

RESEARCH ARTICLE

Open Access



Genetic variation associated with healthy traits and environmental conditions in *Vaccinium vitis-idaea*

Zobayer Alam¹, Julissa Roncal^{1*}  and Lourdes Peña-Castillo^{1,2}

Abstract

Background: Lingonberry (*Vaccinium vitis-idaea* L.), one of the least studied fruit crops in the *Ericaceae* family, has a dramatically increased worldwide demand due to its numerous health benefits. Genetic markers can facilitate the selection of berries with desirable climatic adaptations, agronomic and nutritious characteristics to improve cultivation programs. However, no genomic resources are available for this species.

Results: We used Genotyping-by-Sequencing (GBS) to analyze the genetic variation of 56 lingonberry samples from across Newfoundland and Labrador, Canada. To elucidate a potential adaptation to environmental conditions we searched for genotype-environment associations by applying three distinct approaches to screen the identified single nucleotide polymorphisms (SNPs) for correlation with six environmental variables. We also searched for an association between the identified SNPs and two phenotypic traits: the total phenolic content (TPC) and antioxidant capacity (AC) of fruit. We identified 1586 high-quality putative SNPs using the UNEAK pipeline available in TASSEL. We found 132 SNPs likely associated with at least one of the environmental or phenotypic variables. To obtain insights on the function of the genomic sequences containing the SNPs likely to be associated with the environmental or phenotypic variables, we performed a sequence-based functional annotation and identified homologous protein-coding sequences with functional roles related to abiotic stress response, pathogen defense, RNA metabolism, and, most interestingly, phenolic compound biosynthesis.

Conclusions: The putative SNPs discovered are the first genomic resource for lingonberry. This resource might prove useful in high-density quantitative trait locus analysis, and association mapping. The identified candidate genes containing the SNPs need further studies on their potential role in local adaptation of lingonberry. Altogether, the present study provides new resources that can be used to breed for desirable traits in lingonberry.

Keywords: Antioxidant capacity, Environmental adaptation, Functional annotation, Genetic diversity, Genotyping-by-sequencing, Lingonberry, Phenolic content, Single nucleotide polymorphism

Background

Vaccinium vitis-idaea L. commonly known as lingonberry, is a perennial, evergreen dwarf shrub, which belongs to the *Ericaceae* family, that has high breeding potential for leaf and fruit quality traits such as high concentration of healthy bioactive compounds. *Vaccinium* is a good source of pharmaceutical ingredients because it is rich in phenolic phytochemicals (e.g. flavonoids, phenolic acids), which have proven to reduce the risk of cancer development [1],

hepatitis C [2], cardiovascular disorders [3], diabetes [4], and urinary tract infections [5]. Among the commonly cultivated “berry” fruit such as cranberry, strawberry, raspberry, and blueberry, *V. vitis-idaea* ranks high in antioxidant capacity as conferred by phenolic compounds [6, 7]. Despite its long cultivation history in Scandinavian countries, in North America, breeding is at its developmental stage. Interestingly, the North American *V. vitis-idaea* subsp. *minus* (Lodd) Hult. has a higher anthocyanin content and antioxidant capacity than the Eurasian *V. vitis-idaea* subsp. *vitis-idaea* (L.) Britton [8, 9], increasing the interest to develop a breeding and commercialization program for the North American subspecies.

* Correspondence: jroncal@mun.ca

¹Department of Biology, Memorial University of Newfoundland, St. John's, NL A1B 3X9, Canada

Full list of author information is available at the end of the article



While it is well known that environmental conditions can affect the concentration of phenolic content and antioxidant capacity in *V. vitis-idaea* (e.g. [9, 10]), very few studies have characterized the genetic diversity of wild populations in North America [8, 11, 12]. Neither the complete nuclear genome nor the plastome have been sequenced for this species, and no single nucleotide polymorphisms (SNP) or expressed sequence tags (EST) are available as genetic resources. Genetic markers are therefore needed to facilitate the selection of wild populations with desirable climatic adaptations, agronomic and nutritious characteristics.

In the face of environmental change, it is becoming more important to understand the genetic variation that results from selection to different environmental growing conditions. For several crops, common garden and transplant experiments have revealed strong adaptive clines in growth, phenology traits and physiological responses to abiotic conditions, and sometimes the adaptive genetic variation underlying these differences were revealed (e.g. [13, 14]). Phenolic compound biosynthesis genes have been identified in different plants especially as regulated by the excess or deficiency of the phenylalanine ammonia lyase (*PAL*), an enzyme that catalyzes the first step of the phenylpropanoid pathway, which produces precursors to several important secondary metabolites [15]. For example, the expression of nine phenolic acid biosynthesis pathway genes was closely related to phenolic acids accumulation during grain filling in wheat [16]. Polymorphism of genes encoding isoflavone synthase, an enzyme involved in the phenylpropanoid pathway, was associated with isoflavone concentrations in soybean seeds [17].

Genotyping-by-sequencing (GBS, [18]) is a practical and inexpensive method for high-throughput SNP discovery and genotyping through next-generation sequencing (NGS) technologies. This approach conducts a multiplex sequencing of restriction site-associated DNA, and has been successfully used in crops for numerous applications like quantitative trait loci marker identification [19], characterization of germplasm diversity and conservation [20]. In *Vaccinium*, this technique successfully identified markers for root architecture trait breeding [21].

Knowledge of the genetic basis of phenotypic variation and adaptation to local environmental conditions is necessary for the success of a nutrition-oriented breeding program for lingonberry. Thus, the goals of the current study were: 1) to use the GBS approach for high-throughput identification of intraspecific putative SNPs in *V. vitis-idaea* subsp. *minus*, 2) to assess whether or not identified putative SNPs were correlated with environmental or phenotypic variables, as a potential signature of adaptation, and 3) to identify and functionally

annotate protein-coding genomic sequences containing SNPs exhibiting correlation with environmental or phenotypic variables, as an insight to their potential role in local adaptation.

Methods

Figure 1 illustrates the workflow followed in this study. In the following subsections we describe each step in detail.

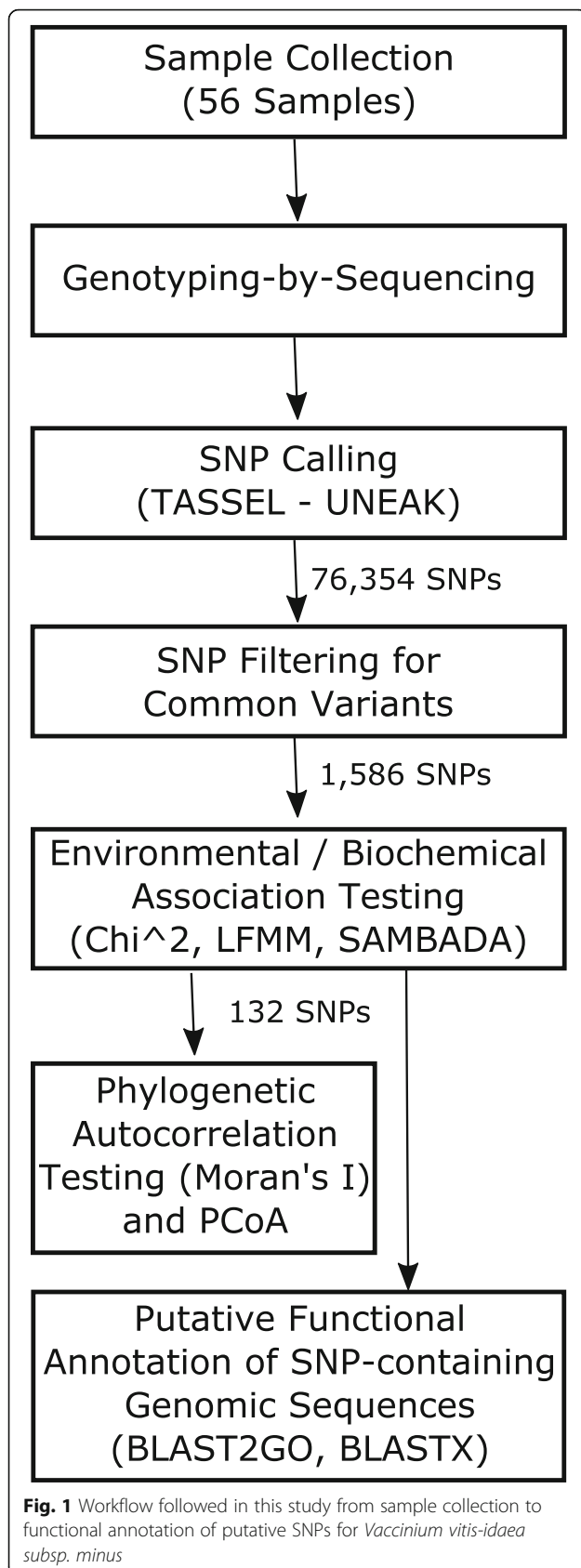
Wild site selection, environmental variables, and plant material collection

