 functional groups. The comparative GO assignment analyses of three libraries assigned unigenes to 15 terms of molecular function category. In this

category, the highest number of genes was related to ‘binding’ and ‘catalytic activity’ functions (~70%). Among the unigenes assigned to 14 terms of cellular component category, the unigenes required for cell and cell part (~80%) were dominant followed by ‘organelle’ (~50%) and ‘organelle part’ (~20%). In biological process category that contained 23 terms, genes for ‘cellular process’ and ‘metabolic process’ (~70%) were found most abundant, followed by genes involved in ‘response to stimulus’ (~8%) and ‘biological regulation’ (~10%). In the three major categories, the lowest number of unigenes was classified under ‘chemoattractant’ (0.05% in FL library) and ‘protein tag’ (0.02% in ML library and 0.04% in BS library).

The distribution pattern of unigenes across different pathways indicated that 50% unigenes were related to metabolism pathways (Fig. 3). It was found that 3390 (6.8%) from BS, 2609 (5.6%) from FL, and 3317 (6.9%) from ML-annotated sequences were mapped to reference pathways described in KEGG. In total, 25 KEGG pathways were assigned among which highest percentage was assigned to ‘translation’ followed by ‘fold sorting and degradation’ and

Fig. 2 WEGO plot for three GO categories among BS, FL, and ML transcriptome libraries

Fig. 3 Barplot showing distribution of unigenes across different pathways (KO term)

signal transduction'. The unigenes that were mapped to metabolic pathway 'translation' included 558 unigenes from BS, 414 unigenes from FL, and 591 unigenes from ML libraries, whereas unigenes that were mapped to 'fold sorting and degradation' included 490 unigenes from BS, 354 unigenes from FL, and 428 unigenes from ML libraries. Likewise, unigenes mapped to 'signal transduction' comprised of 499 unigenes from BS, 379 unigenes from FL, and 434 unigenes from ML libraries.

Identification of differential expressed genes

For comparative analysis and differential gene expression study, we performed pairwise comparisons of the three libraries as BS vs FL, BS vs ML, and FL vs ML. To study the differential gene expression (DEG), total of 71,460,302 reads from BS library, 72,834,700 reads from FL library, and 70,436,080 reads from ML library were mapped against 49,414, 45,944, and 49,028 unigenes, respectively. DEGs

showing up-regulation and down-regulation in three paired combinations are represented as heatmap in fig. S3.

Based on Log2 fold change ($\text{Log}_2\text{FC} > 0$ up-regulated and $\text{Log}_2\text{FC} < 0$ down-regulated), a total of 21,717 DEGs (7631 up-regulated and 14,086 down-regulated) between BS and FL, 23,369 DEGs (10,240 up-regulated and 13,129 down-regulated) between BS vs ML, and 18,170 DEGs (10,672 up-regulated and 7498 down-regulated) between FL vs ML library were identified (Fig. 4). The DEG statistics showed that 58% of the genes were up-regulated between FL and ML, 43.8% were up-regulated between BS and ML and 35.1% were up-regulated between BS and FL library (Tab. S3). In our analyses, the expression profiles between two different library combinations were also explored and studied more closely by plotting (MA) plot (Fig. 5). These MA plots represent the log fold change (the difference) between two libraries combinations against the average log expression across those libraries (the mean). The MA plot (Fig. 5a) between BS and FL library indicated the noticeable positive skew with a number of very highly up-regulated genes. In

Fig. 4 Bar plot of differential gene expression statistics

Fig. 5 MA plots. Scatter plot of fold change in **a** BS vs FL, **b** BS vs ML, and **c** FL vs ML (y-axis) against mean expression level (x-axis). The horizontal line represents the equal expression of bisexual, female and male samples. Each point represents a gene, and the

red line indicates a log-ratio of zero. The majority of points cluster around the red line. Small triangles at the top and bottom of the plots indicate points that would fall outside of the plotting window

comparison to MA plots of BS vs ML (Fig. 5b) and FL vs ML library (Fig. 5c), the MA plot of BS vs FL library had the number of points in the upper right of the plot corresponding to genes that were both, highly expressed and highly up-regulated. The MA plots of BS vs ML and FL vs ML showed the negative skew indicating a few numbers of up-regulated genes. In particular, there are a number of points in lower right of the plot, corresponding to the genes which are down-regulated.

GO annotation of DEGs between paired library combinations

To analyze the functions of the DEGs and identify differences in the three main categories, viz., biological processes, cellular processes, and molecular functions, GO enrichment analyses were performed on the DEGs between the pair of library combinations. This was accomplished by separating the DEGs into up-regulated ($L2FC > 0$) and down-regulated ($L2FC < 0$) list for pairs of these three combinations and using them for plotting stacked bar plots. The enriched GO terms in the biological processes, cellular processes, and molecular functions for BS vs FL, BS vs ML, and FL vs ML libraries are represented in figures S4, S5, and S6, respectively, and the numbers of genes expressed in GO category for three pairs of libraries are presented in Table S3.

