View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by UTL Repository

Plant Mol Biol (2017) 93:151–170
DOI 10.1007/s11103-016-0553-9

Deep analysis of wild *Vitis* flower transcriptome reveals unexplored genome regions associated with sex specification

Miguel Jesus Nunes Ramos[1] · João Lucas Coito[1] · Joana Fino[2] · Jorge Cunha[3,4] · Helena Silva[5] · Patrícia Gomes de Almeida[6] · Maria Manuela Ribeiro Costa[5] · Sara Amâncio[1] · Octávio S. Paulo[2] · Margarida Rocheta[1]

Received: 29 March 2016 / Accepted: 15 October 2016 / Published online: 24 October 2016
© Springer Science+Business Media Dordrecht 2016

Abstract

Key message **RNA-seq of *Vitis* during early stages of bud development, in male, female and hermaphrodite flowers, identified new *loci* outside of annotated gene models, suggesting their involvement in sex establishment.**

Abstract The molecular mechanisms responsible for flower sex specification remain unclear for most plant species. In the case of *V. vinifera* ssp. *vinifera*, it is not fully understood what determines hermaphroditism in the domesticated subspecies and male or female flowers in wild dioecious relatives (*Vitis vinifera* ssp. *sylvestris*). Here, we describe a de novo assembly of the transcriptome of three flower developmental stages from the three *Vitis vinifera* flower types. The validation of de novo assembly showed a correlation of 0.825. The main goals of this work were the identification of *V. v. sylvestris* exclusive transcripts and the characterization of differential gene expression during flower development. RNA from several flower developmental stages was used previously to generate Illumina sequence reads. Through a sequential de novo assembly strategy one comprehensive transcriptome comprising 95,516 non-redundant transcripts was assembled. From this dataset 81,064 transcripts were annotated to *V. v. vinifera* reference transcriptome and 11,084 were annotated against *V. v. vinifera* reference genome. Moreover, we found 3368 transcripts that could not be mapped to *Vitis* reference genome. From all the non-redundant transcripts that were assembled, bioinformatics analysis identified 133 specific of *V. v. sylvestris* and 516 transcripts differentially expressed among the three flower types. The detection of transcription from areas of the genome not currently annotated suggests active transcription of previously unannotated genomic *loci* during early stages of bud development.

Keywords *Vitis vinifera* ssp. *sylvestris* · RNA-seq · De novo assembly · Differential gene expression · Sex *loci*

Miguel Jesus Nunes Ramos and João Lucas Coito have contributed equally to this work.

Electronic supplementary material The online version of this article (doi:10.1007/s11103-016-0553-9) contains supplementary material, which is available to authorized users.

✉ Margarida Rocheta
rocheta@isa.ulisboa.pt

[1] Universidade de Lisboa, Instituto Superior de Agronomia, LEAF, Linking Landscape, Environment, Agriculture and Food, Tapada da Ajuda, 1359-017 Lisboa, Portugal

[2] Computational Biology and Population Genomics Group, Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

[3] Instituto Nacional de Investigação Agrária e Veterinária, Quinta d'Almoinha, Dois Portos, Portugal

[4] Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal

[5] BioSystems & Integrative Sciences Institute (BioISI), Plant Functional Biology Centre, University of Minho, Campus de Gualtar, 4710-057 Braga, Portugal

[6] Development and Evolutionary Morphogenesis, Centre for Ecology, Evolution and Environmental Change, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

Introduction

Transcriptome analysis is an important tool for the characterization and understanding of the molecular basis of phenotypic diversity between individuals. Thus, among the

complex genetic mechanisms controlling plasticity, sex differentiation can be interpreted from the dynamics of transcriptome. High throughput RNA-sequencing analysis of highly complex transcriptomes, can now be applied to any tissue despite the developmental stage (Wang et al. 2009).

Vitis vinifera ssp. *vinifera* is the domesticated grapevine, a hermaphrodite plant and one of the most valuable cultivated fruit crop in the world, with 67 tons of grapes produced in 2012 (http://faostat.fao.org/). On the other hand, *Vitis* wild relative, *Vitis vinifera* ssp. *sylvestris*, has separate sexes, and is a repository of genetic diversity, essential for the maintenance of genetic variability and to limit genetic erosion of the domesticated variety (García and Revilla 2013). As distinctive feature male flowers present a reduced pistil without style or stigma and female flowers have reflexed stamens with infertile pollen. The development of a flower requires the synchronization of gene activities, environmental conditions and hormones (de Montaigu et al. 2010; Huijser and Schmid 2011). The decision to make a flower encloses sex-specific signals that activate or repress male- or female-specific pathways leading to the differentiation of reproductive organs. Among these signals, plant hormones play an important role. They are involved in various aspects of reproductive organ development in hermaphrodite plants but also in plants with unisexual flowering, such as monoecious and dioecious plants (Lebel-Hardenack and Grant 1997).

Several studies have shown the effects of hormones on sexual specification (Golenberg and West 2013; Manzano et al. 2014; Sekhar and Sawhney 1991). Interestingly, hormone effects are plant species dependent; there are no merely masculinizing or feminizing hormones. For instance, gibberellic acid has a masculinizing effect in cucumber and *Asparagus officinalis* (Fuchs et al. 1977; Lazarte and Garrison 1980) whereas a feminizing effect is observed in maize (Dellaporta and Calderon-Urrea 1994). Also, auxin has a feminizing effect in *Opuntia stenopetala* (Orozco-Arroyo et al. 2012) but a masculinizing effect in *Mercurialis annua* (Durand and Durand 1991), while ethylene has a feminizing role in cucurbit flower development (Boualem et al. 2008) but a masculinizing effect in watermelon (Manzano et al. 2014). Other plant hormones like abscisic acid that have a feminizing effect in *Solanum carolinense* and jasmonate with a masculinizing action in maize also contribute to flower sexual differentiation (Browse 2009; Hartwig et al. 2011). In addition, cytokinins have been related to feminized the phenotype in *Asparagus officinalis* (Lazarte and Garrison 1980) and *Vitis vinifera* (Negi and Olmo 1966).

Additionally, previous studies identify a putative *locus* for flower sex in *V. v. vinifera* in chromosome 2 (chr2) (Dalbó et al. 2000; Lowe and Walker 2006; Marguerit et al. 2009), located in a region that comprise a few genes that may play a role in hormone metabolism (Battilana et al. 2013). Other authors using BAC clones established a physical map of 143 kb (Fechter et al. 2012). Sequence and annotation of this region identified 11 genes possibly responsible for flower sex determination. The most promising of these genes, an *ADENINE PHOSPHORIBOSYL TRANSFERASE* (*APT*), became a marker able to distinguish the female allele from the male/hermaphrodite alleles (Fechter et al. 2012). More recently a new study gave a more detailed insight regarding a possible sex *locus* in chr2 of *V. v. vinifera* suggested a region of 152 kb that comprising the previously referred 143 kb region and included several genes with XY type polymorphism (Picq et al. 2014). All these studies reinforce the existence of a *locus* in chr2 responsible for sex determination in *Vitis v. vinifera*. However, in the most recent *Vitis* annotation (12X_v2.1) (http://genomes.cribi.unipd.it/grape/), some genes in the 152 kb region are absent from the annotation or are located in the chromosome Unknown.

Our aim is to map and annotate unexplored transcripts and to get a list of genes with differential expression all along early flower development of both *Vitis vinifera* subspecies through a new assembly of *Vitis* transcriptome. We reason that some of the undiscovered transcripts may play important roles in *Vitis* sex differentiation. To achieve our goal we de novo assembled the Illumina RNA-seq raw data of flower *Vitis* transcriptome previous obtained (Ramos et al. 2014). The assembly of the transcriptome data of three different stages of flower development from each of the three *Vitis* flower types (hermaphrodite, male and female) produced a non-negligible fraction of unmapped contigs, whose sequences are generally neglected in favor of the mapped contigs despite potentially containing useful information. This analysis of transcriptome dynamics takes advantage of the of RNA-seq approach to discover transcribed *loci* outside of annotated gene models, as well as genes differentially expressed during early *Vitis* flower development.

Materials and methods

Raw data source and strategy of sample analysis

Previous transcriptome sequencing of flower buds in early in early developmental stages was made through RNA-seq of flower buds from *Vitis vinifera sylvestris* female and male plants and from the hermaphrodite *Vitis vinifera vinifera* cv. Touriga Nacional [TN] as previously described (Ramos et al. 2014). Floral buds were collected during the month of April (2012) from a single plant of each flower type and classified according to its developmental stage, B, D, G and H (Baggiolini 1952), and treated separately, adding to a total of twelve samples (Ramos et al. 2014). In order to

evaluate data consistency, two Illumina independent runs were carried out on each sample. The raw data generated in the previous work was used to perform a new assembly. However, in this new assembly, we decided to deep analyze three stages B, D and H. The exclusion of stage G from the analysis was due an unusual lower number of reads in the male flower, comparatively to the other stages and flower types, which could mask the results of differential expression gene analysis, a constrain not present in the previous analysis (Ramos et al. 2014).

Transcriptome de novo assembly and annotation

cDNA libraries construction and transcriptome sequencing were performed by BaseClear, B.V, Netherlands (Ramos et al. 2014) and quality assessment was performed using SeqtrimNEXT (Falgueras et al. 2010). The input configuration was set to remove indeterminations, poly-A-tails, Illumina adapters, contaminant sequences, vector residues, low quality zones and repetitive/complex regions (Online Resource 1).

The nine datasets were assembled in a unique transcriptome in order to obtain a more robust result than with individual assemblies using the multiple k-mer method. Assemblies with k-mers of 21, 25, 29 and 31 were performed with Velvet package (Zerbino 2010): reads prepared with velveth, with options *-fastq* and *-short* were assembled with velvetg with option *-read_trkg*. Corrections were applied with Oases (Schulz et al. 2012) using the ideal cut-off (0.2) suggested by Velveth Optimizer (Victorian Bioinformatics Consortium). The final merge was performed with the same programs with a k-mer of 31, as suggested by Velvet Optimizer as the ideal k-mer. To reduce redundancy and group transcripts, our library was submitted to CD-HIT-EST algorithm (Li and Godzik 2006). CD-HIT-EST works by clustering sequences according to their similarity and outputs the larger sequence of each contig, although some redundancy persisted after CD-HIT-EST. A pipeline based on BLAST was then used to deal with this issue. A blast of our transcriptomic library against itself was conducted with blastn version 2.2.28 (Altschul et al. 1997) with an e-value limit of 1e-5. An in-house python script was used to remove redundant and small contigs, considering 90% of identity which is available upon request. Finally, contigs were matched against the NCBI databases to exclude potential contaminations (e.g. microorganisms).

In order to evaluate the assembly, in addition to N50 calculation, we ran RSEM-EVAL from DETONATE (Li et al. 2014) with the same raw data files and with the non-redundant fasta sequences previously assembled, considering an average of 100 bp per read.

Genome, transcriptome and annotation versions 12X_v2.1 of *Vitis vinifera vinifera* cv. PN40024 were downloaded from Grape Genome Database on CRIBI (http://genomes.cribi.unipd.it/grape/). Non-redundant libraries were blasted on *Vitis v. vinifera* reference transcriptome (Altschul et al. 1997), with an e-value limit of 1e-5. A second in-house python script to map and rename each aligned contig was developed. Reads were classified as exonic if they mapped to an annotated exon; intronic, if they mapped within the 5′ and 3′ boundaries of a given transcript, but outside annotated exons; or intergenic, if they aligned outside known boundaries of annotated genes. In order to get accurate data only contigs with length equal or higher to 150 bp (Celaj et al. 2014) were kept. The remaining sequences were discarded, as they reduce our ability to identify meaningful confident sequence matches. For each sample its expression was determined by calculating its Reads per Kilobase transcript per Million reads (RPKM) for each library described below.