Although there is no cultivation practiced in Canada for lingonberry, the province of Newfoundland and Labrador (NL) is the top wild lingonberry fruit producer in Canada with a harvest range of 40 to 500 tons per year [22]. In August 2014, we collected lingonberry adult leaf samples from 56 wild sites distributed across NL. Leaves from one stem were collected per site and stored in silica gel. To elucidate the genetic variation across environmental conditions, we considered the following six variables: mean annual temperature (MAT), mean summer temperature (MST), mean annual precipitation (MAP), mean annual runoff (MAR), surface water pH (SWp), and surface water sensitivity to acid rain (SWS). Using the most updated information from the Water Resources Division of the Government of Newfoundland and Labrador [23] we partitioned the province into categories within these six variables, thus we worked with discrete environmental variables. The environmental characteristics associated with each collection site, a map, and their geographic coordinates were disclosed in Alam et al. [10] and in Additional file 1: Table S1. A voucher specimen was deposited in the NFLD herbarium (JR500).

GBS assay and SNP calling

We extracted total genomic DNA from silica gel dried leaf samples using the DNeasy Plant Mini Kit (Qiagen). The quality and quantity of extracted DNA was evaluated using a Qubit 2.0 Fluorometer (Life Technologies) and a NanoDrop 1000 spectrophotometer (Thermo Fisher Scientific). An estimated quantity of 100 ng of total genomic DNA was used to prepare each library, in a volume of 10 μ L of EB buffer. Reduced representation GBS libraries were prepared for each DNA sample using a double-digest restriction protocol with enzymes *Pst*I and *Msp*I as described in Poland et al. [24] (University of Laval), and sequenced (single-end reads) in one lane using Illumina HiSeq 2500 (McGill University-Génome Québec Innovation Center).

The raw Illumina data were processed following the protocol of Poland et al. [24]. Raw DNA read quality was investigated using FASTQC (Banraham Bioinformatics,



Cambridge, England). We used the TASSEL-UNEAK v.3.0 network-based de novo SNP discovery pipeline without a reference genome [25] to identify accurate and high quality SNPs. In the TASSEL-UNEAK pipeline, the multiplexed reads were sorted into unique sequence tags by compiling exactly matching reads with at least $\geq 3X$ read depth. To reduce the rare or singleton tag from sequencing error, the created taglist (identical reads classified as a tag) was filtered with $\geq 5X$ tag depth. Pairwise alignments were conducted between tag pairs and tag pairs with 1 bp mismatch were considered as putative SNPs. An homology network was constructed by joining all of the tags that differed by a single base, and then a network filter was applied to find reciprocal tag pairs [25] based on a sequencing error tolerance rate (ETR) of 0.05. The created HapMap file was then filtered using the default minor allele frequency (MAF) of 0.05, and a maximum allele frequency of 0.5 to call SNPs. Using a custom R script, putative SNPs were further filtered for minimum coverage threshold at $3X$, minimum minor allele frequency at 0.30, and call rate higher than 50%. A minimum MAF of 0.30 was selected so that the putative SNPs were highly polymorphic and produced clearly separated genotypic clusters [26].

Putative SNPs associated with environmental and phenotypic variables

To identify some of the loci putatively responsible for adaptive differences among individuals, we applied three methods to test for correlations between allele frequencies and the environmental or phenotypic variables. These methods were Chi-square test [27], Latent Factor Mixed Models (LFMM, [28]), and a logistic regression-based approach (Sam β ada, [29]). LFMM estimates the influence of population structure on allele frequencies by introducing unobserved variables as latent factors [28]. Sam β ada uses logistic regressions to model the probability of observing a SNP genotype given the environmental conditions at the sampling location [29]. Fruit total phenolic content (TPC) and antioxidant capacity (AC) for *V. vitis-idaea* sampled in this study were analyzed previously in Alam et al. [10]. To be able to apply Chi-square test with the phenotypic variables, samples were categorized into three groups according to their TPC as low (< 650 mg GAE/100 g FW), medium (650–850 mg GAE/100 g FW), and high (> 850 mg GAE/100 g FW); and samples were categorized into two groups according to their AC as low (< 650 mg TE/100 g FW), and high (> 650 mg TE/100 g FW). For Chi-square test, Monte Carlo simulations (10,000 replicates) were performed. To run LFMM, the function `lfmm` available in the R package LEA (version 1.6.0) was executed with two as the number of latent factors, ten repetitions, and 20 k total and 10 k burnin iterations in the Gibbs

Sampling algorithm. All other parameters were set to their default values. Samβada (version 0.5.3) was executed with DIMMAX equal to 3, SAVETYPE set to “end best”, SPATIAL set to “Longitude Latitude SPHERICAL NEAREST 2” and AUTOCORR “BOTH BOTH 1000”. A *p*-value of 0.01 was used to call significant associations identified by each approach. Only putative SNPs identified as significantly associated with at least one environmental or phenotypic variable by at least two approaches were used for further analyses.

Phylogenetic autocorrelation with environmental/phenotypic variables and principal coordinates analysis (PCoA)

To gather further support to the potential association of the putative SNPs deemed to be significantly associated with each variable, we conducted a Moran's I test and PCoA. For each variable, the pairwise genetic distances between samples taking into account only those SNPs found to be significantly associated with the given variable were calculated using the function `dist.gene` available in the R package ‘ape’ [30]. To calculate the Moran's I autocorrelation coefficient, samples were clustered using the BIONJ algorithm using the function `bionj` in ‘ape’, then their pairwise cophenetic distances were computed from the pairwise genetic distances using the `ape`'s function `cophenetic`, and finally `ape`'s function `Moran.I` was executed. A cutoff *p*-value of <0.05 was used to determine whether a given phenotypic or environmental variable had a significant phylogenetic autocorrelation. If the observed Moran's I value is significantly greater than the expected I value, then it is considered as positively autocorrelated, while if the observed Moran's I value is smaller than the expected I value, then it is negatively autocorrelated [30]. To further observe the discrimination of samples according to health traits and environmental variables as revealed by their associated SNPs, we conducted a PCoA using the function `dudi.pco` available in the R package `adegenet` (version 2.0.1) [31]. To perform PCoA, the pairwise genetic distances were converted to Euclidean distances using the `adegenet`'s function `quasieulid`. PCoA plots were created using the `adegenet`'s function `s.class` and `add.scatter.eig`.

Morales [32] evaluated the genetic similarities of these *Vaccinium vitis-idaea* samples using the consensus sequences of DNA reads towards the reconstruction of a Bayesian phylogenetic tree. He found three genetic groups in the tree showing a geographic structure according to ecoregion and temperature, with genetically close individuals also being geographically close.

Putative function of reads with SNPs

To obtain a putative functional annotation of the protein-coding genomic regions where putative SNPs were found,

we performed BLASTX (version 2.6.1+) searches against the NCBI Non-Redundant (NR) protein database using BLAST2GO [33] (version 4.1.7) with an E-value <0.001. The first step in BLAST2GO was to find proteins with the highest DNA sequence similarity to our query DNA sequence and retrieve them for further analysis. The next step was to provide a functional analysis of proteins by classifying them into families and predicting domains using the InterPro function of BLAST2GO. Mapping was performed to retrieve gene ontology (GO) terms associated to the hits obtained after the BLASTX search. GO terms assigned were restricted to the plant slim GO subset. Additionally, we searched for similar sequences in RNAcentral [34] (release 7), a comprehensive database of non-coding RNA sequences, using `nhmmer` [35] with an E-value <0.001 and all other parameters set to default values.