GO annotation of DEGs between BS and FL flowers

In the biological process category between BS and FL library, the dominant terms were ‘oxidation reduction process’, followed by term ‘metabolic process’. A total of 463 genes were up-regulated and 613 genes were down-regulated for ‘oxidation reduction process’, whereas 364 genes were up-regulated and 226 genes were down-regulated for the term ‘metabolic process’ (Table S4 and Fig. S4A). In the cellular component category, a total of 565 genes were up-regulated and 1756 genes were down-regulated for GO term ‘integral component of membrane’, while 253 genes were up-regulated and 587 genes were down-regulated for term ‘membrane component’ (Table S4 and Fig. S4B). In the molecular function category, 1204 genes were down-regulated and 412 genes were found up-regulated for term ‘ATP-binding activity’, while 256 genes were up-regulated and 581 genes were down-regulated for the ‘term nucleic acid-binding activity’ (Table S4 and Fig. S4C).

GO annotation of DEGs between BS and ML flowers

When comparison was made between BS and ML library, it was found that, in biological process category, a total of 490 genes were up-regulated and 560 genes were down-regulated for the term ‘oxidation–reduction process’, and

258 genes were up-regulated and 361 genes were down-regulated for the term ‘regulation of transcription DNA template’ (Table S4 and Fig. S5A). In the cellular component category, a total of 1341 genes were up-regulated and 1598 genes were down-regulated for the predominant GO term ‘integral component of membrane’, whereas a total of 452 genes were up-regulated and 566 genes were down-regulated for the second most dominant term ‘membrane component’ (Table S4 and Fig. S5B). In the molecular function category, a total of 791 and 1096 genes were up-regulated and down-regulated, respectively, for the dominant term ‘ATP-binding capacity’, while 404 genes were up-regulated and 513 genes were down-regulated for the term ‘metal ion binding’ respectively (Table S4 and Fig. S5C).

GO annotation of DEGs between FL and ML flowers

Similarly, the GO annotations of DEGs between FL and ML libraries indicated that 500 genes were up-regulated and 448 genes were down-regulated for the term ‘oxidation reduction’ in the biological process category (Table S4 and Fig. S6A). In cellular component category, 1402 genes were found up-regulated and 540 genes were down-regulated for the most dominant GO term ‘integral component of membrane’, while 477 genes were up-regulated and 338 genes were down-regulated for the term ‘membrane component’ (Table S4 and Fig. S6B). In molecular function category, 827 genes were found up-regulated and 406 were down-regulated for most dominant term ‘ATP binding’, while 417 genes up-regulated and 280 genes down-regulated for the term ‘metal binding’ (Table S4 and Fig. S6C).

Furthermore, we analyzed the expression profiles of DEGs in studied flower types by generating the heat map (Fig. S3). This analysis indicated that maximum numbers of genes were abundantly expressed in BS vs ML followed by BS vs FL and FL vs ML. It was observed that, in these three DEG libraries, 600 genes were up-regulated, whereas 562 genes were down-regulated. The genes like ubiquitin-conjugating enzyme (gil1026073648), Exocyst subunit exo70 family (gil1035916121), importin subunit alpha-2 like (gil1098838214), pectin esterases (gil147782151), polypeptide *N*-acetylgalactose aminyltransferase (gil255552539), typeIIIa membrane cp-wap13 (gil566213126), ringbox Ia (gil593215283), and 26S proteasome non-ATP regulatory subunit (gil658022239) were distinctly up-regulated in BS vs ML library (Table S5).

Identification of genes associated with sex expression

Based on prior knowledge of the putative involvement of the genes in sex determination, the expression levels and functions of DEGs between paired library combinations were

analyzed. In this analysis, several subgroups were found related to sex expression. We identified 33 genes for ‘auxin response factor’ and 50 genes for ‘ethylene-responsive transcription factor’ whose expressions were found changed significantly in all three library combinations (Table S6). Genes associated with tissue or organ developmental processes, signal transduction, auxin transport, phytohormone biosynthesis, and metabolism (such as polyamine and abscisic acid) were identified. These include genes like Dynamin 2A, auxin response factor and Spermidine synthase (Table S7).

Auxin response factor is known to respond hormone stimulus which participates in hormone-mediated signaling pathways, and also plays a role in tissue and organ developmental processes (Tan et al. 2016). In addition, 12 gene (arginine/serine-rich splicing factor) associated with sex differentiation were identified. In BS vs ML and FL vs ML libraries, the genes were distinctly up-regulated, but down-regulated in BS vs FL. Similarly, genes like DNA replication helicase (*dna2*) and *sam50* which are known to be associated with the developmental process in reproduction were found up-regulated in BS vs ML and FL vs ML (Table S8) (Tan et al. 2016).

Identification of orthologous pairs and analysis of positive selection

A reciprocal blast-hit (RBH) method was used to identify putative orthologs between the pairs of three libraries. To understand the influence of stabilizing and diversifying selection between BS, FL, and ML flower types, the comparison of synonymous base substitutions (Ks) and non-synonymous base substitutions (Ka) was calculated (Wang et al. 2010). Table S10 shows the statistics of KaKs calculation between three combination and figure S7 shows distribution of KaKs ratio for three combinations. Following method reported by Novaes et al. (2008) we further classified genes with Ka/Ks. Two binary variables (‘purifying’ and ‘diversifying’) were classified according to their Ka/Ks values were Ka/Ks smaller than 0.15 were classified as purifying, while those with greater than 0.5 were classified as diversifying. By this analysis, we obtained 75 genes from FL vs ML, 257 genes from BS vs FL, and 83 genes from BS vs ML flower genes which were orthologous to three flower types.