The non-aligned contigs were blasted to *Vitis v. vinifera* reference genome with the same parameters described above. To map contigs to the reference genome, three different approaches were applied. On the first approach only the contigs that mapped to a gene were considered; on the second approach, contigs mapping within 1000 bp upstream or downstream of a gene were considered as being part of that gene; finally, on the third approach, the flanking gene regions were increased to 4000 bp upstream or downstream of the annotated gene (Zenoni et al. 2010). Sequences and its flanking regions were manually inspected and the second approach was considered the best. Reads that mapped inside this region were assigned as unannotated exons or untranslated regions (UTRs), while reads that mapped outside these regions were denominated as putatively novel transcripts. Additionally, only contigs with RPKM equal or higher than one were kept.

Finally, a third python script was used to map and rename sequences that aligned to the genome. All in-house programs are available upon request.

Moreover, cis association on adjacent genes was also studied on all sequences that mapped successfully to the reference transcriptome and genome. Sequences associated to unknown or X-random chromosomes were excluded, since it was not possible to correctly infer the real localization of the genes in the chromosome. Moreover, in order to reduce false positive associations, only transcripts with expression correlations higher than 0.8 were considered.

In addition to *Vitis* reference genome, the unmapped sequences were blasted against seven databases present in Ensemble (*Arabidopsis thaliana, Medicago truncatula, Oryza sativa, Populus trichocarpa, Solanum lycopersicum*, *Solanum tuberosum* and *Zea mays* - http://plants.ensembl.org) considering an e-value limit of 1e-5. The sequences that did not map with these databases were considered putative contaminations or novel genes. A general blast (e-value

limit of 1e-5) against all NCBI database was performed and sequences from bacteria, fungi and similar organisms were discarded as contaminants.

Contigs that did not map in databases were split into sequences of 100 bp and each sequence treated separately. These fragmented sequences were blasted against the NCBI nt database. This strategy allowed the identification of chimeric sequences that could not be mapped previously. Additionally, a self-blast was applied to these 100 bp sequences in order to confirm the similarity to other unmapped sequences. For these two blasts an e-value cutoff of 1e-5 was considered. The remaining sequences were considered as "completely unknown" and were submitted to Repbase (http://www.girinst.org/), with default parameters, in order to predict if these sequences belong to mobile elements.

Transcriptome de novo assembly validation and gene expression reproducibility through RT-qPCR

To determine the accuracy of the de novo assembly, a correlation was established between the de novo assembly and the previous one (Ramos et al. 2014) (Online Resource 2). For gene validation through RT-qPCR, four genes with annotation *ABNORMAL FLORAL ORGANS* (VIT_202s0154g00070), *FLAVIN-CONTAININGMONOOXYGENASE-LIKE*(VIT_202s0154g00170), *FLAVIN-CONTAINING MONOOXYGENASE LIKE-3* (VIT_202s0154g00180) and *AGAMOUS-LIKE 19* (*AGL 19*; VIT_202s0025g04650) related to flower development present in chromosome 2 were used (Online Resources 3, 4). Additionally, some of the completely unknown transcripts (G_30546; G_80794; G_89407) were validated in order to determine if their expression is maintained in new biological samples collected in 2015 (Online Resources 3, 4).

The cDNA concentration in RT-qPCR was achieved through a serial of decimal dilutions from 2.1 to 0.021 µg of cDNA. Amplification reactions were performed in triplicates of 20 µL containing 5 µL of master mix (SsoFast_EvaGreen Supermix, Bio-Rad, Hercules, CA), 0.4 µM of specific primers, 0.21 µg of cDNA and autoclaved water. The following program was applied: initial polymerase activation, 95 °C, 2 min; then 40 cycles at 95 °C, 15 s (denaturation); 57 °C, 30 s (annealing); 76 °C, 30 s (extension) with a single fluorescence reading taken at the end of each cycle. Each run was completed with a melting curve analysis to confirm the specificity of amplification and the lack of primer dimmers. To confirm amplicon size, products were run on 1.7 % (w/v) agarose gels.

RT-qPCR runs were perform, for each gene, with a series of decimal dilutions of a precisely calculated number of plasmid copies (3000–3 pg) to create a calibration ruler. Cqs (threshold cycles) obtained from RNA samples was matched against the calibration ruler to estimate the copy number of each transcript on samples. The absolute copy number was calculated using the following formula: Copy number = C × NA/M; where Copy number, number of molecules/µL contained in the purified cDNA; C, concentration of the purified cDNA (g/µL); M, the molecular weight of the cDNA gene fragment; NA, Avogadro's number = 6.023 × 1023 molecules/mole. The coefficient of correlation (r) between de novo assembly samples and RT-qPCR were also calculated.

Differentially expressed genes (DEGs)

The number of RPKM was quantified and assumed as an expression value. RPKMs were calculated individually for each sample, matching the cleaned reads library from each sample to the total assembly contigs. This assessment was achieved using RSEM package (Kelley et al. 2012). Contigs file was formatted using rsem-prepare-reference algorithm and quantification was performed with rsem-calculate-expression. A mismatch of three nucleotides was permitted (option—*bowtie-n 3*). The analysis of differential gene expression was also performed with RSEM package (Li and Dewey 2011). The data matrix was exported from the rsem-calculate-expression output, using rsem-generate-data-matrix algorithm. The analysis itself was performed by rsem-run-ebseq algorithm.

The transcripts falsely detected as differentially expressed were corrected using rsem-control-fdr algorithm with a false discovery rate (FDR) of 0.05. Fold change was manually filtered according its $\log_2$, in Excel (Microsoft). Transcripts with $\log_2$ between −1 and 1 were considered false positives and discarded. Additionally, in the absence of a control sample that could be used as a reference for all samples, DEGs in female and male flowers were determined using the hermaphrodite as a reference, while the male transcriptome was used as a reference when female and male were compared. We opted for the use of "up-expressed" and "down-expressed" for differential gene expression when comparing two flower transcriptomes. In addition to the previous filters applied, for the expression analysis we only considered contigs with RPKM equal or higher than 1 to be expressed.

Results and discussion

Transcriptome de novo assembly and validation

One of the objectives of this work was to identify exclusive and unidentified transcripts sequences with a putative role in sex specification in *Vitis vinifera sylvestris* by comparison with *Vitis vinifera vinifera* (Touriga Nacional) and the reference genome (*Vitis vinifera vinifera* cv. PN40024). To achieve this purpose, a de novo assembly of *Vitis vinifera* transcripts was performed (described in "Materials and methods" section).

From our previous study (Ramos et al. 2014), we used 368,898,213 sequence reads of 45–50 bp covering 15–17 Gb of sequence data. Each flower developmental stage was represented by an average of 30 million reads (Online Resource 1). The trimming/filtering with SeqtrimNEXT removed 6% of the initial reads, and the remaining 345,641,042 reads from the nine samples were pooled and assembled, resulting in 231,342 contigs (Table 1). After removing redundancy, only 41% (95,516) of the contigs remained, with an average size of 390 bp. The N50 statistic of 887 bp was used in further analysis (Table 1). N50 is a quality measure value, which describes the point of hall mass of our data, meaning that 50% of the sequenced bases were assembled into contigs with this length or higher (Zerbino and Birney 2008). Also, the program RSEM-EVAL from DETONATE (Li et al. 2014) was used to evaluate the reliability of the assembly. The assembly was scored as—754,768,755, which is not a comparable value to other studies since RSEM-EVAL score is dependent of the number of reads found on a RNA-seq library (Li et al. 2014). In order to get a clear support to this assembly, the "contig impact score" was evaluated to each contig. This parameter is an indication that the contig is supported by the original reads. On average, the contigs produced on this assembly were scored as 1487, which, by being a positive value, provides confidence on this assembly (Li et al. 2014).

To determine the accuracy of the de novo assembly, a correlation with the previous RNA-seq was established (Ramos et al. 2014). Data showed a correlation of r = 0.825, which guaranteed the quality of data comparison (Online Resource 2). However, the expression of some transcripts is relatively different in the comparison of assemblies. This situation results from improvements done to the reference transcriptome. Initially, when all reads were mapped against the reference genome, there was only one isoform described per gene (version 12X_v1) (Ramos et al. 2014). As a consequence each gene was evaluated only by one RPKM value. The most recent transcriptome annotation (12X_v2.1) has multiple isoforms assigned to each gene and allows assigning a RPKM value to each isoform. The match between both assemblies translates differences in the expression of different isoforms of the same gene. Despite these specific situations, the correlation between both assemblies is high and positive, proving that the de novo assembly is consistent with the previous one (Ramos et al. 2014).

Additionally, gene expression validation was performed through RT-qPCR with seven transcripts and the correlation between RT-qPCR and a de novo assembly shows and r between 0.989 and 0.831 which confirm data precision (Online Resource 4).

Table 1 New assembly parameters and raw results of nine transcriptomes, developmental stages B, D and H, from *V. v. sylvestris* flowers (male and female) and *V. v. vinifera* (hermaphrodite)

Assembly parameters	
Optimal k-mer	31
Optimal cut-off	0.2
Obtained contigs	
Obtained	231,342
After CD-HITS-EST	177,105
Non redundant	95,516
Blast to *V. v. vinifera* transcriptome (cDNA)	
Mapped	
Total	81,064
One hit	28,466
More hits	52,598
Unmapped	14,452
Blast to *V. v. vinifera* genome (DNA)[a]	
Mapped	
Total	11,084
One hit	8161
More hits	2923
Unmapped	3368

[a]Contigs that have no match with the *V. v. vinifera* transcriptome have been mapped against *V. v. vinifera* genome version 12X_v2.1

Mapping and annotation to *Vitis* reference transcriptome and genome

The mapping of non-redundant contigs against the *Vitis vinifera vinifera* reference transcriptome allowed us to predict the role of 85% of the contigs that matched with *Vitis* reference transcriptome (81,064) (Table 1). The contigs that did not map to the reference transcriptome were then blasted to the reference genome and 12% of the total was successfully mapped (11,084) (Table 1) in intronic or intergenic regions. The remaining 3368 contigs did not map to the *Vitis* available genome (Table 1). When redundancy was removed former numbers decreased but the highest proportion of contigs (72%) map to previously annotated exons of the *Vitis* transcriptome with 7% of the reads aligning to intronic regions or UTRs, 18% aligning outside of the boundaries of annotated genes, i.e. intergenic areas and 3% are considered as unmapped sequences (Fig. 1). This could be due to the active transcription of not yet annotated genes or, alternatively, to the genomic differences (such as polymorphisms) between the samples and the reference genome or even to the potential erroneous base calling by the sequencing technology (Oshlack et al. 2010). Additionally, such sequences might also represent non-coding RNAs, repetitive sequences such as mobile elements or miRNAs precursor (pre-miRNA) playing a role in *Vitis* flower development.