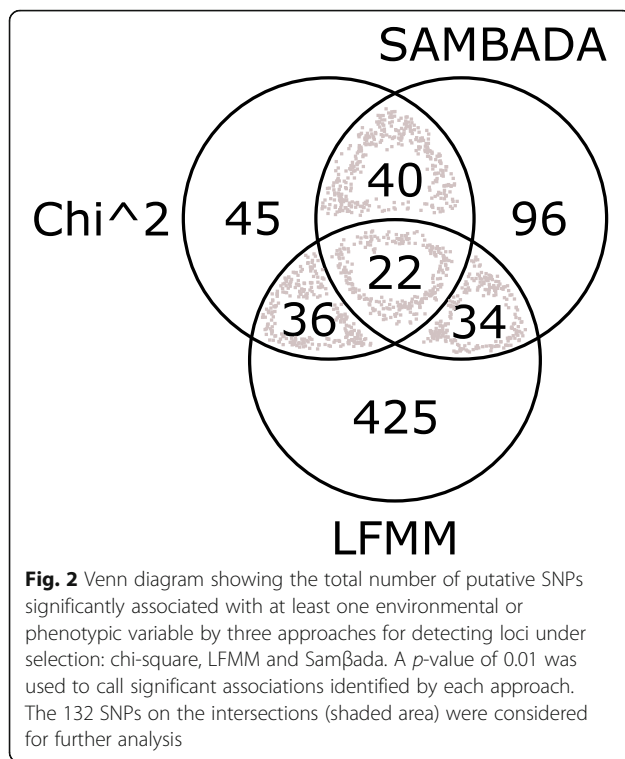
Results

Library sequencing, SNP discovery and selection

Sequencing of a 56-plex GBS library yielded about 142 million barcoded reads 100 bp long, corresponding to an average of 2,53 million reads per sample and ranging from 0,94 to 6,02 million reads per sample. GBS data is available at the NCBI with the BioProject accession number PRJNA377297. A quality analysis using FastQC reported an average Phred quality score of 34, and % GC content of 53. The TASSEL-UNEAK pipeline yielded 76,354 SNP loci putatively located at a single locus, with MAF of 0.05. Using a custom R script, a set of 1586 SNPs with a read depth coverage of at least 3X, genotyped in at least half the samples, and a minor allele frequency greater than 0.30, was identified and selected for further genetic characterization analysis. On average, there was $13.4 \pm 7.3\%$ of missing data per sample across the selected dataset of 1586 SNP loci.

Putative SNPs associated with phenotypic and environmental variables

The Chi-square test, LFMM, and Samβada identified 143, 517, and 192 SNPs respectively, significantly associated with phenotypic and/or environmental variables (Fig. 2). We found 132 SNPs significantly associated with at least one variable using at least two approaches. Of these 132, the maximum number of associated SNPs (53) was identified for the variable MAR, and the minimum (38) for the variable TPC (Fig. 3 and Additional file 1: Table S2). 102 (or 82%) of the 132 SNPs were correlated with more than one variable (Fig. 3 and Additional file 1: Table S3). On average, there were 17.4 ± 5.5 SNPs shared between each pair of variables. Surface water variables (SWS and SWp) shared the highest number of SNPs (37), while the phenotypic variables (TPC and AC) shared the lowest number of SNPs (10).



On average, 3 ± 2.1 SNPs were uniquely associated to one variable. The variable MAT has the largest number of SNPs (7) exclusively associated to it. The variables TPC and SWS have both only one SNP exclusively associated to them.

Phylogenetic autocorrelation with environmental/phenotypic variables and PCoA

Moran's *I* statistics indicated a statistically significant positive phylogenetic autocorrelation for all environmental and phenotypic variables supporting the results of the Chi-square, LFMF, and Samβada. Moran's *I* expected and observed values, and *p*-values are shown in Additional file 1: Table S4. PCoA plots showed that individuals collected from the same environmental condition or showing similar phenotype were genetically more similar when considering the SNPs associated with the corresponding variable. Fig. 4 illustrates this for MAT and AC. Principal coordinate values for each sample appear in Additional file 1: Table S5 and S6. PCoA plots for the remaining variables are provided in Additional file 2.

Putative function of reads with SNPs

To identify protein coding genes harbouring the 132 SNPs associated with environmental and phenotypic variables, the corresponding SNP-containing reads were used as query for BLASTX searches to the NCBI NR protein database. BLAST2GO was used for assigning a

functional annotation based on sequence homologies. Most of *V. vitis-idaea* sequences containing the 132 SNPs had no significant sequence alignment or hits in the NCBI NR database. The BLAST2GO annotation produced GO term assignments for 10 sequences out of the 14 genomic sequences with homologous proteins. These 14 genomic sequences with homologous proteins are listed in Table 1, and Fig. 5 indicates the GO term assignments obtained.

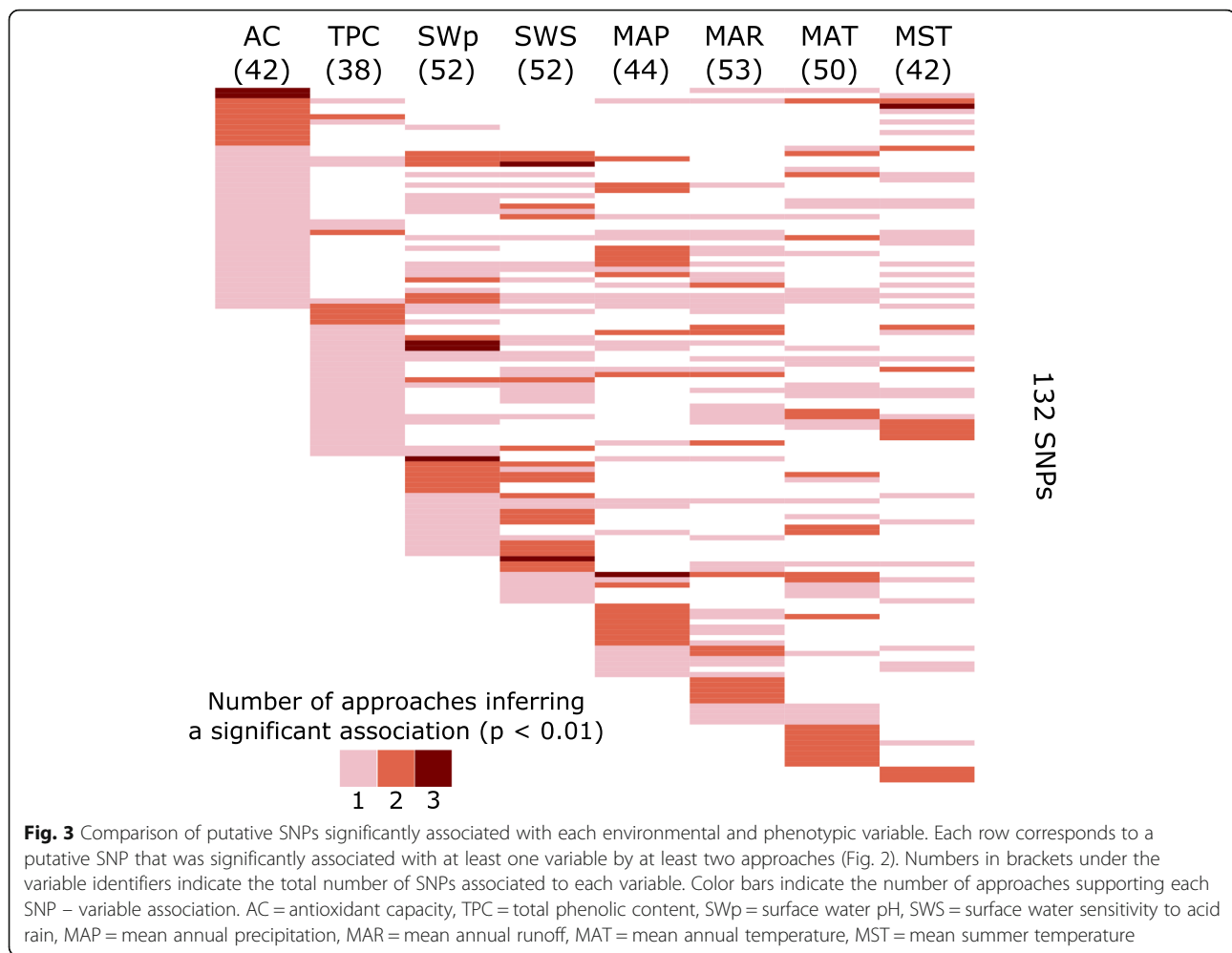
The gene ontology encompasses terms describing genes and gene product roles in cells [36]. These terms are in three categories: 1) biological process – referring to the biological objective of the genes or gene products; 2) cellular component – referring to the place in the cell where the gene product is active; and 3) molecular function – referring to the biochemical activity of the genes or gene products. In the biological process category, the most common GO term was biosynthetic process (4 proteins). The cellular component was dominated by membrane (4 proteins). The molecular function category was comprised of protein coding genes involved in compound binding, transferase activity, and DNA-binding. This GO analysis suggested that *V. vitis-idaea* SNP-containing sequences code for diverse proteins involved in structural, regulatory, and metabolic processes (Fig. 5).