Analysis of MADS-box genes

The statistics of MADS-box gene-specific expression analysis is summarized in table S11, and Fig. 6 represents heatmap showing MADS-box gene expression in three libraries, whereas Table 1 shows the list of highly expressed MADS-box genes in three libraries. With these analyses, we identified 18 MADS-box genes which were specific for BS, while 9 for FL and 22 for ML flower. The analysis of DEGs based

on L2FC showed differential expression of a few MADS-box gene in three libraries. ‘Floral homeotic’ (*gil1000937356*) regulator for APETALA 2 was found up-regulated (L2FC 3.75) between BS vs FL and down-regulated (L2FC -3.85) between FL vs ML. ‘Ethylene-responsive transcription factor ERF071’ (*gil1009148309*), a regulator for APETALA2 was highly up-regulated for BS vs FL (L2FC 4.039) and down-regulated for FL vs ML (L2FC -3.337), whereas no expression of this gene was observed in BS vs ML. Similarly, ‘ethylene-responsive transcription factor ERF071 family’, a regulator for APETALA2, showed up-regulation in FL vs ML (L2FC 2.56) and BS vs ML (L2FC 0.152), whereas it was down regulated in BS vs FL (L2FC -1.93). AGAMOUS-like MADS-box AGL8 (*gil1012358607*), MEF2 like factor was up-regulated in both BS vs ML (L2FC 3.299) and FL vs ML (L2FC 3.27), while no expression was observed in BS vs FL. Another ‘AGAMOUS-like MADS-box AGL11, an isoform X1’ (*gil645224327*) was down-regulated in both BS vs FL (L2FC -3.15) and BS vs ML (L2FC -3.26). Likewise, MADS-box transcription factor 6 (*gil743896757*) or MEF2-like factor was highly down-regulated in BS vs FL (L2FC -3.65) and BS vs ML (L2FC -5.12).

This analysis of MADS-box genes in the libraries indicated that five genes were found expressed only in FL library. These five genes included MADS AGL11 (XM_019226355.1) homologous to *V. vinifera* MADS-box transcription factor 15, CRS2-associated factor chloroplastic (AT5G54890.1) homologous to *A. thaliana* RNA-binding CRS1/CRM domain-binding protein, conserved hypothetical protein (AT1G79730.1) homologous to ELF7 protein involved in flowering time, uncharacterized protein LOC104422218 (AT1G06980.1) homologous to *A. thaliana* required for molecular function such as expression in sepal, male gametophyte, and carpel formation and MADS-box JOINTLESS-like isoform X3 (XM_002513991.2) homologous to *R. communis* MADS-box SVP.

Gene expression of the other two libraries showed higher expression of five genes for BS and four genes for ML libraries. The highly expressed genes that could be playing important role in flower identity included MADS-box transcription factor 6-like isoform X1 (NM_001281182.1) homologous to *V. vinifera* MADS3 protein, hypothetical protein B456_009G227500 (XM_006472343.2) homologous to *C. sinensis*, MADS-box transcription factor 23 like (LOC102630220) and AGAMOUS-like MADS-box AGL1 (XM_020677533.1) homologous to AGAMOUS-like MADS-box protein AGL1 (LOC105629543), hypothetical protein JCGZ_26749 (AT4G39040.1) homologous to *A. thaliana* RNA-binding CRS1/YhbY domain protein, and chloroplastic group IIA intron-splicing facilitator chloroplastic gene (AT5G16180.1) homologous to *A. thaliana* CRS1, ATCRS1 ortholog of maize chloroplast-splicing factor CRS1. In case of *G. indica*, the highly expressed

Fig. 6 A heatmap showing MADS-box gene expression in three samples. The colors represent log₁₀-transformed FPKM values, where red represent high expression and blue represents no or very low expression

genes possibly required for male identity could be chromosome region maintenance 1 (AT5G17020.1) for XPO1A, ATCRM1, ATXPO1, XPO1 HIT2 exportin 1A, myb family transcription factor family (AT5G41020.1) for Myb family transcription factor showing DNA-binding activity, MADS-box gene (NM_001281281.1) homologous to *V. vinifera* APETALA1 (AP1), and K-box region and MADS-box transcription factor family isoform 1 (XM_012230319.2) homologous to *J. curcuas* Agamous-like MADS-box protein AGL9 (LOC105644835).

In addition, genes like developmental SEPALLATA 3-like isoform X1 (XM_002514847.2) homologous to

AGL9 of *R. communis* showed slightly higher expression in ML than FL and BS library. The regulator of nonsense transcripts UPF3-like isoform X1 (XM_015530186.1) homologous to *C. sinensis* AGAMOUS-like MADS-box protein AGL11 showed higher expression in BS than ML and FL library, whereas MADS-box transcription partial (NM_001281031.1) homologous to *V. vinifera* flowering-related B class MADS-box protein APETALLA3 (AP3) was slightly expressed in FL library. The floral homeotic DEFICIENS isoform 1 (NM_001281031.1) homologous to *V. vinifera* APETALA 1 (AP1) was exclusively expressed in ML, K-box region and MADS-box transcription factor

Table 1 Highly and exclusively expressed MADS-box gene in three flower transcriptome libraries analyzed