In order to predict possible loci regulated by cis associations, the expression of sequences that unambiguously match to the *Vitis* reference genome and transcriptome was

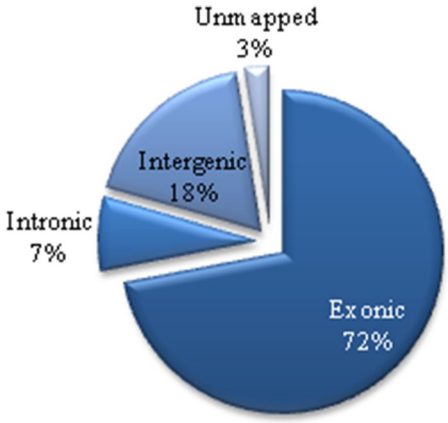

Fig. 1 Overview of sequences transcribed in the genomes in three developmental *V. vinifera* flower stages. The *Vitis* data used to perform annotation was 12X_v2.1. Redundancy has been removed and only contigs with sequence length ≥150 bp were considered

analyzed. To perform this analysis some sequences were discarded: those that did not map to the reference genome, since it was not possible to predict which transcripts were adjacent; and sequences that mapped to the unknown or random chromosomes, as those are virtual chromosomes where the position of transcripts was arbitrarily assigned. Overall, from the 21,745 considered transcripts, there are 688 loci where two or more adjacent transcripts show strong positive (>0.8) co-expression, comprehending a total of 1479 transcripts (Table 2). Only one locus was composed by six adjacent transcripts putatively related, located on the chromosome 12 (chr12:1,628,012–1,630,325). These transcripts are short (157–253 bps) and expressed only in hermaphrodite flowers at developmental stages D and H. The first transcript found in this cluster (chr12:1,628,012–1,628,238) was successfully blasted to a non-coding RNA, predicted by automated computational analysis, but with no additional information regarding its function (NCBI accession, XR_786969.1). For the remaining five transcripts on this cluster, this is the first time that they are reported as being actively transcribed, still with unknown function.

Table 2 Summary of putative cis-associated transcripts. Cluster size, represents the number of transcripts that form the cluster; loci count, indicates the number of loci found with strong positive (>0.8) gene expression association and the total number of transcripts represent the total number of transcripts associated

	Loci count	Cluster size (no of associated transcripts)	Total number of transcripts (loci count × cluster size)
	607	2	1214
	64	3	192
	13	4	52
	3	5	15
	1	6	6
Total	688		1479

Expression analysis in exonic, intronic and intergenic regions

The number of transcripts that map to the exonic regions of *Vitis* annotated gene (*Vitis* reference transcriptome) increases during flower development in the three flower types (Fig. 2a). The female, male and hermaphrodite flower have an increase of expressed genes of 2, 1.9 and 3.8%, respectively, from early stage B to late stage H (Fig. 2a).

The number of the transcripts that align to intronic regions does not follow the same pattern as the exonic transcripts (Fig. 2a). The transcriptome of the female, male and hermaphrodite flower increase 19.4, 1.7 and 11.9% respectively, from early stage B to late stage H (Fig. 2a). However, the female flower shows a slight decrease in stage D, while at the same stage the male one presents an increase of expressed genes. As a whole, this dynamic highlight the requirement of the hermaphrodite flower type to has more active genes for sex specification in later flower developmental stages. Interestingly, there is increased gene activity in intronic regions. In the male flower, the levels of transcription in intronic regions do not always mirror the pattern of exonic regions across flower developmental stages (Fig. 2a). Our data point that changes in intronic expression directly measure changes in transcriptional activity, a fact reported in other works (Gaidatzis et al. 2015; Sultan et al. 2008). Additionally, it has been pointed that intronic expression are a stand-in for nascent transcription (Ameur et al. 2011; Hendriks et al. 2014). We do not rule out the hypothesis that this systematic study in early flower development mimics what has been reported for mammals development (Hogenesch and Ueda 2011; Mohawk et al. 2012) and in *Vitis* flower development there is a rhythmic balance between exonic and intronic gene expression, however, still unknown in its function.

The Venn diagram allows to compare the number of exclusive and shared genes between the three flower types, in each flower developmental stage (Fig. 2b). It is possible to observe that the number of expressed common genes to the three flower types increases during flower development. The number of exclusive transcripts in the female drops at stage D increasing in the stage H; the male shows a progressive decrease in exclusive transcripts as flower development proceeds; the number of genes exclusively expressed in hermaphrodites does not change significantly throughout flower development (Fig. 2b). At later stages (H) female and hermaphrodite flowers share the highest number of transcripts, which could be an indication that similar biological pathways are taking place. Notice that a lower

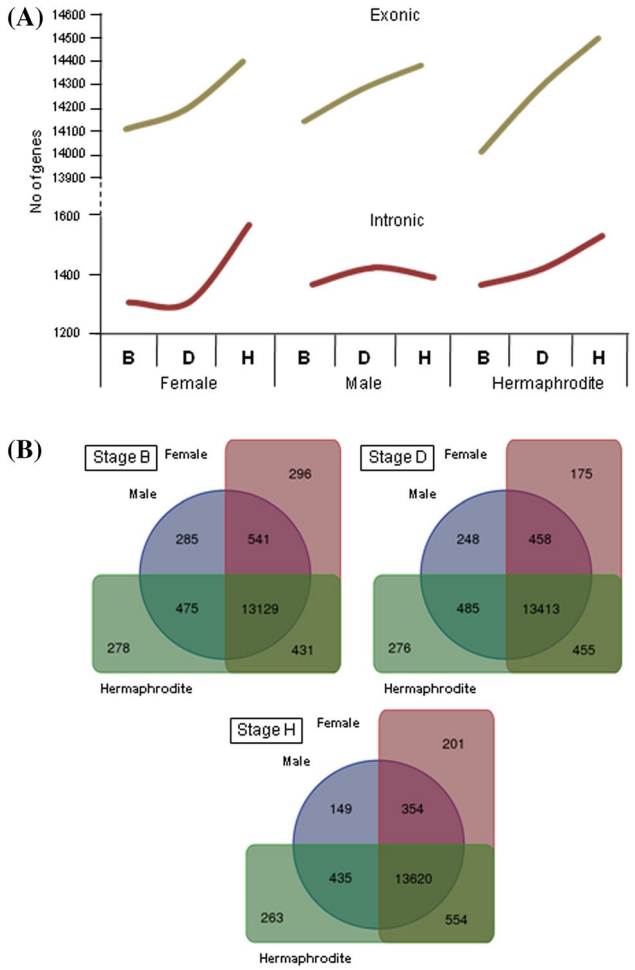

Fig. 2 Analysis of genes transcribed in the genomes that match with *V. v. vinifera* reference transcriptome and genome. **a** Graphical tendency of the number of sequences transcribed in exonic and intronic regions between grapevine flower developmental stages and flower types. **b** Venn diagrams with the number of genes expressed in each flower developmental stage (B, D and H). Redundancy has been removed and only contigs with RPKM ≥1 and sequence length ≥150 bp were considered

number of expressed genes were obtained when compared with our previous work (Ramos et al. 2014), mainly due to the straight filters applied to get an accurate analysis. Additionally, the new assembly allowed the identification of a significant number of contigs that map to intergenic regions (Fig. 3a). In general, an increase in the number of these unidentified expressed genes is observed throughout flower development for all flower types. However, the stage H is the one that presents more expressed genes in female and hermaphrodite transcriptome. The number of transcripts in male flower samples increases along development, particularly in the early stages of flowering; however, fewer transcripts from intergenic regions are present in the stage H when compared with the two other flower types (Fig. 3a). This result highlights the presence of polyadenylated transcripts from presumed intergenic regions with a different dynamic in male flower compared with the female and hermaphrodite. The occurrence of intergenic transcription point out for the possibility that there are more functional genes yet to be discovered. But we cannot exclude the possibility that these regions could drive expression in genes (Bondino and Valle 2009; Moghe et al. 2013) involved in flower development. However, it is not yet clear whether these transcripts represent novel non-coding or protein-coding genes, or the importance of their role during *Vitis* flower development.

The analysis of transcripts mapped in the intergenic regions that are shared or gender exclusive shows fluctuations between the developmental stages (Fig. 3b). Notice that the number of genes shared between female (*V. v. sylvestris*) and the hermaphrodite (*V. v. vinifera*) it is higher when compared with other flower types and increase across development which suggest a higher proximity between the development of these two types of flowers then between female and male (*V. v. sylvestris*).

One advantage of RNA-seq is the possibility to detect the expression of non-annotated genomic *loci*, allowing the detection of novel transcripts. Such areas, initially dubbed as transcribed "dark matter" (Ponting and Belgard 2010), have been the subject of increasing attention. Our results contribute to the increased recognition of the phenomenon of non-exonic, "dark matter" transcription (Kapranov and St Laurent 2012). The non-exonic transcription could be an effect of alternative splicing that lead to the inclusion of new exons not yet annotated.

The high proportion of non-exonic reads may also be explained by uncharacterized gene expression specific to *Vitis* flower development. Taken together, our data suggest that the development of a *Vitis* flower is associated with an increase in expression of intronic and intergenic regions. Such regions may represent novel exons, or genes whose inclusion into mature transcripts is induced during flower development. These genes would thus be previously undescribed in *Vitis* tissues, and are therefore not included in the *Vitis* transcriptome annotation.

Expression analysis in unexplored *Vitis* regions

The remaining contigs that did not map to *V. v. vinifera* reference transcriptome nor to the reference genome named "unmapped" were blasted against NCBI databases that include other works on *V. v. vinifera* besides the reference (http://genomes.cribi.unipd.it/grape/). Depending on the blast results, those contigs were divided in three classes: (I) mapped to other *Vitis* species, (II) homologous to other plant species genes or (III) "completely unknown" if they did not have homology with any sequence present in databases.

Moreover, considering that the existence of completely unknown sequences can result from chimeras, an artifact

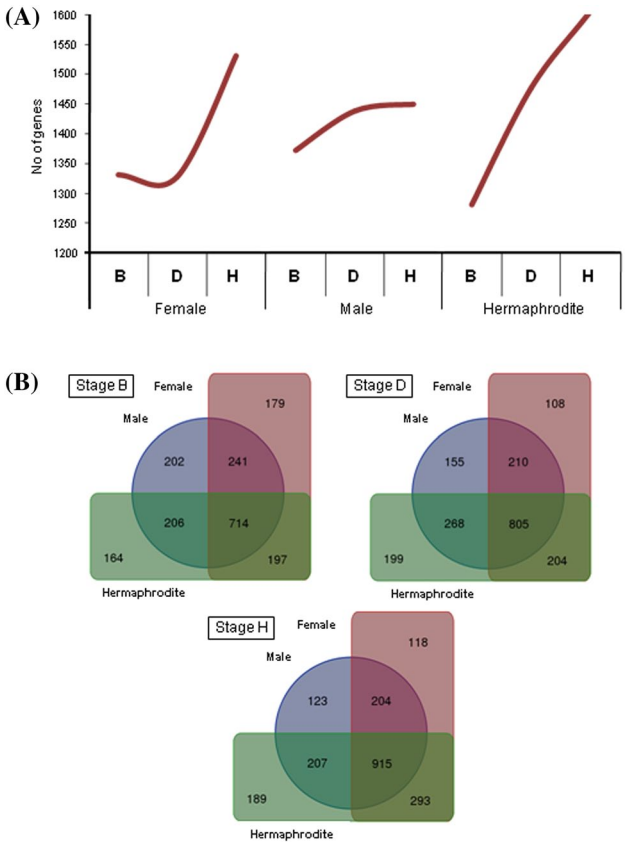

Fig. 3 Analysis of intergenic sequences transcribed in the genomes that match with *V. v. vinifera* reference genome. **a** Graphical tendency of the number of sequences transcribed between grapevine flower developmental stages and flower types. **b** Venn diagrams with the number of exclusive intergenic sequences, as well as, sequences shared by flower types, in each flower developmental stage (B, D and H). Only contigs with RPKM ≥1 and sequence length ≥150 bp were considered

that could lead to false sequences, the 1286 completely unknown transcripts were split into 100 bp, originating 3081 sequences that were treated individually and subjected to two different blastn approaches. A blastn was performed against all NCBI nt database, to predict sequences that had not map previously. On a second approach, the split sequences were blasted against themselves, in order to identify possible self and multi-gene chimeras. Using the first approach, 67 out of 1286 contigs had one or more sequences with positive results against NCBI nt database. This outcome suggests that some of the original completely unknown sequences had not been properly blasted, perhaps because they were too large. A blast of each short sequence against all the 3081 transcripts reveals situations where contigs were formed by the successive repetition of the same reads. This approach revealed eight putative chimeras (four self and other four multi-gene), but only three were found using both strategies simultaneously. A total of 72 identified chimeras were removed from further analysis, while 1214 completely unknown transcripts remained and were considered. Additionally, to reinforce that those "completely unknown" sequences are *Vitis* transcripts and not contaminants or artifacts created on the assembly process, three of them (G_30546; G_80794; G_89407) were amplified in cDNA samples from the three flower types collected in different years, using the same phenological developmental stage B (Online Resource 5). The amplification of each one of these transcripts resulted in a clear fragment, with the expected size in different flower types and sampling years. Overall, these results confirm that the completely unknown transcripts are not a result of chimeras.