As there were only 14 genomic sequences with homologous proteins, we additionally searched for homologous non-coding RNAs to the sequences containing the 132 SNPs in RNACentral. Two SNP-containing sequences have a homologous non-coding RNA: SNP ID TP14976 has significant sequence similarity to *Hevea brasiliensis* miscellaneous RNA URS0000A1EE58 (E-value = $8.3E-6$), and SNP ID TP52730 has significant sequence similarity to *Hevea brasiliensis* miscellaneous RNA URS0000A2A2E6 (E-value = $3.9E-5$). Both homologous non-coding RNAs lack functional annotation.

Discussion

Putative SNPs associated with phenotypic and environmental factors

Different methods to identify correlations between environmental variables and allele frequencies have their own advantages and weaknesses [37], and various factors can confound inferences of local adaptation [38]. To be more confident about our findings, we used three complementary approaches (i.e., Chi-square, LFMF, and Samβada) based on different principles. We interpreted SNPs found to be significantly associated with an environmental or phenotypic variable by only one of the approaches as spurious correlations (false positives); although, by doing this we are increasing the number of false negatives. As environmental variables may be highly correlated, approaches were not required to agree on the same associated variable (Fig. 3). The large proportion



(82%) of the 132 SNPs that were associated to more than one variable suggests that indeed variables are correlated.

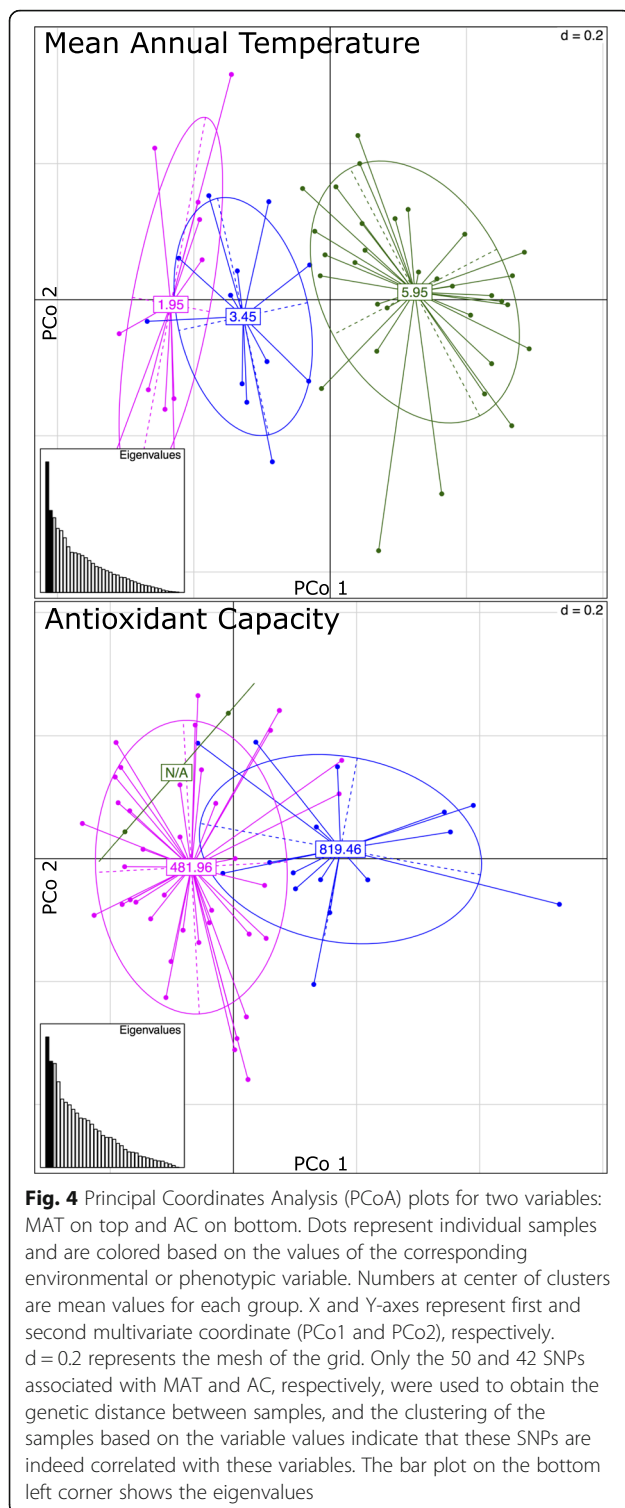
Our finding of only ten putative SNPs associated with both TPC and AC suggests that the genetic basis for the differential TPC and AC of fruit is not the same, and that the AC of fruit extracts depends not only on the specific phenolic compound profile but also on other non-phenolic molecules. This lack of positive correlation between TPC and AC has previously been reported in several studies [10, 39, 40]. Our understanding of how exactly phenolic compounds contribute to the antioxidant activity of different plant species is still limited. The contribution of each individual phenolic compound to the total AC may vary [41], thus variation in the specific phenolic compound profile can affect the overall AC. For example, the AC of different flavonoids depends on the number and position of hydroxyl substitutions on their backbone structure [42], and glycosylation might result in lower AC [43]. Plants possess different enzymatic (e.g. superoxide dismutase, catalase, ascorbate peroxidase) and non-enzymatic (e.g. glutathione, phenolic compounds, alkaloids, non-protein amino acids, and

α -tocopherols) antioxidant defense systems, which protect them from oxidative damage by scavenging reactive oxygen species [44].

Potential functions of putative SNPs associated with environmental and phenotypic variables

Among the 132 putative SNPs that were found associated with environmental or phenotypic variables, only 14 SNP-containing DNA sequences (10.6%) yielded hits in the NCBI NR protein database. These results suggest that most of these putative SNPs were located in unknown proteins (i.e., proteins not yet in the database) or in non-coding genomic regions. We discuss the gene identity and putative function of each of the 14 SNP-containing sequences as revealed from BLAST2GO and the literature search except for SNPs ID TP72119, TP75314, and TP71702 because their functional annotation rendered a protein of unknown function (Table 1).

The most significant blastx hits (95% identical) for SNP ID TP49117 corresponded to calmodulin-binding transcription activator 4 (CAMTA 4). In plants, the calcium ion is the most ubiquitous secondary messenger,



participating in the transduction of various signals such as cold, wind, touch, UV light, pathogen attack, and phytohormones [45]. The mechanisms of Ca^{2+} -dependent transcription regulation include various signal transducers such as CAMTAs. It is hypothesized that CAMTAs

are involved in biotic and abiotic stress responses by modulating hormone signalling [46]. More specifically for our study, CAMTA 4 positively regulates auxin transport and homeostasis, upregulates specific toxin metabolic processes, and together with CAMTA 5 and 6 regulates more than 1000 downstream genes linking environmental cues to growth responses [47]. This SNP-containing region might provide insight on CAMTA-based regulation of cold response, since it was associated with MAT.

The most significant blastx hits yielded from the sequence containing SNP ID TP72740 were annotated as probable galacturonosyltransferase 6 (GAUT). GAUTs are among the enzymes responsible for pectin and xylan biosynthesis in cells walls and seed testa [48]. Pectins are highly heterogeneous polymers that control cell wall porosity, cell adhesion, plant defense, and resistance to extracellular freezing [49, 50]. Cold acclimation in *Brassica napus* has been associated with increased cell wall thickness, modifications in leaf stiffness, cell wall and pectin contents, pectin methyltransferase activity, and the low-methylated pectin content [51–53]. Therefore, this correlated SNP with MAT and SWS is worth further evaluation especially because few pectin biosynthesis genes have been identified [48].

A protein similar to a receptor-like protein kinase (RLK) was annotated for an homologous sequence to SNP ID TP25921. In *Arabidopsis*, the RLK gene family contains more than 600 members, and it is the largest gene family in plants [54]. Many RLKs are known to be involved in several biological functions, including the regulation of plant development, growth, environmental stress response, or pathogen defense [55, 56]. Among the abiotic stress factors that induce RLKs are cold, drought, salt, toxic metals, and mineral deficiency [57]. Since this SNP was associated with TPC and several environmental variables (SWS, SWp, MAR, MAT, MST), and the fact that few RLKs have an experimentally verified function, we cannot hypothesize on a potential adaptive role of this particular locus until further biochemical and functional characterization are conducted.