Library	MADS-box gene	Gene ID	Homologues to	Function
Bisexual	Hypothetical protein JCGZ_26749	AT4G39040.1	<i>Arabidopsis thaliana</i>	RNA-binding CRS1/YhbY domain protein
	Chloroplastic group IIA intron-splicing facilitator chloroplastic	AT5G16180.1	<i>Arabidopsis thaliana</i>	CRS1, ATCRS1 ortholog of maize chloroplast-splicing factor CRS1
	MADS-box transcription factor 6-like isoform X1	NM_001281182.1	<i>Vitis vinifera</i>	MADS-box protein 3 (MADS3) mRNA
	Hypothetical protein B456_009G227500	XM_006472343.2	<i>Citrus sinensis</i>	MADS-box transcription factor 23 like (LOC102630220)
	Agamous-like MADS-box AGL1	XM_020677533.1	<i>Jatropha curcas</i>	Agamous-like MADS-box protein AGL1 (LOC105629543), mRNA
Female	MADS AGL11	XM_019226355.1	<i>Vitis vinifera</i>	MADS-box transcription factor 15
	CRS2-associated factor chloroplastic	AT5G54890.1	<i>Arabidopsis thaliana</i>	RNA-binding CRS1/CRM domain-binding protein involved in biological process
	Conserved hypothetical protein	AT1G79730.1	<i>Arabidopsis thaliana</i>	For ELF7 protein, Encodes a PAF1 homolog that is involved in the control of flowering time
	Uncharacterized protein LOC104422218	AT1G06980.1	<i>Arabidopsis thaliana</i>	Molecular function, expressed in sepal, male gametophyte, carpel
	24-methylenesterol C-methyltransferase 2	AT1G20330.1	<i>Arabidopsis thaliana</i>	Encodes a sterol-C24-methyltransferases involved in sterol biosynthesis
Male	Peroxidase 63-like	AT1G24110.1	<i>Arabidopsis thaliana</i>	Peroxidase activity, heme binding
	MADS-box JOINTLESS-like isoform X3	XM_002513991.2	<i>Ricinus communis</i>	MADS-box SVP
	Chromosome region maintenance 1	AT5G17020.1	<i>Arabidopsis thaliana</i>	XPO1A, ATCRM1, ATXPO1, XPO1 HIT2 exportin 1A, functions as receptors for nuclear export
	Myb family transcription factor family	AT5G41020.1	<i>Arabidopsis thaliana</i>	Myb family transcription factor, DNA binding, sequence-specific DNA-binding transcription factor activity
	MADS-box	NM_001281281.1	<i>Vitis vinifera</i>	Apetala1 (AP1), mRNA, MEF2 like transcriptional regulator binds DNA
	K-box region and MADS-box transcription factor family isoform 1	XM_012230319.2	<i>Jatropha curcas</i>	Agamous-like MADS-box protein AGL9 (LOC105644835) transcript variant X1, mRNA, binds DNA

family isoform 1 (XM_012230319.2) homologous to *J. curcas* AGAMOUS-like MADS-box protein AGL9 was also exclusively expressed in ML, and AGAMOUS-like MADS-box AGL1 (XM_020677533.1) homologous to *J. curcas* AGAMOUS-like MADS-box protein AGL1 was found expressed in BS (Table S12).

qRT-PCR validation of flowering-related genes

To quantitate the level of expression of flower-related genes that were identified using RNA-sequencing and to investigate if their expression level is in congruence with the expression levels obtained by RNA-seq data, we performed relative quantification qRT-PCR assays with independent samples collected from the flowers. To accomplish this, six flowering-related unigenes were selected to validate the RNA-seq data. The list of genes selected and the primer sequences designed are listed in table S13. The qRT-PCR analyses indicated that expression levels of AP3, AGL9,

AGL11, AG, SEP1, and AG pathway unigenes matched well with the expression levels of NM_001281031.1 XM_002514847.2, XM_015530186.1, XM_006467514.1, XM_012232674.2, and AT5G23150.1 unigenes indicated by RNA-sequencing (Fig. 7).

Discussion

In polygamodioecious trees such as *G. indica*, the genetic variation resulting in reproductive system differences between flower gender types can be studied by establishing large-scale transcriptome and identifying the number of differentially expressed genes. This would help to identify key regulator genes of floral development and gender determination.

Our annotation results indicated that more than 31,152 (64.5%) of unigenes had homologs in the Nr database, whereas only 4300, 3400, 2500, and 1000 of unigenes had

Fig. 7 A comparison of fold changes over time for expression by real-time PCR. Color bars indicate relative expression between floral buds. The scale of each graph was determined individually to maximize the view of the particular pattern associated with expression.

The results represent samples as *M* male, *F* female, and *B* bisexual with fold change expression of six selected candidate genes. qPCR fold change is represented from **a–f**, **a** AP3, **b** AGL9, **c** AGL11, **d** AG, **e** SEP1, and **f** AG pathway

homologs in the Nr database in *J. curcas*, *P. trichocarpa*, *R. communis*, and *V. vinifera*. Among these species, the *J. curcas* showed the highest percentage of matches indicating a closer genetic relationship between *J. curcas* and *G. indica*. In *J. curcas*, some abundant GO terms were identified in the molecular function category like ‘catalytic activity’ and ‘binding activity’ followed by transporter activity, ‘cell’ and ‘cell part’ in cellular component, and ‘cellular and metabolic process’ in biological component. (Xu et al. 2016). In congruence with this observation in *J. curcas*, we also observed similar pattern of GO annotation and classification in *G. indica*. These may reflect conserved characteristics in angiosperms. The KO annotation of unigenes revealed 25 functional categories where functions under metabolism are predominant. A large number of genes were also annotated to ‘signal transduction’. Similar results have been obtained in the studies of genes associated with floral development in Orchidaceae and Bambusoideae (Zhang et al. 2012, 2013). In addition, large numbers of unigenes were annotated to carbohydrate metabolism. Pathways involving carbohydrate transport and metabolism are involved in plant floral induction (Rady and El-Yazal 2013; Vaddepalli et al. 2015). These pathway assignments provide a valuable resource for investigating specific processes, functions, and pathways in *G. indica*.