The number of transcripts that mapped to *V. v. vinifera* and those that map against other plant species have a slight variation among flower types and flower developmental stages (Fig. 4a). Conversely, the number of contigs with no homology to any database drops during flower development in each gender (Fig. 4a).

Following the dynamics of these "completely unknown" transcripts shared between flower types during flower development shows a male flower with a lower percentage of exclusive transcripts as the development proceeds, while female and hermaphrodite flowers have an opposite behavior (Fig. 4b). In addition, the higher percentage of expressed genes shared between female and male flowers in stage B is a good indication that similar processes are taking place at this stage in both flower types. On the other hand, female and hermaphrodite flowers shared the highest percentage of expressed genes in the later flower developmental stage (stage H). A profile of these "completely unknown" transcripts (Da Silva et al. 2013) shows some clusters of promising transcripts that could have a role in sex specification (Fig. 5). Novel gene information from *Vitis* flower transcriptomes can be used for further gene expression studies, to elucidate and understanding molecular processes related with flower development and to gain a deeper functional annotation of *Vitis* genome and transcriptome.

DEGs analysis

Gene expression of specific genes may affect flower sex specification, not only due to tissue specificity but also to differential transcription levels. Therefore, we analyzed the differential gene expression between flower types along flower developmental stages. It is clear that the majority of genes with differential expression are found in the earlier stage (B) and in the later one (H) (Fig. 6). When male and female samples are compared with the hermaphrodite, the majority of genes differentially expressed at stage B are found in the male sample. At the same stage, the comparison between male and female transcriptomes shows female with less up-expressed

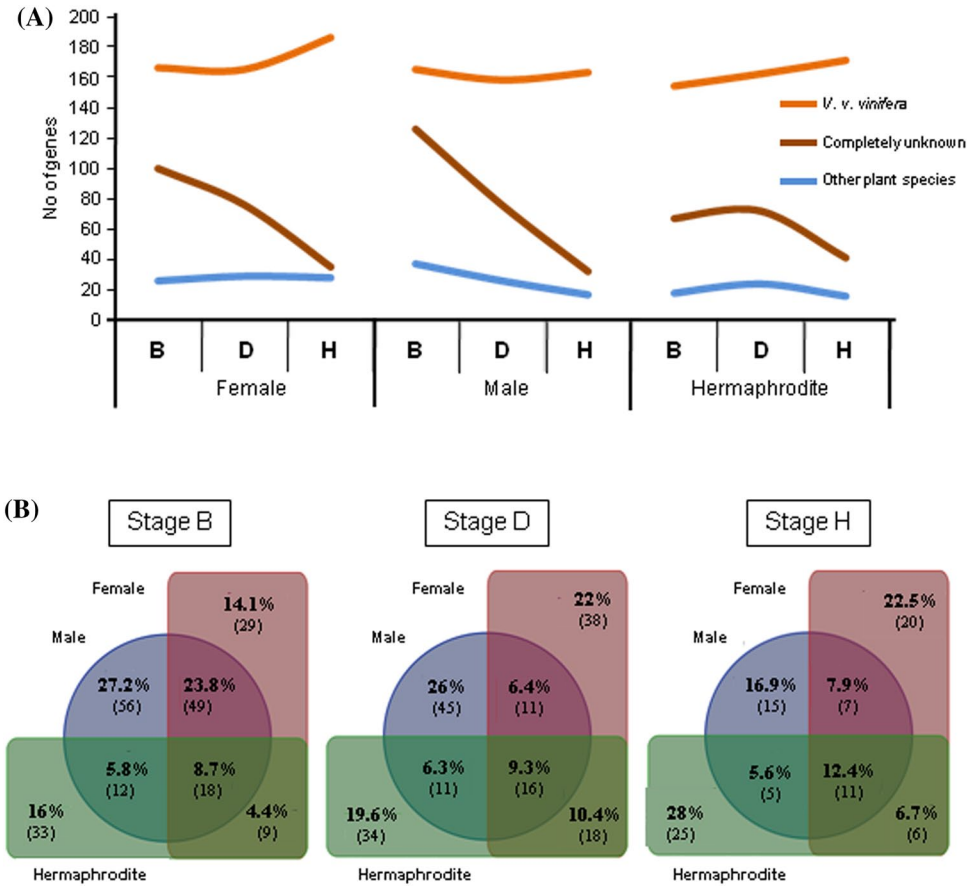

Fig. 4 Analysis of transcribed sequences that do not match with *Vitis* reference genome (12X_v2.1). **a** Number of transcripts that mapped with *V. v. vinifera* outside the reference genome annotation (*orange*); genes that mapped with other plant species (*blue*) and genes that not mapped against any known database (*brown*). **b** Venn diagrams with the percentage and number of "completely unknown" transcripts shared by flower types, in each flower developmental stage (B, D and H). Filters applied were: RPKM ≥1 and sequence length ≥150 bp

genes (Fig. 6). At stage D, the number of differentially expressed genes decreases in all samples while at developmental stage H, the highest number of DEGs was found in female when compared to the hermaphrodite (Fig. 6). At the same stage the number of up- and down-expressed genes is similar in male and hermaphrodite flower (Fig. 6). Taking the results as a whole an opposite dynamics between male and female flowers development is evident across flower development. Male flowers in the initial phase of flower development need more DEGs, while the female flower requires more DEGs in the last flower stage analyzed (stage H). This behavior translates differences between sexes and may include unknown mechanisms suppressing carpel development (early) in male flowers, or preventing the correct formation of male organs in female plants (late). The D stage where rudimentary leaves start to appear still collected in a rosette (Baggiolini 1952), transcription slowed in all three types of flowers for some reason not yet known. Our results are in agreement with a previous study applied to *Arabidopsis thaliana* (Ryan et al. 2015), where temporal gene expression was analyzed during different flowering phases. This work reinforces an extensive view of gene expression activities between flower types during developmental stages.

Differentially expressed genes assigned to gene functional categories

The data reveal that most DEGs belong to the Primary Metabolism category, followed by the Transport Overview category with second highest number of DEGs detected (Fig. 7; Online Resource 6). Nevertheless, when analyzing the percentage of number of genes relative to the total number of annotated genes in each category, the highest percentage of DEGs is found in the Cellular Metabolism, Secondary Metabolism and Response to Stimulus categories (Fig. 7). Notice that in the Development functional category, when comparing male and hermaphrodite flowers, male has the highest percentage of up-expressed genes in stage B, but then shows more down-expressed genes in the later developmental stages. In the female flower, the number of up-expressed genes is almost constant until stage H, where it shows the highest percentage. In this category no DEGs were observed between female and male transcriptome in early stage B while in development stage H the female flower has no down-expressed genes when compared with the male or hermaphrodite transcriptome (Fig. 7). At individual level we analyzed several genes, present in the Development functional category, differentially expressed between flower types. These genes are reported to be important and relevant

◀**Fig. 5** Expression profile of Unknown transcripts. Expression profile of completely unknown transcripts present in the three *Vitis* flower types in development stages B, D and H. The *color profile* translates the expression value measured in RPKM

in flower sex development. Among these genes, *VviSPLAYED* (Vvi*SYD* - VIT_205s0020g02000) is required for flower development and patterning (Wagner and Meyerowitz 2002). Its interaction with *LEAFY* and *SEPALLATA3* regulate the activation of *AP3* and *AG* during flower development (Wagner and Meyerowitz 2002). Additionally, *SYD* is also an activator of *WUS* and required for the maintenance of shoot apical meristem (Kwon et al. 2005; Wu et al. 2012). In our work *VviSPLAYED* is down-regulated in development stage B in male plants when compared with hermaphrodite plants in the same development stage. However, there are no obvious differences in gene expression in subsequent stages and phenotypically male plants bloom early than female or hermaphrodite plants. This might indicate a slight delay in male flower development at initial stages. Other gene is *Flowering Time Control* gene (*VviFCA* - VIT_205s0077g00710). *FCA* gene is a RNA-binding protein that promotes *microRNA172* gene activity which in turn represses *AP2* from the forth whorl during flower development (Aukerman and Sakai 2003; Jung et al. 2012). Our analysis show that male plants in development stage H display a down-expression of *VviFCA* when compared with female and hermaphrodite plants. This down-regulation of *VviFCA* might be relevant taking into consideration the abortion of carpel structures in late development stages in male plants. However, no differences were detected between the three flower types in the expression of the gene (VIT_206s0061g00290) that codes for *microRNA172* which suggest another *VviFCA* regulation in *Vitis*. A relevant set of genes with differential expression is *NODULIN MtN3*, from the Lipoxygenase family that plays an important role in exine pattern determination in microspores (Guan et al. 2008). This gene was covered in our previous work (Ramos et al. 2014) and in this new assembly *VviNODULIN MtN3* is, again, highly up-expressed in development stage H of female plants when compared with both male and hermaphrodite plants in the same development stage. Female plants display infertile pollen with and oval shape and without pores (Gallardo et al. 2009). This up-regulation of *VviNODULIN MtN3* might play a role in excessive deposition of exine in female pollen making it infertile and thus contributing to the feminizing phenotype in *Vitis vinifera sylvestris* female plants. It is not to exclude that key player(s) for sex specification in *Vitis* might be located in the Unknown functional category, in the current *Vitis* annotation. The differential expression analysis of this functional category shows

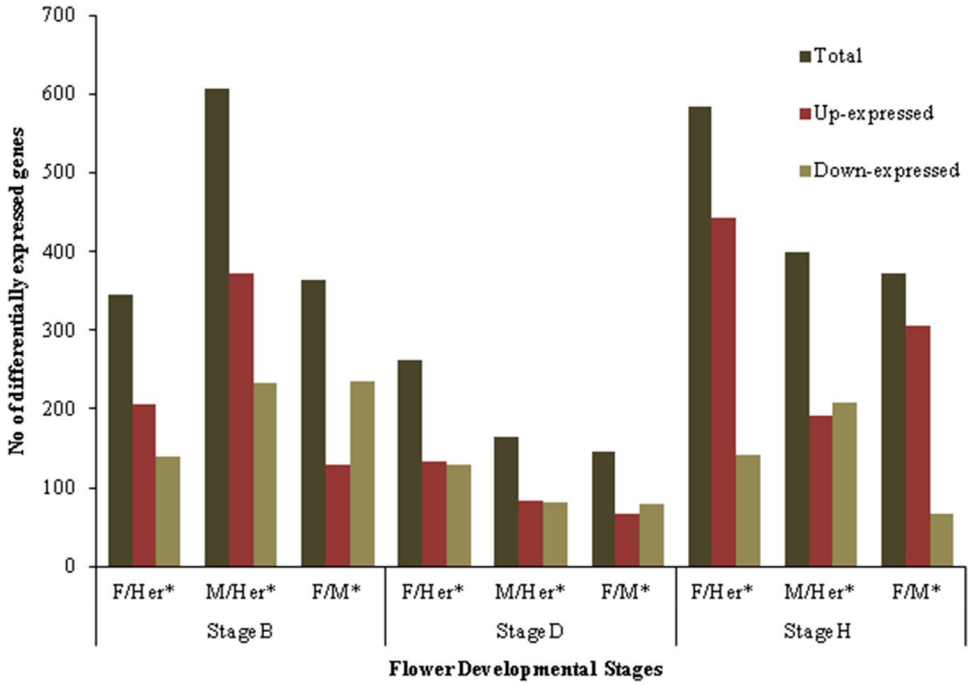

Fig. 6 Differentially expressed genes between flower types in each developmental stage. DEGs in female and male were calculated by using the hermaphrodite transcriptome as reference. DEGs between female and male, the male transcriptome was the reference. Filters used: FDR = 0.05 and $-1 < \log2$ (fold change) > 1, *F* female, *M* male, *Her* hermaphrodite. (*) Sample used as reference