The most significant blastx hits for SNP ID TP26200 were 100% identical to a segment (21 aminoacids) of glucose-6-phosphate/phosphate translocator (GPT). GPTs are involved in delivery of glucose 6-phosphate to plastids from non-green tissues as carbon sources for starch biosynthesis and/or to the oxidative pentose phosphate pathway [58]. Of the two GPT genes in *A. thaliana*, expression of GPT2 seems highly variable and sensitive to impaired carbon metabolism [59], sugar-induced senescence [60], phosphate starvation [61], and increases in carbon fixation due to increased light [62]. This latter study implied that GPT2 plays a role in sugar sensing either directly or indirectly by affecting the

Table 1 Functional annotations of 14 SNP-bearing sequences significantly associated with phenotypic and environmental variables

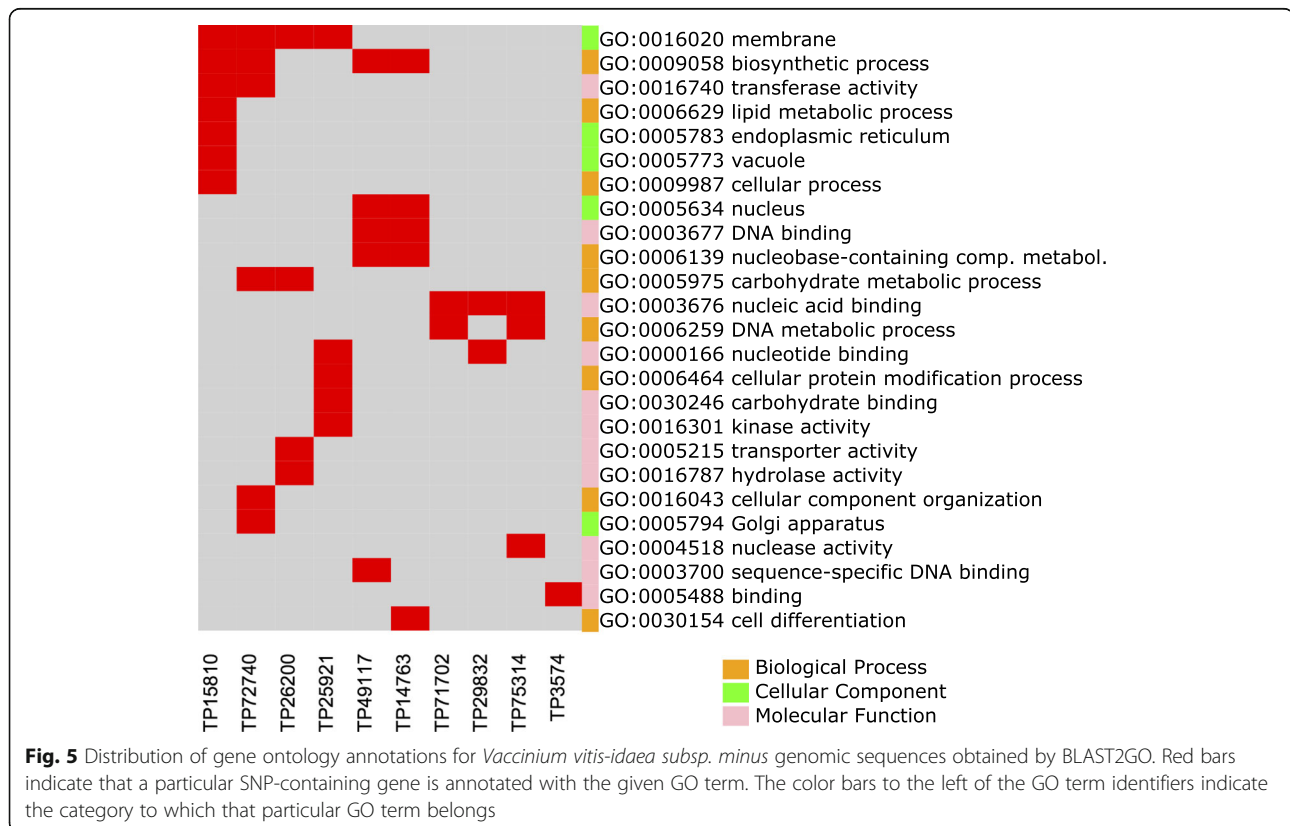
| SNP ID | Associated Environmental Variable | BLAST2GO Functional Annotation | E-value | % Identity |
|---------|-----------------------------------|---|------------|------------|
| TP3574 | MAR | Pentatricopeptide repeat-containing mitochondrial | 1.79042E-4 | 96 |
| TP14763 | TPC | Transcription factor (contains a MYB DNA-binding domain) | 3.94276E-5 | 93 |
| TP15810 | SWp, MAP, MAR | Phospholipid:diacylglycerol acyltransferase 1-like | 7.36315E-5 | 100 |
| TP25921 | TPC, SWp, SWS, MAR, MAT, MST | Leaf rust 10 disease-resistance locus receptor-like protein kinase-like | 1.57258E-5 | 86 |
| TP26200 | TPC, SWS | Glucose-6-phosphate phosphate translocator chloroplastic-like | 1.28322E-5 | 100 |
| TP29832 | MAP, AC | RNA-binding CP31B chloroplastic | 3.68209E-4 | 100 |
| TP45867 | SWp, SWS | DEAD-box ATP-dependent RNA helicase 37-like | 6.29099E-4 | 100 |
| TP49117 | MAT | Calmodulin-binding transcription activator 4 isoform X3 | 1.35641E-6 | 95 |
| TP60073 | MAT | Pre-mRNA-splicing factor ATP-dependent RNA helicase DEAH1-like | 6.21354E-4 | 100 |
| TP69971 | MAP, SWp, SWS | NAD(P) FAD-dependent oxidoreductase | 3.22832E-5 | 100 |
| TP71702 | MAR, SWS, TPC | Unnamed protein product, partial | 4.30509E-4 | 88.33 |
| TP72119 | AC | Uncharacterized protein | 3.3681E-5 | 100 |
| TP72740 | MAT, SWS | Galacturonosyltransferase | 3.34427E-6 | 100 |
| TP75314 | MST, SWp, SWS | Hypothetical protein VITISV_032719 | 2.16821E-4 | 93 |

Homologous sequences were retrieved by BLAST2GO using blastX with an E-value <0.001

balance of metabolites among cellular compartments [62]. GPT connection with TPC and SWS - the two variables associated with this SNP - is unclear.

The functional annotation of the homologous sequence to that containing SNP ID TP15810 revealed a phospholipid-diacylglycerol acyltransferase (PDAT). PDAT1 is a critical enzyme involved in triacylglycerol

(TAG) synthesis [63], a high-energy compound. In plants, TAGs are generally synthesized on specialized cells of developing seeds, fruit, and pollen; and they are stored in lipid droplets, which are mobile during active periods of metabolism [64, 65]. Aging and stress factors such as freezing and desiccation can induce accumulation of TAGs in plant tissues [66–69] but the role of



PDAT on this accumulation was not reported or was unknown since other enzymes like those of the Kennedy pathway can also lead to TAG biosynthesis. The connection between PDAT and the environmental variables that were associated to SNP ID TP15810 (SWp, MAP, and MAR) is therefore not evident.

The SNP ID TP69971 is located in a protein 100% identical to a segment (20 aminoacids) of ferredoxin–NADP(+) reductase in *Methylosinus*, *Methylocystis*, and *Paramesorhizobium*. These organisms are bacteria within the order Rhizobiales, an order known to contain nitrogen fixing bacteria in symbiotic relationship with plant roots [70] and genera like *Agrobacterium* known to have the capacity of transferring their DNA to their host plants [71]. We therefore, hypothesize that this SNP-containing locus might represent an old rhizobial DNA transfer to *V. vitis-idaea*. Ferredoxin–NADP(+) reductase (FNR) is an enzyme that participates in photosynthesis, but in non-photosynthetic organisms like bacteria, it provides reduced ferredoxin to other metabolic pathways including nitrogen fixation, terpenoid biosynthesis, steroid metabolism, oxidative stress response, and iron-sulfur protein biogenesis [72–74]. The link between this enzyme and the water-related environmental variables (MAP, SWp, SWS) is uncertain especially because plant dehydration did not induce significant changes in FNR activity in wheat [75].