To study and identify the key regulator genes of floral development and gender determination, sex-related genes were screened from the DEGs in three libraries. The molecular functions assigned included three biological process subcategories such as biological adhesion, viral reproduction pigmentation, and three cellular component subcategories such as extracellular matrix, extracellular matrix part, and membrane-enclosed lumen. This indicated that these molecular functions are related to the sex differentiation of flowers, such as carpel fusion and stamen degeneration. The molecular functions and biological pathways suggest that they contribute to the late period of sex differentiation, such as the extracellular matrix, biological adhesion, protein-binding transcription factor activity, active base and amino acid metabolism, glycan degradation, protein biosynthesis, brassinosteroid biosynthesis, terpenoid biosynthesis, phosphatidylinositol signaling system, and ubiquitin-mediated proteolysis appear to participate in sex determination (such as the development of stamen and carpel fusion) (Xu et al. 2016). In the present study, we found amino acid metabolism (277 genes), glycan biosynthesis and metabolism (93 genes), metabolism of terpenoids and polyketides (75 genes), and biological adhesion related genes were abundant in *G. indica*.

Sex expression in plants is controlled by genetic factors and non-genetic fluctuations (Shifriss 1956; Ainsworth

1998) and phytohormones play important roles in sex determination in some species (Yin and Quinn 1995; Tanurdzic and Banks 2004). In cucumber, ethylene, a true sex hormone, was proven to be a 'female' hormone that exerts a strong feminizing effect (Dellaporta and Calderon-Urrea 1993). Auxin also promotes cucumber femaleness, potentially playing an indirect role in sex determination by promoting the action of ethylene (Yamasaki et al. 2000). A previous study showed that ethylene and auxin were involved in floral organogenesis (Ruonala et al. 2006). Similarly, in *G. indica*, the majority of DEGs related to ethylene and auxin were up-regulated in BS vs ML and FL vs ML. Our results suggested that DEGs related to pathways of plant hormone response, hormone-mediated biosynthesis, and metabolic processes may have roles to participate in the regulation of sex expression and reproduction by changing the hormone level and transport. Similar to our observation, the genes like auxin response factor were also found to contribute to embryo sac development in instaminate flower in *J. curcas* (Xu et al. 2016). However, validation and confirmation of the roles of this DEG in sex determination of *G. indica* would require further research efforts. In addition, auxin-signaling pathways were thought to involve in stamen development and also were important for the embryo sac development (Xu et al. 2016). Similarly, in *R. communis*, the genes involved in auxin transport, auxin response factor, and other involved in plant hormone-mediated biosynthetic and metabolic process were down-regulated in leaf apices and young racemes, but up-regulated during the transformation from leaf apices to floral bud. Furthermore, the expression profile of phytohormones (auxin response factor) and signal transduction exhibited similar expression pattern which may specify their role in sex determination (Tan et al. 2016). In plants, phytohormones modulate various growth and developmental events through signal transduction pathways (Bleecker and Kende 2000; Davies 2004).

We also detected programmed cell death (PCD)-related gene (cysteine protease) which was significantly up-regulated in BS vs FL followed by FL vs ML and BS vs ML (Table S9). Plants like castor, maize, cucumber, and campion (Ye et al. 1991; Dellaporta and Calderon-Urrea 1993; Kater et al. 2001; Tan et al. 2016) follow a sex determination pathway involving arrest of preformed sexual organs in bisexual primordia in which few PCD-related genes participate (DeLong et al. 1993). These results are consistent with the recent finding of PCD-related genes in *R. communis* which are significantly up-regulated in female apices at the 5–7 leaf stage, but down-regulated at the two other stages (early leaf apices and initial inflorescence) (Tan et al. 2016). Similarly, the recent findings in PCD-related genes, including cysteine protease, are more highly expressed at the peak of anther abortion in cytoplasmic male sterility CMS and genetic male sterility GMS lines of cotton (Lorrain et al. 2003; Wei et al.

2013). It is possible that PCD-related genes might play an important role in *Garcinia* apical development and, thus, be associated directly or indirectly with sex expression.

Many of the key regulators of floral development are members of the MADS-box family and transcription factors (reviewed in Causier et al. 2010; Litt and Kramer 2010). A total of 45 common and 327 differentially expressed MADS-box genes were identified in *G. indica* transcriptome, many of which are previously identified in *A. thaliana* (Pa enicova et al. 2003), *J. curcas* (Chen et al. 2016), and *V. vinifera* (Grimplet et al. 2016). Most of the MADS-box gene were highly expressed in FL than ML and BS flowers. Among these, our study identified FLC (FLOWERING LOCUS C), SVP (SHORT VEGETATIVE PHASE), AP1 (APETALA1), and AP3 (APETALA3) genes which might be responsible for sex determination in *G. indica*. Studies in *Arabidopsis* show that MADS-box genes identified in *G. indica* belong to MIKC MADS-box gene family (Pa enicova et al. 2003). Similarly, some identified intermediate genes such as AGL2, AGL4, and AGL9 act as mediator between floral meristem and organ identity genes (Pnueli et al. 1994; Savidge et al. 1995), but, now, they are interpreted as class E genes possibly involved in the formation of floral quartets in whorl two, three, and four organs (Becker and Theissen 2003). These findings indicate that AGL2/4/9 have redundant functions required for petal, stamen, and carpel development, and to prevent the indeterminate growth of the flower meristem (Pelaz et al. 2000). The floral organ identity genes specify the organ identity within each whorl of the flower by activating 'realizator genes'. After fertilization of the flower, MADS-box genes such as AGL1, AGL5, and FUL are required for proper fruit development (Gu et al. 1998; Liljegren et al. 1998). The present study of flower transcriptome of *G. indica* supports the earlier findings.