429 DEGs in male flower at developmental stage B (298 up-expressed and 131 down-expressed) (Fig. 8a). Differences between male and female flowers are most visible at developmental stage B, where the female has 272 differential expressed genes (77 up-expressed and 195 down-expressed) (Fig. 8a). However, the differences between female and hermaphrodite flowers are most evident in stage H, where 273 genes are differentially expressed (203 up-expressed and 70 down-expressed) (Fig. 8a). The analysis of the transcripts that have no correspondence in any database show differential expression between flower developmental stages and flower types (Fig. 8b). It allowed to identify a decrease in the number of these differentially expressed transcripts during flower development with *V. v. sylvestris* presenting the largest number of up-expressed transcripts when compared to *V. v. vinifera* (Fig. 8b). These novel sequences could be ancestral genes, targets for genome evolution (Chen et al. 2013). Some of them are still shared between the wild type (*V. v. sylvestris*) and the domesticated grapevine while gene diversification proceeds until a new gene function evolves. New genes could arise by genome/gene duplication or by retroelements that transcribed through an RNA intermediate (Chen et al. 2013; Senerchia et al. 2013; Zou et al. 2009). The blast against the Repbase (http://www.girinst.org/), a database of mobile elements revealed homologies with mobile elements: transposon sequences (30), LTR retrotransposon (38) and Non-LTR retrotransposon (12) (Table 3). We cannot discard the possibility that these mobile elements are players in plant sex, modulating the genome in order specify the sex.

Transcription factors

The DEG analysis reveals 36 differentially expressed genes annotated as Transcription Factors (Online Resource Table 1). One transcription factor with strong differential expression is the gene VIT_201s0127g00860 annotated as *TRANSCRIPTION FACTOR ABORTED MICROSPORES-LIKE (AMS)*, which shows no expression in any sample until the inflorescence reaches the developmental stage H. In this stage, VIT_201s0127g00860 is strongly expressed in the female inflorescence with differential expression relative to male and hermaphrodite inflorescences. The gene VIT_212s0057g00440 is a GCN5 *N*-acetyltransferase (GNAT) and could be associated to the transference of an acetyl group, from acetyl coenzyme A (CoA) to a substrate. However, no effective function has been associated to this gene yet. This transcription factor has a basal expression on all samples, but in female inflorescences its action seems to be crucial in flower developmental stage H, as showed by the higher expression when compared with male and hermaphrodite samples (Online Resource Table 1). The transcription factor VIT_205s0029g00130 annotated as High mobility group (HMG1) has an opposite expression pattern, with no expression in female plants at developmental stage H and higher expression in male and hermaphrodite inflorescences.

The MADS-box gene family is a large and important group that promotes changes in gene expression pathways acting as master regulators of developmental switches to define the diversity of floral organs, and to control organ differentiation

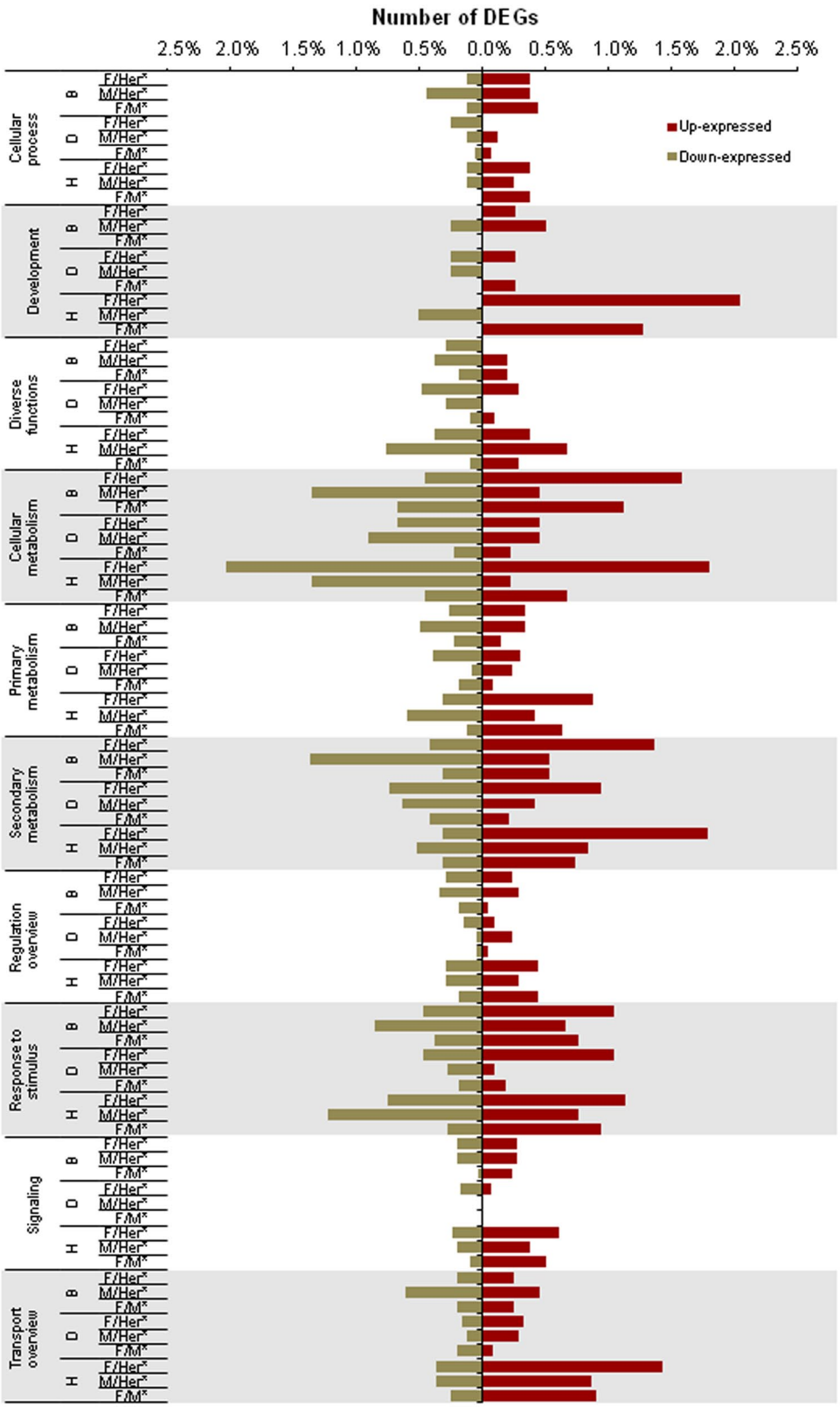

Fig. 7 DEGs between two flower types in the same developmental stage in ten functional categories. Percentage of DEGs present in the transcriptome relatively to the total number of annotated genes by functional category. DEGs were determined in female and a male sample using the hermaphrodite as reference. To compare female and male the male transcriptome was the reference. Filters used: FDR = 0.05 and −1 > log2 (fold change) > 1, *F* female, *M* male, *Her* hermaphrodite. (*) Sample used as reference

and development (Smaczniak et al. 2012). Three genes annotated as MADS box transcription factors were found to have different behaviors between flower types: *AGL16_AGL21* (VIT_218s0001g07900), *AGL21* (VIT_200s0211g00180) and *SOC1-like* (VIT_216s0022g02380). The first two genes (VIT_2 18s0001g07900 and VIT_200s0211g00180) belong to the *AGAMOUS* clade. The first one is allocated to chromosome 18 and has higher expression in the hermaphrodite

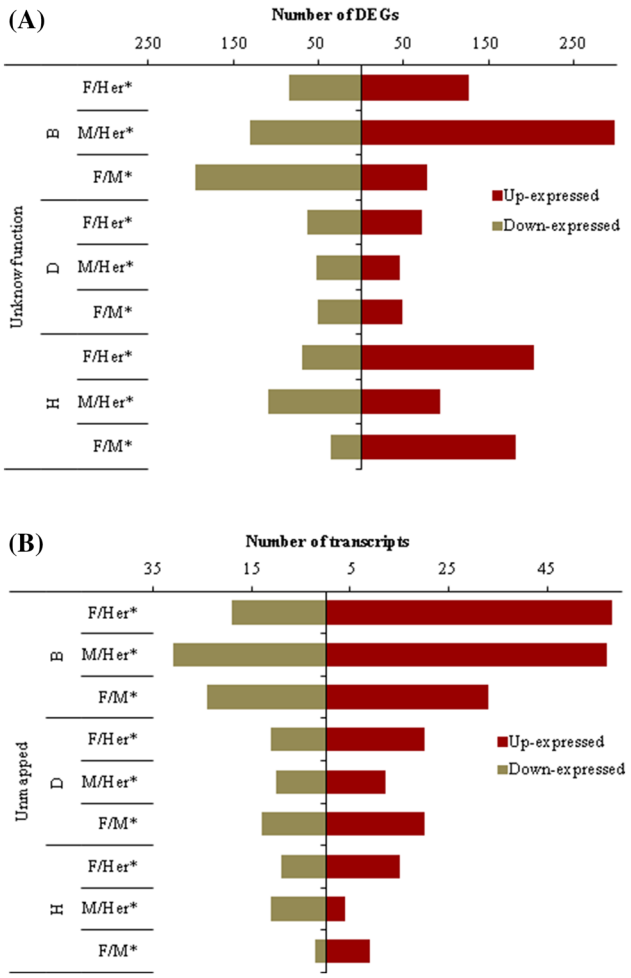

Fig. 8 Overview of differentially expressed genes up and down expressed. **a** Genes with Unknown function, **b** Transcripts with no reference in any database. DEGs were determined in female and a male sample using the hermaphrodite as reference. To compare female and male the male transcriptome was the reference. Filters used: FDR = 0.05 and $-1 > \log2$ (fold change) > 1, *F* female, *M* male, *Her* hermaphrodite. (*) Sample used as reference

Table 3 Fragments of mobile elements present in sequences with no homologies in any database

Repeat class	Sequences	Length (bp)[a]
DNA transposon	30	3255
LTR retrotransposon	38	2601
Non-LTR retrotransposon	12	811

[a]Length; number of the total base pairs present in all sequences

inflorescence when compared with female and male in flower development stage H and with male developmental stage B. The second one (VIT_200s0211g00180) still belongs to the chromosome Unknown and is highly expressed in male stage B when compared with the same stage in the female inflorescence. Both may be alleles of *AGL21*. It is known that *AGL21* is involved in root development and are influenced by hormone signals (Yu et al. 2014). Its expression in several stages of *Vitis* inflorescence development and flower types suggest a possible role in sex specification. Another MADS-box gene is the *SOC1-like* gene (VIT_216s0022g02380) that has differential expression in female developmental stage B when compared with the hermaphrodite at the same stage. *SOC1* gene plays an important role in flower development, acting as an integrator of signals from different flowering pathways (Lee and Lee 2010; Zhong et al. 2012). In *Vitis, VviSOC1* transcript was found in buds and vegetative organs (Díaz-Riquelme et al. 2009); however, its early expression in inflorescences of female *Vitis* plants when compared with the hermaphrodite indicates that, directly or indirectly, *VviSOC1* is responsible for intrinsic characteristics of the female phenotype.