A protein similar to a DEAD-box ATP dependent RNA helicase was annotated for an homologous sequence to that containing SNP ID TP45867. The DEAD-box RNA helicase family plays a role in DNA and RNA metabolism such as replication, repair, recombination, transcription, translation, ribosome biogenesis and splicing [76, 77]. These enzymes act in the growth and development of plants under stress by regulating stress-induced transcription and translation [78, 79]. For example, the overexpression of a gene in this family was associated with disease resistance and tolerance to oxidative stress [80, 81]. DEAD-box RNA helicases can regulate the function of transcripts involved in salt tolerance [82] and cold stress [81]. The role of these enzymes on alkalinity stress (SWp and SWS were the correlated variables), however, is unknown.

SNP ID TP29832 was associated with MAP and AC and the most significant blastx hits correspond to a chloroplast RNA binding protein 31B, also known as chloroplast ribonucleoprotein CP31 (cpRNP CP31). cpRNPs are nuclear encoded, highly abundant RNA binding proteins involved in chloroplast RNA processing and stabilization [83]. They react to external and internal signals, particularly light, but also cold stress by influencing multiple chloroplast RNA processing steps [84, 85]. It has also been shown that all or most cpRNP RNAs were repressed following attack by *Pseudomonas*,

deprivation of nitrate, hypoxia or osmotic stress [84]. The role of cpRNPs on drought tolerance or the synthesis of phenolics is unknown.

The most significant blastx hits (96% identical) for SNP ID TP3574 corresponded to a mitochondrial pentatricopeptide repeat (PPR) protein. PPR proteins belong to a large gene family of sequence-specific RNA-binding proteins that regulate gene expression at the post-transcriptional level mainly in organelles but also in the nucleus [86]. Their functions comprise a wide range of physiological and developmental processes including cytoplasmic male sterility, photosynthesis, respiration and embryogenesis [86, 87]. Few functional studies of PPR proteins in plants have revealed a biotic and abiotic stress response. For example, they can respond to salt, abscisic acid, oxidative stress, and pathogen infection [86, 88, 89]. More importantly for our case study, some PPR proteins in maize, sugarcane and *Arabidopsis* were upregulated or downregulated in response to drought stress [90, 91]. TP3574 therefore deserves further exploration in light of its association with runoff.

A protein similar to a pre-mRNA splicing factor ATP dependent RNA helicase from the DEAH box family was annotated for an homologous sequence to that containing SNP ID TP60073. As with all other RNA helicases, the DEAH box subfamily plays a crucial role in plant growth, development, and stress response. Since TP60073 was associated with MAT, the following examples of temperature dependent control of gene expression by DEAH-box proteins motivates further exploration of this locus. A DEAH-box protein in *Arabidopsis* was found to regulate the efficiency of pre-mRNA splicing, and mutants of this gene showed severe defects in hypocotyl dedifferentiation and de novo meristem formation in tissue culture under high temperature [92]. In yeast, a DEAH-box RNA helicase mutant showed cold sensitive growth defects and impaired RNA unwinding activity in vitro [93].

SNP ID TP14763 was associated with TPC and the most significant blastx hits correspond to proteins containing a Myeloblastosis (MYB) DNA-binding domain. MYB is a large gene family that controls several processes like responses to biotic and abiotic stress, nutrient deficiency, hormone responses, trichome development, cellular proliferation and differentiation, cell shape and petal morphogenesis, primary and secondary metabolism, and defense [94]. Most notably for our study, several MYB transcription factors (TFs) have been reported to regulate the phenylpropanoid pathway, which produces several important phenolic compounds such as flavonoids [95–97]. In *Vaccinium* species such as bilberry, highbush blueberry and bog bilberry, potential R2R3 MYB genes involved in flavonoid biosynthesis have been found [98–102]. This SNP-bearing DNA region is

worth of more in depth studies on the genetic changes related with different phenolic compound phenotypes. TFs are good targets for modifying complex traits in plants, thus we can expect that a technology based on TFs could lead the future of successful biotechnology crops [94]. The advent of transcriptome and genome databases for some *Vaccinium* species [99–101, 103–105] has revealed different families of TFs with potential roles in flavonoid biosynthesis. These databases, including ours, constitute an important resource to understand the signaling network involved in the regulation of these compounds' synthesis.

Antioxidant properties and phenolic compounds have been measured in some European and North American cultivars (e.g. [9, 106]). Overall, the antioxidant properties, total phenolic, flavonoids, and anthocyanin content of these cultivars are lower than those of wild clones, which have been explained by phenology-related physiological differences between them. Cultivars do not show a significant variation of antioxidant activity (<50 GAE mg/g FW), whereas wild clones can vary more (125–200 GAE mg/g FW) [9]. North American cultivars can vary significantly with respect to total anthocyanin, total phenolic, and total tannin content [106]. Cultivars must be added to this genomic assay for a more comprehensive discovery of health-associated SNPs. For these markers to be used for lingonberry breeding, the differential expression of these SNP-containing genes should be first tested in response to environmental conditions in order to better understand their regulation mechanism. The heritability mode of these health-associated SNPs should also be investigated. A genetic linkage map should be constructed in a standard reference population for identifying markers that are close to the target genes associated with healthy traits. As the success of a breeding program depends on how closely associated these markers are to the healthy traits, other methods to achieve this could include quantitative trait loci analysis, association mapping, classical mutant analysis, and bulk segregant analysis [107].

We acknowledge the current discussion on the utility of restriction site associated DNA sequencing such as RADseq or GBS to detect loci under selection [108–110] and recognize that very likely there are still undetected potential adaptive loci in *V. vitis-idaea*. While accounting for the extent of linkage disequilibrium in a study system is an important consideration, the alternative methods proposed by Lowry et al. [108] – transcriptome sequencing, exon capture, whole genome sequencing and pool-seq – are also subject to equivalent limitations and require a larger genomic resource investment which is not available for many non-model plant species such as *V. vitis-idaea*. While the 132 putative SNPs identified in this study could be useful for the selection of genotypes

with desirable climatic adaptations for breeding in a region with similar environmental conditions, further studies are needed to identify SNPs associated to abiotic and biotic stress. Finally, empirical validation of SNPs and the function of the annotated genes are required to corroborate the identified loci as adaptive.

Conclusions

This study generated the first published genomic resource for *V. vitis-idaea*, and highlighted promising functional markers for molecular breeding in this species. Further studies are needed to validate the identified 132 putative loci associated to environmental variables and phenotypic traits, and to decipher the molecular basis of their role in local adaptation.

Additional files

Additional file 1: Table S1. Sampled *V. vitis-idaea* sites in Newfoundland and Labrador, Canada. Each site is coded by ecoregion, mean annual temperature, mean summer temperature, mean annual precipitation, mean annual runoff, surface water pH, surface water sensitivity to acid rain, and proximity to the coast (Government of Newfoundland and Labrador 1992, and Alam et al. 2016). **Table S2.** Number of methods (LFMM, SamBada and Chi²) that predicted a significant association ($p < 0.01$) between each SNP and biochemical/environmental variable. SNPs are the ones inferred to be significantly associated with a variable by at least two methods. **Table S3.** Number of SNPs in common between each pair of biochemical/environmental variables. Only SNPs inferred to be significantly associated with at least one variable by at least two methods are considered. **Table S4.** Moran.I test results per variable. **Table S5.** Data visualized in Fig. 4 Top. **Table S6.** Data visualized in Fig. 4 Bottom. (XLS 62 kb)

Additional file 2: PCoA plots. (PDF 104 kb)

Abbreviations

AC: Antioxidant capacity; GBS: Genotyping-by-sequencing; MAP: Mean annual precipitation; MAR: Mean annual runoff; MAT: Mean annual temperature; MST: Mean summer temperature; NR: Non redundant; SNP: Single nucleotide polymorphism; SWp: Surface water pH; SWS: Surface water sensitivity to acid rain; TPC: Total phenolic content

Acknowledgements

The authors thank Brian Boyle for providing valuable insight on genotyping-by-sequencing, and Raul Mendez for field support.

Funding

Research and Development Corporation (RDC) of the Government of Newfoundland and Labrador (grant No. 5404.1722.101 to J.R.), and NSERC Discovery Grant (grant No. 402087–2011) to LPC. RDC and NSERC had no role in the design of the study, data collection, analysis, interpretation and in writing the manuscript.

Availability of data and materials

The GBS data supporting the conclusions of this study are available in the NCBI repository, <https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA377297>.