Several class of genes from organ identity likes AGAMOUS (AG), APETALA2 (AP2), APETALA3 (AP3), and PISTILATA (PI) (Yanofsky et al. 1990; Jack et al. 1992; Goto and Meyerowitz 1994; Jofuku et al. 1994) have been identified in the current study. The homologous genes of A, B, and C function were also identified in the present study. AGL11 is expressed in the early stage of ovule development and maintained after pollination for seed development (Steven et al. 1995). Similarly, at later stage of flower development, AGL1, AGL11, and AGL13 RNAs get accumulated in developing ovules or specific tissue type of ovule (Ma et al. 1991). SVP is considered as the one of the strong candidates for sex determination from *J. curcas*. SVP and TFL1 are inhibitors of the transition from vegetative growth to flowering (Hartmann et al. 2000). The previous study supports the findings related to transcription analysis in *G. indica*.

In summary, a large number of DEGs were identified in *G. indica* for flower development. Flower development-related genes were characterized based on DEGs and

MADS-box heat map; some were validated by qRT-PCR analysis. Furthermore, the expression level of MADS-box gene was compared by real-time PCR analysis. The identification and analysis of these genes will aid us in elucidating the regulatory mechanisms in *G. indica*.

Our study has shown that *G. indica* possess a unique set of genes that are involved in growth and developmental process including a rich set of candidate genes related to sex determination. These unique set of genes specifically expressed in only one of the library are important source for understanding the sex determination mechanism and regulation and will eventually contribute to *Garcinia* breeding improvement techniques. The DEG library analysis shows that different molecular and biological pathways contribute for late period of sex determination, whereas the expression pattern of identified MADS-box gene may have role to play in sex determination. Overall, our data provide valuable information about sex differentiation of flowers of *G. indica* and further validation study on candidate genes will be helpful in revealing the sex differentiation mechanism in *G. indica*.

Acknowledgements We would like to thank Xcleris Ltd. Ahmedabad, India for transcriptomics data analysis and DST-SERB (YSS/2015/000319) for their financial support.

Compliance with ethical standards

Conflict of interest There is no conflict of interest.