Hormones and sex specification

The development of a flower is intrinsically connected to the reception of numerous environmental input signals that coordinate the expression of many genes and interaction of their products.

At first place, the floral meristem identity genes activate floral organ identity genes, whose expression pattern is required for the developmental fate of floral organ primordia (Wellmer et al. 2014).

Numerous hormones including auxin, ethylene, jasmonates, gibberellins (GA) and abscisic acid (ABA) affect floral development (Golenberg and West 2013). DEGs involved in hormone biosynthesis and metabolism could correspond to players in the regulation of flower development as was the case for our data in stage B and H of the three types of flowers (Table 4; Online Resource Table 2).

Regarding the phytohormone auxin, in the early stage B there is one gene up (VIT_212s0134g00230) and one gene down-expressed (VIT_209s0002g01340) in female, and two genes down-expressed male flowers (VIT_207s0104g00800 and VIT_205s0049g00460) when compared with the hermaphrodite transcriptome. The later stage H only has up-expressed genes in female (VIT_212s0059g00610 and VIT_201s0150g00300) and male (VIT_201s0150g00300) when compared with the hermaphrodite. Although, we could not assign one unique function to these genes it is not to exclude the hypothesis that the genes involved in the auxin biosynthesis in *Vitis* could have a putative role in masculinizing or feminizing the flower, although such effects seem to depend on the *Vitis* gender and the developmental flower stage (Table 4; Online Resource Table 2).

Genes involved in ethylene signaling are down-expressed in female (VIT_204s0023g02820) compared to male in stage B, and up-expressed in male (VIT_213s0084g00080 and VIT_218s0089g01030) when compared to the

Table 4 Genes with differential expression involved in hormone signaling functional category. Flower developmental stages B, D and H. *F* Female, *M* Male, *Her* Hermaphrodite (Touriga Nacional)

Gene ID	Functional category	Stage B			Stage D			Stage H		
		F/Her[a]	M/Her[a]	F/M[a]	F/Her[a]	M/Her[a]	F/M[a]	F/Her[a]	M/Her[a]	F/M[a]
VIT_212s0059g00610	Signaling.Auxin Signaling.Auxin-mediated Signaling pathway							Up		
VIT_212s0134g00230	Signaling.Auxin Signaling.Auxin-mediated Signaling pathway	Up								
VIT_209s0002g01340	Signaling.Auxin Signaling.Auxin-mediated Signaling pathway	Down								
VIT_207s0104g00800	Signaling.Auxin Signaling.Auxin metabolism		Down							
VIT_201s0150g00300	Signaling.Auxin Signaling.Auxin metabolism							Up	Up	
VIT_205s0049g00460	Signaling.Auxin Signaling.Auxin metabolism		Down							
VIT_204s0023g02820	Signaling.Ethylene Signaling.Ethylene-mediated Signaling pathway			Down						
VIT_213s0084g00080	Signaling.Ethylene Signaling.Ethylene-mediated Signaling pathway								Up	
VIT_218s0089g01030	Signaling.Ethylene Signaling.Ethylene-mediated Signaling pathway								Up	
VIT_213s0047g00450	Signaling.Jasmonate salicylate signaling				Down			Down	Down	
VIT_215s0048g01320	Metabolism.Secondary metabolism.Terpenoid metabolism							Up		
VIT_219s0177g00030	Metabolism.Secondary metabolism.Terpenoid metabolism	Down		Down						
VIT_209s0002g07110	Signaling.ABASignaling.ABA-mediated Signaling pathway							Up		Up
VIT_212s0034g00110	Signaling.ABASignaling.ABA-mediated Signaling pathway	Up								
VIT_219s0015g01020	Signaling.ABASignaling.ABA-mediated Signaling pathway				Down					

[a]DEGs were determined in female and a male sample using the hermaphrodite has a reference and using male sample has a reference when compared with the female sample

hermaphrodite in stage H. This fact points for a possible masculinizing effect of ethylene in *Vitis* male plants (Table 4; Online Resource Table 2).

Other genes like 1-aminocyclopropane-1-carboxylate synthase (ACS) involved in ethylene biosynthetic pathway (Wang et al. 2002) regulates unisexual flower development in cucumber (Knopf and Trebitsh 2006; Yamasaki et al. 2003). Two of these genes are differentially expressed in male and female cucumber flowers with high expression in female flowers (Knopf and Trebitsh 2006). However, this gene family has no differential expression in *Vitis* which suggest another ethylene regulation in early flower development.

The only differentially expressed gene is VIT_213s0047g00450, which is involved in jasmonate signaling. This gene is down-expressed in female flowers in stages D and H when compared to the hermaphrodite, and also in males relatively to the hermaphrodite in stage H (Table 4; Online Resource Table 2). Jasmonate signaling is required for stamen and pollen maturation as show by *Arabidopsis* mutants that are defective in the synthesis or signaling of jasmonate present male sterile (Park et al. 2002). Jasmonate biosynthesis genes play many roles in regulating flower development. In our expression data we could not assign this gene to a male sterility function as reported for *Arabidopsis*, but the fact that it is up-expressed in hermaphrodite flowers in stages D and H hints for an Unknown function of this gene in this type of flowers (Table 4; Online Resource Table 2).

Genes that code for gibberellins show down-expression in female when compared with hermaphrodite and male flowers in stage B (VIT_219s0177g00030), and up-expression in stage H when compared to the hermaphrodite (VIT_215s0048g01320). For instance, gibberellins have been shown to be important during stamen development, since *Arabidopsis* mutants with no gibberellins production have compromised stamen development with reduced pollen fertility (Cheng et al. 2004). Gibberellins can also directly affect *LEAFY* expression, a floral meristem gene (Blazquez et al. 1998), as well as the expression of floral organ identity genes, like *AGAMOUS (AG), APETALA3* (*AP3*), and *PISTILLATA* (*PI*) (Yu et al. 2004). Although, none of these genes were affected by gibberellins in the three *Vitis* flower types; however, the two differentially expressed genes in *Vitis* involved in the gibberellins pathway could act at different stages in flower development through a dual role: a masculinizing effect in stage B and a possible feminizing action in stage H, as evident by up-expression in female flowers (Table 4; Online Resource Table 2).

Several abscisic acid (ABA) response *loci* are directly involved in controlling flowering time (Kurup et al. 2000), or indirectly involved in meristem function (Kuhn et al. 2007; Lu and Fedoroff 2000). In our data ABA related genes were up-expressed in stage B (VIT_212s0034g00110) and in stage H (VIT_209s0002g07110) and down-expressed in stage D (VIT_219s0015g01020) of female flowers when compared to the hermaphrodite. Genes involved in ABA metabolism may act in the development of stamens in *V. v. sylvestris* female plant as it does in tomato (Sekhar and Sawhney 1991).

The differential expression of genes related to hormone biosynthesis seems to be associated to the timing of flower formation in different flower types, rather than with the development of the specific structures of each flower type.

The sex chromosome?

A region in chromosome 2 spawning from 4,907,434 to 5,037,597 bp has been considered as important for sex determination in *V. v. vinifera*, since it presents a duplication of 122 bp inserted in the gene *ADENINE PHOSPHORIBOSYL TRANSFERASE* (*APT*) that acts as a genetic marker for differentiating female from male and hermaphrodite plants (Fechter et al. 2012). More recently, a new study enlarged the region from the marker *VVIB23* at 4,781,551 to 5,037,597 bp and found several transcripts with XY type polymorphism (Picq et al. 2014). Considering the importance of this region for sex determination, it was given special attention to the transcripts mapped to this region. It must be stressed that the 12X_v2.1 annotation several transcripts described in previous works as relevant for sex determination in *Vitis* are absent from chromosome 2: *ETHYLENE OVERPRODUCER LIKE 1* (*ETO1*) is located in chromosome Unknown (VIT_200s0233g00090) in the *Vitis* database, *ADENINE PHOSPHORIBOSYL TRANSFERASE* (*APT*) is also located in the chromosome Unknown (VIT_200s1847g00010), and *PHOSPHATIDIC ACID PHOSPHATASE* (*PAP2*) is located in the chromosome 12 (VIT_212s0057g01540) (Table 5). In addition, in the 12X_v2.1 annotation, a new gene (VIT_202s0154g00230) was allocated to chromosome 2 between 5,036,984 and 5,037,952 bp at the end of the sex determination region with a sequence annotated as *Pinus taeda* anonymous locus. This newly allocated *Pinus taeda* anonymous locus did not have a matching transcript in our assembly. Also several other genes did not have a matching transcript, for example, out of the four *FLAVIN-CONTAINING MONOOXYGENASE* (*FMO*) genes, only one was mapped and had its expression quantified (VIT_202s0154g00190) (Table 5). Moreover, in the 4,781,551 to 5,037,597 bp regions, several other genes did not have a matching transcript in this assembly. The region spawning from 4,851,659 to 4,875,700 bp harbors four transcripts, including an *ABNORMAL FLOWER ORGANS* gene (VIT_202s0154g00070), that when blasted with TAIR database (https://www.arabidopsis.org/) retrieves a *YABBY* transcription factor family transcript. Also without a matching

Table 5 Sex genes present in chromosome 2 sex region. Annotation, position in the chromosome 2 (in bp) and gene ID based on CRIBI annotation 12X_v2.1. Expression in RPKM in female (F), male (M) and hermaphrodite (Her) in several flower development stages, B, D and H

Annotation	Position chr02	Gene ID	FB	FD	FH	MB	MD	MH	HerB	HerD	HerH
Serine threonine kinase	4762039–4762742	VIT_202s0241g00140	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unknown	4766577–4768238	VIT_202s0241g00150	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Esterase/lipase/thioesterase family/unknown	4768428–4782138	VIT_202s0241g00160	90.1	105.9	95.4	95.7	99.5	102.0	101.5	106.5	104.1
Unknown	4783908–4789369	VIT_202s0241g00170	13.5	11.5	13.1	12.5	11.8	11.7	14.9	15.1	18.7
GAE5/nucleotide sugar epimerase	4802060–4803142	VIT_202s0241g00180	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Unknown	4810829–4811634	VIT_202s0241g00190	0.0	1.5	0.0	0.0	0.0	0.0	2.3	0.0	4.2
Saur family protein/SAUR11	4822437–4823283	VIT_202s0154g00010	2.9	10.6	7.4	2.7	5.6	3.9	7.4	5.0	7.9
Vascular related NAC-domain protein 4	4836970–4839472	VIT_202s0154g00020	1.8	5.4	5.4	3.1	4.6	5.8	4.7	4.9	9.5
Unknown	4841365–4841723	VIT_202s0154g00030	2.0	1.0	3.3	2.2	2.4	3.7	0.8	3.1	2.4
Thylakoid lumenal protein	4845959–4851301	VIT_202s0154g00040	2.0	5.5	9.6	1.5	4.6	5.8	1.4	0.0	10.0
Hydrolyzing o-glycosyl	4851659–4852238	VIT_202s0154g00050	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Thylakoid lumenal protein	4852239–4855821	VIT_202s0154g00060	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Abnormal floral organs (TAIR:transcription factor YABBY)	4861970–4864802	VIT_202s0154g00070	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Monocopper oxidase-like protein sku5	4870422–4875700	VIT_202s0154g00080	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Acid invertase/GIN2	4882029–4884360	VIT_202s0154g00090	73.9	101.6	64.7	91.5	80.1	70.2	127.7	106.1	116.3
Transaldolase	4887133–4897511	VIT_202s0154g00100	1.3	4.3	2.4	3.3	2.7	1.5	3.7	3.0	5.0
Trehalose-6-phosphate phosphatase	4897621–4903784	VIT_202s0154g00110	52.2	66.5	76.0	62.4	60.0	95.0	78.1	80.6	105.7
Unknown	4914168–4914984	VIT_202s0154g00120	0.6	0.5	6.7	0.7	0.9	1.7	0.6	0.9	2.2
Exostosin (xyloglucan galactosyltransferase KATAMARI 1)	4921780–4923581	VIT_202s0154g00130	22.8	35.3	37.8	19.9	26.9	26.4	17.7	22.8	26.5
3-oxoacyl-[acyl-carrier-protein] synthase 3 A	4921664–4935576	VIT_202s0154g00140	125.4	134.4	85.3	110.4	108.8	95.5	103.3	108.6	87.3
PLATZ transcription factor (Blast TAIR)	4949210–4950534	VIT_202s0154g00150	10.5	7.1	18.4	33.2	39.3	89.0	14.1	14.8	33.3
Ethylene overproducer like 1 (ETO1)[a]		VIT_200s0233g00090[a]	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Flavin-containing monooxygenase family protein	4951841–4956070	VIT_202s0154g00160	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Flavin-containing monooxygenase 3	4957715–4960913	VIT_202s0154g00170	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Flavin-containing monooxygenase 3	4962382–4965728	VIT_202s0154g00180	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Flavin-containing monooxygenase 3	4974431–4978234	VIT_202s0154g00190	44.4	39.8	42.1	28.3	24.7	40.3	43.1	39.3	39.1
Unknown protein	4981728–4986851	VIT_202s0154g00200	62.3	51.9	30.9	66.7	60.5	63.3	61.0	51.8	40.1
WRKY DNA-binding protein 21	4989461–4990306	VIT_202s0154g00210	2.9	0.0	2.5	1.5	3.0	3.5	2.5	2.9	1.6
Adenine phosphoribosyltransferase APT3/APT1		VIT_200s1847g00010[a]	7.6	4.7	6.8	3.3	6.8	5.0	13.2	3.6	7.4
Unknown locus	5025223–5026178	VIT_202s0154g00220	73.4	66.6	32.7	49.6	54.0	36.5	62.5	70.3	32.1
Phosphatidic acid phosphatase/PAP2		VIT_212s0057g01540[a]	27.2	22.7	32.0	42.2	36.2	47.5	40.1	45.6	38.2
Pinus taeda anonymous locus 0_16347_01	5036984–5037952	VIT_202s0154g00230	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
Oxysterol-binding protein 2a-like	5062292–5069255	VIT_202s0154g00240	9.2	4.3	3.1	3.0	3.1	2.9	6.5	7.7	0.0