Authors' contributions

ZA conducted DNA extractions, data analysis, and wrote the manuscript. JR designed the research, obtained funding, helped with interpretation of results, and wrote the manuscript. LPC provided guidance on computational analysis of the GBS sequence data, conducted data analysis, and wrote the manuscript. All authors read and approved the final manuscript.

Authors' information

ZA was a MSc student at MUN interested on agronomy and horticulture under the supervision of JR and LPC. JR is a plant evolutionary biologist and LPC is a bioinformatician at MUN.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

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

Author details

¹Department of Biology, Memorial University of Newfoundland, St. John's, NL A1B 3X9, Canada. ²Department of Computer Science, Memorial University of Newfoundland, St. John's, NL A1B 3X5, Canada.

Received: 29 August 2017 Accepted: 19 December 2017

Published online: 02 January 2018

References

- Neto CC. Cranberry and blueberry: evidence for protective effects against cancer and vascular diseases. *Mol Nutr Food Res*. 2007;51(6):652–64.
- Takeshita M, Ishida Y, Akamatsu E, Ohmori Y, Sudoh M, Uto H, Tsubouchi H, Kataoka H. Proanthocyanidin from blueberry leaves suppresses expression of subgenomic hepatitis C virus RNA. *J Biol Chem*. 2009;284(32):21165–76.
- Mckay DL, Blumberg JB. Cranberries (*Vaccinium macrocarpon*) and cardiovascular disease risk factors. *Nutr Rev*. 2007;65(11):490–502.
- Wang L, Zhang XT, Zhang HY, Yao HY, Zhang H. Effect of *Vaccinium bracteatum* Thunb. leaves extract on blood glucose and plasma lipid levels in streptozotocin-induced diabetic mice. *J Ethnopharmacol*. 2010;130(3):465–9.
- Perez-Lopez FR, Haya J, Chedraui P. *Vaccinium macrocarpon*: an interesting option for women with recurrent urinary tract infections and other health benefits. *J Obstet Gynaecol Res*. 2009;35(4):630–9.
- Zheng W, Wang S. Oxygen radical absorbing capacity of phenolics in blueberries, cranberries, chokeberries, and lingonberries. *J Agric Food Chem*. 2003;51(2):502–9.
- Bakowska-Barczak AM, Marianchuk M, Kolodziejczyk P. Survey of bioactive components in western Canadian berries. *Can J Physiol Pharmacol*. 2007;85(11):1139–52.
- Debnath Samir C, Mathilde S. Genetic diversity, antioxidant activities, and anthocyanin contents in Lingonberry. *International Journal of Fruit Science*. 2009;9:185–99.
- Vyas P, Curran NH, Igamberdiev AU, Debnath SC. Antioxidant properties of lingonberry (*Vaccinium vitis-idaea* L.) leaves within a set of wild clones and cultivars. *Can J Plant Sci*. 2015;95(4):663–9.
- Alam Z, Morales HR, Roncal J. Environmental conditions affect phenolic content and antioxidant capacity of leaves and fruit in wild partridgeberry (*Vaccinium vitis-idaea*). *Botany*. 2016;94(7):509–21.
- Debnath SC. Inter simple sequence repeat (ISSR) to assess genetic diversity within a collection of wild lingonberry (*Vaccinium vitis-idaea* L.) clones. *Can J Plant Sci*. 2007;87(2):337–44.
- Balsdon JL, Smith TW, Lundholm JT. Phenotypic and genotypic differentiation of *Vaccinium vitis-idaea* between coastal barrens and forests in Nova Scotia, Canada. *Botany*. 2011;89(3):147–55.
- Ballesteros DC, Mason RE, Addison CK, Acuna MA, Arguello MN, Subramanian N, Miller RG, Sater H, Gbur EE, Miller D, Griffey CA, Barnett MD, Tucker D. Tolerance of wheat to vegetative stage soil waterlogging is conditioned by both constitutive and adaptive QTL. *Euphytica*. 2015;201(3):329–43.
- Raina SK, Govindasamy V, Kumar M, Singh AK, Rane J, Minhas PS. Genetic variation in physiological responses of mungbeans (*Vigna radiata* (L.) Wilczek) to drought. *Acta Physiol Plant*. 2016;38(11).
- Huang J, Gu M, Lai Z, Fan B, Shi K, Zhou Y, Yu J, Chen Z. Functional analysis of the *Arabidopsis* PAL gene family in plant growth, development, and response to environmental stress. *Plant Physiol*. 2010;153(4):1526–38.
- Ma D, Li Y, Zhang J, Wang C, Qin H, Ding H, Xie Y, Guo T. Accumulation of phenolic compounds and expression profiles of phenolic acid biosynthesis-related genes in developing grains of white, purple, and red wheat. *Front Plant Sci*. 2016;7:528.
- Cheng H, Yu O, Yu D. Polymorphisms of IFS1 and IFS2 gene are associated with isoflavone concentrations in soybean seeds. *Plant Sci*. 2008;175(4):505–12.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 2011;6(5):e19379.
- Lin M, Cai S, Wang S, Liu S, Zhang G, Bai G. Genotyping-by-sequencing (GBS) identified SNP tightly linked to QTL for pre-harvest sprouting resistance. *Theor Appl Genet*. 2015;128(7):1385–95.
- Romay MC, Millard MJ, Glaubitz JC, Peiffer JA, Swarts KL, Casstevens TM, Elshire RJ, Acharya CB, Mitchell SE, Flint-Garcia SA, McMullen MD, Holland JB, Buckler ES, Gardner CA. Comprehensive genotyping of the USA national maize inbred seed bank. *Genome Biol*. 2013;14(6):R55.
- Nunez GH, Rodriguez-Armenta HP, Darnell RL, Olmstead JW. Toward marker-assisted breeding for root architecture traits in southern highbush blueberry. *J Am Soc Hort Sci*. 2016;141(5):414–+.
- Burt L, Penhallegon R. Economic evaluation of lingonberry production in Oregon. <http://extension.oregonstate.edu/lane/sites/default/files/documents/em8847.pdf>. 2003, 1–27.
- Government of Newfoundland and Labrador. Water resources atlas of Newfoundland. http://www.canal.gov.nl.ca/reports/Water_Resources_Atlas/Water_Resources_Atlas_of_Newfoundland_1992.pdf. Water Resources Division, St. John's. 1992, 1–87.
- Poland J, Endelman J, Dawson J, Rutkoski J, Wu S, Manes Y, Dreisigacker S, Crossa J, Sanchez-Villeda H, Sorrells M, Jannink J. Genomic selection in wheat breeding using genotyping-by-sequencing. *Plant Genome*. 2012;5(3):103–13.
- Lu F, Lipka AE, Glaubitz J, Elshire R, Cherney JH, Casler MD, Buckler ES, Costich DE. Switchgrass genomic diversity, ploidy, and evolution: novel insights from a network-based SNP discovery protocol. *PLoS Genet*. 2013;9(1):e1003215.
- Maughan PJ, Smith SM, Fairbanks DJ, Jellen EN. Development, characterization, and linkage mapping of single nucleotide polymorphisms in the grain amaranths (*Amaranthus* sp.). *Plant Genome*. 2011;4(1):92–101.
- Griffiths AJF, Wessler SR, Carroll SB, Doebley J. An introduction to genetic analysis. 7th ed. New York: W. H. Freeman; 2010.
- Frichot E, Schoville SD, Bouchard G, Francois O. Testing for associations between loci and environmental gradients using latent factor mixed models. *Mol Biol Evol*. 2013;30(7):1687–99.
- Stucki S, Orozco-Terwengel P, Forester BR, Duruz S, Colli L, Masembe C, Negrini R, Landguth E, Jones MR, Bruford MW, Taberlet P, Joost S. High performance computation of landscape genomic models including local indicators of spatial association. *Molecular Ecology Resources*. 2016;17(5):1072–89.
- Paradis E. Analysis of phylogenetics and evolution with R: second ed. New York: Springer; 2012.
- Jombart T. Adegnet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*. 2008;24(11):1403–5.
- Morales HR. Population genomics of *Vaccinium vitis-idaea* L. and links to environmental conditions, total phenolics, and antioxidant capacity in Newfoundland and southern Labrador. Memorial University of Newfoundland, St. John's: MSc; 2017.
- Conesa A, Gotz S, Garcia-Gomez J, Terol J, Talon M, Robles M. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics*. 2005;21(18):3674–6.
- Petrov AI, SJE K, Kalvari I, Howe KL, Gray KA, Bruford EA, Kersey PJ, Cochrane G, Finn RD, Bateman A, Kozomara A, Griffiths-Jones S, Frankish A, Zwiab CW, Lau BY, Williams KP, Chan PP, Lowe TM, Cannone JJ, Gutell RR, Machnicka MA, Bujnicki JM, Yoshihama M, Kenmochi N, Chai B, Cole JR, Szymanski M, Karlowski WM, Wood V, Huala E, Berardini TZ, Zhao Y, Chen R, Zhu W, Paraskevopoulou MD, Vlachos IS, Hatzigeorgiou AG, Ma L, Zhang Z, Puetz J, Stadler PF, Mc Donald D, Basu S, Fey P, Engel SR, Cherry JM, Volders P, Mestdagh P, Wower J, Clark M, Quek XC, Dinger ME. RNAcentral Consortium, SILVA Team: RNAcentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res*. 2017;45(D1):D128–34.