References

- Ainsworth CC (1998) Sex determination in plants. *J Exp Bot* 49:70–72
- Andersen JR, Lubberstedt T (2003) Functional markers in plants. *Trends Plant Sci* 8:554–560
- Becker A, Theissen G (2003) The major clades of MADS-box genes and their role in the development and evolution of flowering plants. *Mol Phylogenet Evol* 29:464–489
- Ben L, Cole T, Mihai P, Steven S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* 10:25
- Berkhout AT, Havekes ML, Pearce JU, Groot PH (1990) The effect of (–) HCA on the activity of the low density lipoprotein receptor and 3-hydroxy-3-methylglutaryl-CoA reduction levels in the human hepatoma cell line the Hep42. *Biochem J* 272:181–186
- Bleecker AB, Kende H (2000) Ethylene: a gaseous signal molecule in plants. *Annu Rev Cell Dev Biol* 16:1–18
- Bouck A, Vision T (2007) The molecular ecologist's guide to expressed sequence tags. *Mol Ecol* 16:907–924
- Causier B, Schwarz-Sommer Z, Davies B (2010) Floral organ identity: 20 years of ABCs. *Semin Cell Dev Biol* 21:73–79
- Chen M, Pan B, Fu Q, Tao Y, Herrera JM, Niu L, Ni J, Dong Y, Zhao ML, Xu ZF (2016) Comparative transcriptome analysis between gynoeious and monoecious plants identifies regulatory networks controlling sex determination in *Jatropha curcas*. *Front Plant Sci* 7:1953
- Chuck G (2010) Molecular mechanisms of sex determination in monoecious and dioecious plants. *Adv Bot Res* 54:53–83
- Coen ES, Meyerowitz EM (1991) The war of the whorls: genetic interactions controlling flower development. *Nature* 353:31–37
- Conesa A, Götz S, García-Gómez M, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21:3674–3676
- Davies PJ (2004) Plant hormones: biosynthesis, signal transduction, action! Kluwer academic publishers, Dordrecht
- Dellaporta SL, Calderon-Urrea A (1993) Sex determination in flowering plants. *Plant Cell* 5:1241–1251
- DeLong A, Calderon-Urrea A, Dellaporta SL (1993) Sex determination gene TASSELSEED2 of maize encodes a short-chain alcohol dehydrogenase required for stage-specific floral organ abortion. *Cell* 74:757–768
- George GL, Lakshmanan R, Mathew KM, Rao YS (2002) Review of research in kokum camboge and relates species. *J Spices Aromat Crop* 11:1–12
- Goto K, Meyerowitz EM (1994) Function and regulation of the *Arabidopsis* floral homeotic gene PISTILLATA. *Genes Dev* 8:1548–1560
- Gramzow L, Theissen G (2010) A hitchhiker's guide to the MADS world of plants. *Genome Biol* 11:214
- Grimplet J, Martinez-Zapater JM, Carmona MJ (2016) Structural and functional annotation of the MADS-box transcription factor family in grapevine. *BMC Genom* 17:80
- Gu Q, Ferrándiz C, Yanofsky MF, Martienssen R (1998) The *FRUIT-FULL* MADS-box gene mediates cell differentiation during *Arabidopsis* fruit development. *Development* 125:1509–1517
- Hartmann U, Hohmann S, Nettekheim K, Wisman E, Saedler H, Huijser P (2000) Molecular cloning of SVP: a negative regulator of the floral transition in *Arabidopsis*. *Plant J* 21:351–360
- Iovene M, Yu Q, Ming R, Jiang J (2014) Evidence for emergence of sex-determining gene (s) in a centromeric region in *Vasconcellea parviflora*. *Genetics* 199:413–421
- Jack T, Brockman LL, Meyerowitz EM (1992) The homeotic gene *APETALA3* of *Arabidopsis thaliana* encodes a MADS-box and is expressed in petals and stamens. *Cell* 68:683–697
- Jena S, Jayaprakasha GK, Singh P, Sakariah KK (2002) Chemistry and biochemistry of (–) hydroxy citric acid. *J Agric Food Chem* 50:10–22
- Jofuku KD, den Boer Bart GW, Montagu MV, Okamura JK (1994) Control of *Arabidopsis* flower and seed development by the homeotic gene APETALA2. *Plant Cell* 6:1211–1225
- Joseph KS, Murthy HN (2015) Sexual system of *Garcinia indica* Choisy: geographic variation in trioecy and sexual dimorphism in floral traits. *Plant Syst Evol* 301:1065–1071
- Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T et al (2008) KEGG for linking genomes to life and the environment. *Nucleic Acids Res* 36:480–484
- Kater MM, Franken J, Carney KJ, Colombo L, Angenent GC (2001) Sex determination in the monoecious species cucumber is confined to specific floral whorls. *Plant Cell* 13:481–493
- Kaur R, Chattopadhyay SK, Tandon S, Sharma S (2012) Large scale extraction of the fruits of *Garcinia indica* for the isolation of new and known polyisoprenylated benzophenone derivatives. *Ind Crop Prod* 37:420–426
- Liljgren SJ, Ferrandiz C, Alvarez-Buylla ER, Pelaz S, Yanofsky MF (1998) MADS-box genes involved in fruit dehiscence. *Flower News Lett* 25:9–19
- Litt A, Kramer EM (2010) The ABC model and the diversification of floral organ identity. *Semin Cell Dev Biol* 21:129–137
- Lorrain S, Vaillau F, Balague C, Roby D (2003) Lesion mimic mutants: keys for deciphering cell death and defense pathways in plants. *Trends Plant Sci* 8:263–271