[a]Genes placed in the *Vitis* database in chromosome Unknown or in other chromosome than the chromosome 2

transcript in our assembly and, therefore, no quantifiable expression value (Table 5). The *ETO1* absent in the 12X_v2.1 does not match to any transcript in the assembly, even before filters were applied.

More recently, by using a new heterozygous mapping strategy, it was possible to include five scaffolds from the chromosome Unknown into the chromosome 2, surrounding scaffold_0154 of this chromosome (Fechter et al. 2012; Hyma et al. 2015). Considering the genes comprising these scaffolds, only VIT_200s0229g00190 was differentially expressed in female inflorescences in development stage D, which displays a higher expression value when compared with male and hermaphrodite at the same developmental stage. Also, scaffolds_1344 and scaffold_1682, which comprise a total of three genes (VIT_200s1344g00010, VIT_200s1682g00010 and VIT_200s1682g00020) were also placed in the sex region of chromosome 2 (Hyma et al. 2015). However, none of these three genes were expressed in our de novo assembly.

Admitting that this region of chromosome 2 is important for *Vitis* sex determination, it would be expected to find disparities in transcript expression between the different flower types. In the DEG analysis none of the genes that comprise this region was differently expressed. However, flower development is known to involve a myriad of genes and pathways that could interact with other genes for sex specification. The focus on gene expression revealed that most expression disparities occur between development stages, although, few stand out. From 4,810,829 to 4,811,634 bp there is an Unknown transcript (VIT_202s0241g00190), without any matching transcripts in the male plants in any of the three developmental stages. However, VIT_202s0241g00190 has a matching transcript with low expression at stage D in female flowers, while it displayed matching transcripts in the hermaphrodite flower at stages B and H, with expression values ranging from 2 (in stage B) to 4 (in stage H) (Table 5). The gene just downstream of this unknown transcript is annotated as a *SAUR FAMILY PROTEIN* (*SAUR11*, VIT_202s0154g00010), which stands out for its expression in female stage D more than doubling when compared with male and hermaphrodite plants at the same stage. Also, male stage H expression of this transcript is considerably lower than female and hermaphrodite expression at the same stage.

Further downstream, a *TREHALOSE-6-PHOSPHATE PHOSPHATASE* (VIT_202s0154g00110) was a prominent gene in a previous study (Picq et al. 2014) by displaying XY type polymorphism. Although this work does not focus on polymorphisms, an increasing expression was observed across developmental stages in the three flower types with a slight decrease in male flower stage D (Table 5). Another gene regarded as having a XY type polymorphism is the exostosin family protein locus (VIT_202s0154g00130) (Picq et al. 2014), which has a noticeable higher expression in the female flower compared to the male and hermaphrodite flowers, underlining a potentially interesting gene to study in *Vitis* sex specification. Two loci downstream of *EXOSTOSIN* locate a *PLATZ* transcription factor (VIT_202s0154g00150), whose expression in male flowers is considerably higher than in the other two flower types. Between 5,026,178 and 5,036,984 bp resides the *locus* of the *PHOSPHATIDIC ACID PHOSPHATASE* (PAP2, VIT_212s0057g01540) (Table 5). In the previous *Vitis vinifera* annotation this gene was allocated to chromosome 2, however in *Vitis* reference annotation (12X_v2.1) it is located in the chromosome 12. This is an interesting *locus*, since its expression in early developmental stages is doubled in both male and hermaphrodite flowers when compared to female flowers, suggesting a possible role in sex determination.

Conclusion

We carried out a de novo assembly of the transcriptome of three flower developmental stages in three *Vitis* flower types. This new assembly detected novel areas of transcription, mapping to regions currently not annotated as gene models (intra and intergenic regions) that may represent novel genes. It also suggests, during early *Vitis* flower development that there is active transcription of previously unannotated genomic loci bioinformatic analysis identified 133 transcripts specific to *V. v. sylvestris* and 516 transcripts differentially expressed among the three flower types, when the three developmental stages were merged. Additionally, some of these regions are differentially expressed during *Vitis* flower development and may represent novel candidate *loci* involved in sex establishment such non-coding RNAs, repetitive sequences, mobile elements or miRNAs. Although the specific role of hormones in *Vitis* sex establishment and flower development has not yet been analyzed in depth, our data show differential gene expression during flower development of genes that control hormone behavior among the three *Vitis* flower types. Also, analyzing the region of chromosome 2 which has been considered relevant for *Vitis* sex specification we found that several genes reported as part of that region are still located in chromosome Unknown, suggesting that the latest *Vitis* annotation (12X_v2.1) needs further confirmation. However, the expression of some of those genes was not detected in the development stages covered by our study. Nevertheless, other genes in the same region displayed interesting expression patterns between the three flower types and could be involved in sex establishment in *Vitis*.

Acknowledgments This work was supported by the funded project PTDC/AGR-GPL/119298/2010 from Fundação para a Ciência e

Tecnologia (FCT, Portugal). MRocheta, JLCoito, JCunha, MCosta, MRamos and HGSilva are supported by FCT fellowships SFRH/BPD/64905/2009, SFRH/BD/85824/2012, SFRH/BPD/74895/2010, SFRH/BSAB/113781/2015, SFRH/BD/110274/2015 and SFRH/BD/111529/2015, respectively. We are also grateful to Eng. Eiras-Dias, curator from Portuguese Ampelographic Collection (property of Instituto Nacional de Investigação Agrária e Veterinária, Dois Portos) where sampling was performed, for the collaboration in this work allowing the access to the *Vitis* collection.

Author contributions MR, MNJR and JLC conceived and designed the experiments. MNJR, JLC and MR performed the experiments. MJNR, JLC, MR, JF and OSP analyzed the data. MNJR, JLC, HGS, MMRC and MR wrote the paper. PGA revised and correct the manuscript. SA read and discussed the manuscript. JC established *Vitis vinifera sylvestris* collection and collected plant tissues samples. All authors read and approved the final manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L (2011) Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. Nat Struct Mol Biol 18:1435–1440. doi:10.1038/nsmb.2143

Aukerman MJ, Sakai H (2003) Regulation of flowering time and floral organ identity by a microRNA and its *APETALA2*-like target genes. Plant Cell 15:2730–2741. doi:10.1105/tpc.016238

Baggiolini M (1952) Les stades repères dans le developpement annuel de la vigne et leur utilisation pratique. Rev Romande Agric Vitic Arbor 8:4–6

Battilana J, Lorenzi S, Moreira FM, Moreno-Sanz P, Failla O, Emanuelli F, Grando MS (2013) Linkage mapping and molecular diversity at the flower sex locus in wild and cultivated grapevine reveal a prominent SSR haplotype in hermaphrodite plants. Mol Biotechnol 54:1031–1037

Blazquez MA, Green R, Nilsson O, Sussman MR, Weigel D (1998) Gibberellins promote flowering of *Arabidopsis* by activating the *LEAFY* promoter. Plant Cell 10:791–800

Bondino HG, Valle EM (2009) A small intergenic region drives exclusive tissue-specific expression of the adjacent genes in Arabidopsis thaliana. BMC Mol Biol 10:95. doi:10.1186/1471-2199-10-95

Boualem A et al (2008) A conserved mutation in an ethylene biosynthesis enzyme leads to andromonoecy in melons. Science 321:836–838. doi:10.1126/science.1159023

Browse J (2009) Jasmonate: preventing the maize tassel from getting in touch with his feminine side Sci Signal 2(59):pe9. doi:10.1126/scisignal.259pe9

Celaj A, Markle J, Danska J, Parkinson J (2014) Comparison of assembly algorithms for improving rate of metatranscriptomic functional annotation. Microbiome 2:39. doi:10.1186/2049-2618-2-39

Chen S, Krinsky BH, Long M (2013) New genes as drivers of phenotypic evolution. Nat Rev Genet 14:645–660

Cheng H et al (2004) Gibberellin regulates *Arabidopsis* floral development via suppression of DELLA protein function. Development 131:1055–1064. doi:10.1242/dev.00992

Da Silva C et al (2013) The high polyphenol content of grapevine cultivar tannat berries is conferred primarily by genes that are not shared with the reference genome. Plant Cell 25:4777–4788. doi:10.1105/tpc.113.118810

Dalbó MA, Ye GN, Weeden NF, Steinkellner H, Sefc KM, Reisch BI (2000) A gene controlling sex in grapevines placed on a molecular marker-based genetic map. Genome 43:333–340

de Montaigu A, Toth R, Coupland G (2010) Plant development goes like clockwork. Trends Genet 26:296–306. doi:10.1016/j.tig.2010.04.003

Dellaporta SL, Calderon-Urrea A (1994) The sex determination process in maize. Science 266:1501–1505

Díaz-Riquelme J, Lijavetzky D, Martínez-Zapater JM, Carmona MJ (2009) Genome-wide analysis of MIKCC-type MADS box genes in grapevine. Plant Physiol 149:354–369

Durand B, Durand R (1991) Sex determination and reproductive organ differentiation in *Mercurialis*. Plant Sci 80:49–65. doi:10.1016/0168-9452(91)90272-A

Falgueras J, Lara AJ, Fernández-Pozo N, Cantón FR, Pérez-Trabado G, Claros MG (2010) SeqTrim: a high-throughput pipeline for pre-processing any type of sequence read. BMC Bioinformatics 11:38. doi:10.1186/1471-2105-11-38