- Ma H, Yanofsky MF, Meyerowitz EM (1991) AGL1-AGL6, an *Arabidopsis* gene family with similarity to floral homeotic and transcription factor genes. *Genes Dev* 5:484–495
- Malik SK, Chaudhary R, Abraham Z (2005) Desiccation-freezing sensitivity and longevity in seeds of *Garcinia indica*, *G. combogia* and *G. xanthochymus*. *Seed Sci Technol* 33:723–732
- Matsunaga S (2006) Sex chromosome-linked genes in plants. *Genes Genet Syst* 81:219
- Miko I (2008) Sex chromosomes and sex determination. *Nat Educ* 1:108
- Nayak CA, Rastogi NK, Raghavarao KSMS (2010) Bioactive constituents present in *Garcinia Indica* Choisy and its potential food applications: a review. *Int J Food Prop* 13:441–453
- Noavaes E, Drost DR, Farmerie WG, Pappas GJ, Grattapaglia D, Sederoff RR, Kirst M (2008) High-throughput gene and SNP discovery in *Eucalyptus grandis*, an uncharacterized genome. *BMC Genom* 9:312
- Paenicova L, de Folter S, Kieffer M, Horner DS, Favalli C, Busscher J, Cook HE, Ingram RM, Kater MM, Davies B et al (2003) Molecular and phylogenetic analyses of the complete MADS-box transcription factor family in *Arabidopsis*: new openings to the MADS world. *Plant Cell* 15:1538–1551
- Pelaz S, Ditta GS, Baumann E, Wisman E, Yanofsky MF (2000) B and C floral organ identity functions require *SEPALLATA* MADS-box genes. *Nature* 405:200–203
- Perl-Treves R (1999) Male to female conversion along the cucumber shoot: approaches to studying sex genes and floral development in *Cucumis sativus*. In: Ainsworth C (ed) Sex determination in plants. Bios Scientific Publishers, Oxford, pp 189–216
- Pnueli L, Hareven D, Broday L, Hurwitz C, Lifschitz E (1994) The TM5 MADS box gene mediates organ differentiation in the three inner whorls of tomato flowers. *Plant Cell* 6:175–186
- Preuss HG, Bagchi D, Bagchi M, Rao CVS, Satyanarayana S, Dey DK (2004) Efficiency of novel natural extract of (–) hydroxycitric acid (HCA-SX) and a combination of HCA-SX niacin bound chromium and *Gymnema sylvestre* extract in weight management in human volunteers: a pilot study. *Nutr Res* 24:55–58
- Rady MM, El-Yazal MAS (2013) Response of ‘Anna’ apple dormant buds and carbohydrate metabolism during floral bud break to onion extract. *Sci Hortic* 155:78–84
- Reddy SY, Prabhakar JV (1994) Cocoa butter extenders from kokum (*Garcinia indica*) and phulwara (*Madhuca butyraceae*) butter. *J Am Oil Chem Soc* 71:217–219
- Ruonala R, Rinne PLH, Baghour M, Moritz T, Tuominen H, Kangasjärvi J (2006) Transitions in the functioning of the shoot apical meristem in birch (*Betula pendula*) involve ethylene. *Plant J* 46:628–640
- Savidge B, Rounsley SD, Yanofsky MF (1995) Temporal relationships between the transcription of two *Arabidopsis* MADS box genes and the floral organ identity genes. *Plant Cell* 7:721–733
- Shara M, Ohia SE, Schmidt RE, Yasmin T, Zerdetta-Smith A, Kincaid A, Bagchi M, Chatterjee A, Bagachi D, Stohs SJ (2004) Physiochemical properties of a novel (–) hydroxycitric acid extract and its effect on body weight selected organ weight hepatic lipid peroxidation and DNA fragmentation hematology and clinical chemistry and histopathological changes over a period of 90 days. *Mol Cell Biochem* 260:171–186
- Shetty BV, Kaveriappa KM (2001) An arboretum of endemic plants of Western Ghats at Mangalore University campus, Karnataka, India. *Zoos Print J* 16:431–438
- Shifriss O (1956) Sex instability in ricinus. *Genetics* 41:265–280
- Steven DR, Gary SD, Martin FY (1995) Diverse roles for MADS box genes in arabidopsis development. *Plant Cell* 7:1259–1269
- Tan M, Xue J, Wang L, Huang J, Fu C, Yan X (2016) Transcriptomic analysis for different sex types of *Ricinus communis* L. during development from apical buds to inflorescences by digital gene expression profiling. *Front Plant Sci* 6:1208
- Tanurdzic M, Banks JA (2004) Sex-determining mechanisms in land plants. *Plant Cell* 16:61–71
- Thatte KS, Deodhar MA (2012) Study of flowering behavior and sex determination in *Garcinia indica* (Thomas-Du Pettite) Choisy by means of molecular markers. *Biotechnology* 11:232–237
- Vaddepalli P, Scholz S, Schneitz K (2015) Pattern formation during early floral development. *Curr Opin Genet Dev* 32:16–23
- Vining KJ, Romanel E, Jones RC, Klocko A, Alves-Ferreira M, Hefer CA, Amarasinghe V, Dharmawardhana P, Naithani S, Ranik M et al (2015) The floral transcriptome of *Eucalyptus grandis*. *New Phytol* 206:1406–1422
- Wang YQ, Shen JK, Berglund T, Ohlsson BA, Tang XF, Zhou ZK, Wu RY, Zhou XH, Chen JN (2010) Analysis of expressed sequence tags from Ginkgo mature foliage in China. *Tree Genet Genomes* 6:357–365
- Watson JA, Fang M, Lowenstein JM (1969) Tricarbolylate and hydrooxycitrate substrate and inhibitors of ATP: citrate oxaloacetate lyase. *Arch Biochem Biophys* 135:209–217
- Wei MM, Song MZ, Fan SL, Yu SX (2013) Transcriptomic analysis of differentially expressed genes during anther development in genetic male sterile and wild type cotton by digital gene-expression profiling. *BMC Genom* 14:97
- Xu G, Huang J, Yang Y, Yao Y (2016) Transcriptome analysis of flower sex differentiation in *Jatropha curcas* L. using RNA sequencing. *PLoS One* 11:e0145613
- Yamasaki S, Fujii N, Takahashi H (2000) The ethylene-regulated expression of CS-ETR2 and CS-ERS genes in cucumber plants and their possible involvement with sex expression in flowers. *Plant Cell Physiol* 41:608–616
- Yanofsky MF, Ma H, Bowman JL, Drews GN, Feldmann KA, Meyerowitz EM (1990) The protein encoded by the *Arabidopsis* homeotic gene *AGAMOUS* resembles transcription factors. *Nature* 346:35–39
- Ye D, Oliveira M, Veuskens J, Truong AT, Brown S, Mouras A, Negru-tiu I (1991) Sex determination in the dioecious *Melandrium*. The X/Y chromosome system allows complementary cloning strategies. *Plant Sci* 80:93–106
- Yin TJ, Quinn JA (1995) Tests of a mechanistic model of one hormone regulating both sexes in *Cucumis sativus* (Cucurbitaceae). *Am J Bot* 82:1537–1546
- Zeng F, Yu Q, Hou S, Moore PH, Alam M, Ming R (2014) Features of transcriptome in trioecious papaya revealed by a large-scale sequencing of ESTs and comparative analysis in higher plants. *Plant OMICS* 7:450–460
- Zhang X, Zhao L, Larson-Rabin Z, Li D, Guo Z (2012) De novo sequencing and characterization of the floral transcriptome of *Dendrocalamus latiflorus* (Poaceae: Bambusoideae). *PLoS One* 7:e42082
- Zhang J, Wu K, Zeng S, Teixeira A, Zhao X, Tian C, Xia H, Duan J (2013) Transcriptome analysis of *Cymbidium sinense* and its application to the identification of genes associated with floral development. *BMC Genom* 14:279

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.