Fechter I, Hausmann L, Daum M, Sorensen TR, Viehover P, Weisshaar B, Topfer R (2012) Candidate genes within a 143 kb region of the flower sex locus in *Vitis*. Mol Genet Genomics 287:247–259. doi:10.1007/s00438-012-0674-z

Fuchs E, Atsmon D, Halevy AH (1977) Adventitious staminate flower formation in gibberellin treated gynoecious cucumber plants. Plant Cell Physiol 18:1193–1201

Gaidatzis D, Burger L, Florescu M, Stadler MB (2015) Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. Nat Biotechnol 33:722–729. doi:10.1038/nbt.3269

Gallardo A, Ocete R, López MÁ, Lara M, Riviera D (2009) Assessment of pollen dimorphism in populations of *Vitis vinifera* L. subsp. *sylvestris* (Gmelin) Hegi in Spain. VITIS 48:59–62

García RAA, Revilla E (2013) The current status of wild grapevine populations (*Vitis vinifera* ssp *sylvestris*) in the Mediterranean Basin. In: Poljuha D, Sladonja B (eds). InTech, pp 51–72. doi:10.5772/52933

Golenberg EM, West NW (2013) Hormonal interactions and gene gegulation can link monoecy and environmental plasticity to the evolution of dioecy in plants. Am J Bot 100:1022–1037

Guan YF, Huang XY, Zhu J, Gao JF, Zhang HX, Yang ZN (2008) *RUPTURED POLLEN GRAIN*1, a member of the MtN3/saliva gene family, is crucial for exine pattern formation and cell integrity of microspores in *Arabidopsis*. Plant Physiol 147:852–863. doi:10.1104/pp.108.118026

Hartwig T et al (2011) Brassinosteroid control of sex determination in maize. Proc Natl Acad Sci USA 108:19814–19819. doi:10.1073/pnas.1108359108

Hendriks GJ, Gaidatzis D, Aeschimann F, Grosshans H (2014) Extensive oscillatory gene expression during *C. elegans* larval development. Mol Cell 53:380–392. doi:10.1016/j.molcel.2013.12.013

Hogenesch JB, Ueda HR (2011) Understanding systems-level properties: timely stories from the study of clocks. Nat Rev Genet 12:407–416. doi:10.1038/nrg2972

Huijser P, Schmid M (2011) The control of developmental phase transitions in plants. Development 138:4117–4129. doi:10.1242/dev.063511

Hyma KE et al (2015) Heterozygous mapping strategy (HetMappS) for high resolution genotyping-by-sequencing markers: a case

study in grapevine. PloS One 10:e0134880. doi:10.1371/journal.pone.0134880
Jung JH, Seo PJ, Ahn JH, Park CM (2012) *Arabidopsis* RNA-binding protein FCA regulates microRNA172 processing in thermosensory flowering. J Biol Chem 287:16007–16016
Kapranov P, St Laurent G (2012) Dark matter RNA: Existence, function, and controversy. Front Genet 3:60. doi:10.3389/fgene.2012.00060
Kelley JL, Passow CN, Plath M, Arias Rodriguez L, Yee MC, Tobler M (2012) Genomic resources for a model in adaptation and speciation research: characterization of the *Poecilia mexicana* transcriptome. BMC Genomics 13:652. doi:10.1186/1471-2164-13-652
Knopf RR, Trebitsh T (2006) The female-specific *Cs-ACS1G* gene of cucumber. A case of gene duplication and recombination between the non-sex-specific *1-aminocyclopropane-1-carboxylate synthase* gene and a branched-chain amino acid transaminase gene. Plant Cell Physiol 47:1217–1228
Kuhn JM, Breton G, Schroeder JI (2007) mRNA metabolism of flowering-time regulators in wild-type Arabidopsis revealed by a nuclear cap binding protein mutant, abh1. Plant J 50:1049–1062. doi:10.1111/j.1365-313X.2007.03110.x
Kurup S, Jones HD, Holdsworth MJ (2000) Interactions of the developmental regulator ABI3 with proteins identified from developing *Arabidopsis* seeds. Plant J 21:143–155
Kwon CS, Chen CB, Wagner D (2005) *WUSCHEL* is a primary target for transcriptional regulation by *SPLAYED* in dynamic control of stem cell fate in *Arabidopsis*. Genes Dev 19:992–1003
Lazarte JE, Garrison SA (1980) Sex modification in *Asparagus officinalis* L. J Am Soc Hortic Sci 105:691–694
Lebel-Hardenack S, Grant SR (1997) Genetics of sex determination in flowering plants. Trends Plant Sci 2:130–136. doi:10.1016/S1360-1385(97)01012-1
Lee J, Lee I (2010) Regulation and function of *SOC1*, a flowering pathway integrator. J Exp Bot 61:2247–2254. doi:10.1093/jxb/erq098
Li B, Dewey CN (2011) RSEM: accurate transcript quantification from RNA-Seq data with or without a reference genome. BMC Bioinformatics 12:323. doi:10.1186/1471-2105-12-323
Li W, Godzik A (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. Bioinformatics 22:1658–1659. doi:10.1093/bioinformatics/btl158
Li B, Fillmore N, Bai Y, Collins M, Thomson JA, Stewart R, Dewey CN (2014) Evaluation of de novo transcriptome assemblies from RNA-Seq data. Genome Biol 15:553. doi:10.1186/s13059-014-0553-5
Lowe KM, Walker MA (2006) Genetic linkage map of the interspecific grape rootstock cross Ramsey (*Vitis champinii*)×Riparia Gloire (*Vitis riparia*). Theor Appl Genet 112:1582–1592. doi:10.1007/s00122-006-0264-8
Lu C, Fedoroff N (2000) A mutation in the *Arabidopsis HYL1* gene encoding a dsRNA binding protein affects responses to abscisic acid, auxin, and cytokinin. Plant Cell 12:2351–2366. doi:10.1105/tpc.12.12.2351
Manzano S, Martinez C, Garcia JM, Megias Z, Jamilena M (2014) Involvement of ethylene in sex expression and female flower development in watermelon (*Citrullus lanatus*). Plant Physiol Biochem 85:96–104. doi:10.1016/j.plaphy.2014.11.004
Marguerit E et al (2009) Genetic dissection of sex determinism, inflorescence morphology and downy mildew resistance in grapevine. Theor Appl Genet 118:1261–1278. doi:10.1007/s00122-009-0979-4
Moghe GD et al (2013) Characteristics and significance of intergenic polyadenylated RNA transcription in *Arabidopsis*. Plant Physiol 161:210–224. doi:10.1104/pp.112.205245
Mohawk JA, Green CB, Takahashi JS (2012) Central and peripheral circadian clocks in mammals. Annu Rev Neurosci 35:445–462. doi:10.1146/annurev-neuro-060909-153128
Negi SS, Olmo HP (1966) Sex conversion in a male *Vitis vinifera* L. by a kinin. Science 152:1624–1625
Orozco-Arroyo G, Vazquez-Santana S, Camacho A, Dubrovsky JG, Cruz-Garcia F (2012) Inception of maleness: auxin contribution to flower masculinization in the dioecious cactus *Opuntia stenopetala*. Planta 236:225–238. doi:10.1007/s00425-012-1602-5
Oshlack A, Robinson MD, Young MD (2010) From RNA-seq reads to differential expression results. Genome Biol 11:220. doi:10.1186/gb-2010-11-12-220
Park JH, Halitschke R, Kim HB, Baldwin IT, Feldmann KA, Feyereisen R (2002) A knock-out mutation in allene oxide synthase results in male sterility and defective wound signal transduction in *Arabidopsis* due to a block in jasmonic acid biosynthesis. Plant J 31:1–12
Picq S et al (2014) A small XY chromosomal region explains sex determination in wild dioecious *V. vinifera* and the reversal to hermaphroditism in domesticated grapevines. BMC Plant Biol 14:229. doi:10.1186/s12870-014-0229-z
Ponting CP, Belgard TG (2010) Transcribed dark matter: meaning or myth? Hum Mol Genet 19:R162–R168. doi:10.1093/hmg/ddq362
Ramos MJ, Coito JL, Silva HG, Cunha J, Costa MM, Rocheta M (2014) Flower development and sex specification in wild grapevine. BMC Genomics 15:1095. doi:10.1186/1471-2164-15-1095
Ryan PT et al (2015) Patterns of gene expression during *Arabidopsis* flower development from the time of initiation to maturation. BMC Genomics 16:488. doi:10.1186/s12864-015-1699-6
Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092. doi:10.1093/bioinformatics/bts094
Sekhar KNC, Sawhney VK (1991) Role of ABA in stamen and pistil development in the normal and solanifolia mutant of tomato (*Lycopersicon esculentum*). Sex Plant Reprod 4:279–283. doi:10.1007/BF00200548
Senerchia N, Wicker T, Felber F, Parisod C (2013) Evolutionary dynamics of retrotransposons assessed by high-throughput sequencing in wild relatives of wheat. Genome Biol Evol 5:1010–1020. doi:10.1093/gbe/evt064
Smaczniak C, Immink RG, Angenent GC, Kaufmann K (2012) Developmental and evolutionary diversity of plant MADS-domain factors: insights from recent studies. Development 139:3081–3098. doi:10.1242/dev.074674
Sultan M et al (2008) A global view of gene activity and alternative splicing by deep sequencing of the human transcriptome. Science 321:956–960. doi:10.1126/science.1160342
Wagner D, Meyerowitz EM (2002) *SPLAYED*, a novel SWI/SNF ATPase homolog, controls reproductive development in *Arabidopsis*. Curr Biol 12:85–94
Wang KLC, Li H, Ecker JR (2002) Ethylene biosynthesis and signaling networks. Plant Cell 14:S131–S151
Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet 10:57–63. doi:10.1038/nrg2484
Wellmer F, Graciet E, Riechmann JL (2014) Specification of floral organs in *Arabidopsis*. J Exp Bot 65:1–9. doi:10.1093/jxb/ert385
Wu MF et al (2012) SWI2/SNF2 chromatin remodeling ATPases overcome polycomb repression and control floral organ identity with the *LEAFY* and *SEPALLATA3* transcription factors. Proc Natl Acad Sci USA 109:3576–3581
Yamasaki S, Fujii N, Takahashi H (2003) Characterization of ethylene effects on sex determination in cucumber plants. Sex Plant Reprod 16:103–111
Yu H, Ito T, Zhao Y, Peng J, Kumar P, Meyerowitz EM (2004) Floral homeotic genes are targets of gibberellin signaling in flower development. Proc Natl Acad Sci USA 101:7827–7832. doi:10.1073/pnas.0402377101

Yu LH, Miao ZQ, Qi GF, Wu J, Cai XT, Mao JL, Xiang CB (2014) MADS-box transcription factor *AGL21* regulates lateral root development and responds to multiple external and physiological signals. Mol Plant 7:1653–1669. doi:10.1093/mp/ssu088

Zenoni S et al (2010) Characterization of transcriptional complexity during berry development in *Vitis vinifera* using RNA-Seq. Plant Physiol 152:1787–1795. doi:10.1104/pp.109.149716

Zerbino DR (2010) Using the Velvet de novo assembler for short-read sequencing technologies. Curr Protoc Bioinformatics 11(5):1–13. doi:10.1002/0471250953.bi1105s31

Zerbino DR, Birney E (2008) Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. doi:10.1101/gr.074492.107

Zhong X, Dai X, Xv J, Wu H, Liu B, Li H (2012) Cloning and expression analysis of *GmGAL1, SOC1* homolog gene in soybean. Mol Biol Rep 39:6967–6974. doi:10.1007/s11033-012-1524-0

Zou J, Gong H, Yang T-J, Meng J (2009) Retrotransposons: a major driving force in plant genome evolution and a useful tool for genome analysis. J Crop Sci Biotechnol 12:1–8. doi:10.1007/s12892-009-0070-3