35. Wheeler TJ, Eddy SR. Nhmmer: DNA homology search with profile HMMs. *Bioinformatics*. 2013;29(19):2487–9.
36. Ashburner M, Ball C, Blake J, Botstein D, Butler H, Cherry J, Davis A, Dolinski K, Dwight S, Eppig J, Harris M, Hill D, Issel-Tarver L, Kasarskis A, Lewis S, Matese J, Richardson J, Ringwald M, Rubin G, Sherlock G. Gene ontology: tool for the unification of biology. *Nat Genet*. 2000;25(1):25–9.
37. De Mita S, Thuillet A, Gay L, Ahmadi N, Manel S, Ronfort J, Vigouroux Y. Detecting selection along environmental gradients: analysis of eight methods and their effectiveness for outbreeding and selfing populations. *Mol Ecol*. 2013;22(5):1383–99.
38. Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC. Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *Am Nat*. 2016;188(4):379–97.
39. Rumbaoa RGO, Cornago DF, Geronimo IM. Phenolic content and antioxidant capacity of Philippine sweet potato (*Ipomoea batatas*) varieties. *Food Chem*. 2009;113(4):1133–8.
40. Yuan W, Zhou L, Deng G, Wang P, Creech D, Li S. Anthocyanins, phenolics, and antioxidant capacity of *Vaccinium* L. in Texas, USA. *Pharmaceut. Crops*. 2011;6:11–23.
41. Borges G, Degeneve A, Mullen W, Crozier A. Identification of flavonoid and phenolic antioxidants in black currants, blueberries, raspberries, red currants, and cranberries. *J Agric Food Chem*. 2010;58(7):3901–9.
42. Cao G, Sofic E. Prior R: antioxidant and prooxidant behavior of flavonoids: structure-activity relationships. *Free Radic Biol Med*. 1997;22(5):749–60.
43. Fukumoto L, Mazza G. Assessing antioxidant and prooxidant activities of phenolic compounds. *J Agric Food Chem*. 2000;48(8):3597–604.
44. Gill SS, Tuteja N. Reactive oxygen species and antioxidant machinery in abiotic stress tolerance in crop plants. *Plant Physiol Biochem*. 2010;48(12):909–30.
45. Zhu J. Cell signaling under salt, water and cold stresses. *Curr Opin Plant Biol*. 2001;4(5):401–6.
46. Shen C, Yang Y, Du L, Wang H. Calmodulin-binding transcription activators and perspectives for applications in biotechnology. *Appl Microbiol Biotechnol*. 2015;99(24):10379–85.
47. Galon Y, Snir O, Fromm H. How calmodulin binding transcription activators (CAMTAs) mediate auxin responses. *Plant Signal Behav*. 2010;5(10):1311–4.
48. Caffall KH, Pattathil S, Phillips SE, Hahn MG, Mohnen D. *Arabidopsis thaliana* T-DNA mutants implicate GAUT genes in the biosynthesis of pectin and xylan in cell walls and seed testa. *Mol Plant*. 2009;2(5):1000–14.
49. Ashworth E, Abeles F. Freezing behavior of water in small pores and the possible role in the freezing of plant-tissues. *Plant Physiol*. 1984;76(1):201–4.
50. Vorwerk S, Somerville S, Somerville C. The role of plant cell wall polysaccharide composition in disease resistance. *Trends Plant Sci*. 2004;9(4):203–9.
51. Stefanowska M, Kuras M, Kubacka-Zebalska M, Kacperska A. Low temperature affects pattern of leaf growth and structure of cell walls in winter oilseed rape (*Brassica napus* L., var. *oleifera* L.). *Ann Bot*. 1999;84(3):313–9.
52. Kubacka-Zebalska M, Kacperska A. Low temperature-induced modifications of cell wall content and polysaccharide composition in leaves of winter oilseed rape (*Brassica napus* L., var. *oleifera* L.). *Plant Sci*. 1999;148(1):59–67.
53. Solecka D, Zebrowski J, Kacperska A. Are pectins involved in cold acclimation and de-acclimation of winter oil-seed rape plants? *Ann Bot*. 2008;101(4):521–30.
54. Shiu S, Bleecker A. Receptor-like kinases from *Arabidopsis* form a monophyletic gene family related to animal receptor kinases. *Proc Natl Acad Sci U S A*. 2001;98(19):10763–8.
55. Afzal AJ, Wood AJ, Lightfoot DA. Plant receptor-like serine threonine kinases: roles in signaling and plant defense. *Mol Plant-Microbe Interact*. 2008;21(5):507–17.
56. Zhang YX, Chen L. Overexpression of the receptor-like kinase gene OsNRRB enhances drought-stress tolerance in rice. *Euphytica*. 2017;213(4):UNSP 86.
57. Ye Y, Ding Y, Jiang Q, Wang F, Sun J, Zhu C. The role of receptor-like protein kinases (RLKs) in abiotic stress response in plants. *Plant Cell Rep*. 2017;36(2):235–42.
58. Kammerer B, Fischer K, Hilpert B, Schubert S, Gutensohn M, Weber A, Flugge U. Molecular characterization of a carbon transporter in plastids from heterotrophic tissues: the glucose 6-phosphate phosphate antiporter. *Plant Cell*. 1998;10(1):105–17.
59. Kunz HH, Haeusler RE, Fettke J, Herbst K, Niewiadomski P, Gierth M, Bell K, Steup M, Fluegge U, Schneider A. The role of plastidial glucose-6-phosphate/phosphate translocators in vegetative tissues of *Arabidopsis thaliana* mutants impaired in starch biosynthesis. *Plant Biol*. 2010;12:115–28.
60. Pourtau N, Jennings R, Pelzer E, Pallas J, Wingler A. Effect of sugar-induced senescence on gene expression and implications for the regulation of senescence in *Arabidopsis*. *Planta*. 2006;224(3):556–68.
61. Hammond J, Bennett M, Bowen H, Broadley M, Eastwood D, May S, Rahn C, Swarup R, Woolaway K, White P. Changes in gene expression in *Arabidopsis* shoots during phosphate starvation and the potential for developing smart plants. *Plant Physiol*. 2003;132(2):578–96.
62. Athanasiou K, Dyson BC, Webster RE, Johnson GN. Dynamic acclimation of photosynthesis increases plant fitness in changing environments. *Plant Physiol*. 2010;152(1):366–73.
63. Fan J, Yan C, Zhang X, Xu C. Dual role for phospholipid: diacylglycerol acyltransferase: enhancing fatty acid synthesis and diverting fatty acids from membrane lipids to triacylglycerol in *Arabidopsis* leaves. *Plant Cell*. 2013;25(9):3506–18.
64. Hu Q, Sommerfeld M, Jarvis E, Ghirardi M, Posewitz M, Seibert M, Darzins A. Microalgal triacylglycerols as feedstocks for biofuel production: perspectives and advances. *Plant J*. 2008;54(4):621–39.
65. Zienkiewicz K, Castro AJ, De dios Alche J, Zienkiewicz A, Suarez C, Isabel Rodriguez-Garcia M. Identification and localization of a caleosin in olive (*Olea europaea* L.) pollen during in vitro germination. *J Exp Bot*. 2010;61(5):1537–46.
66. Elhafid L, Thi A, Zuilyfodil Y, Dasilva J. Enzymatic breakdown of polar lipids in cotton leaves under water-stress. 1. Degradation of monogalactosyl-diacylglycerol. *Plant Physiol Biochem*. 1989;27(4):495–502.
67. Yang Z, Ohlrogge JB. Turnover of fatty acids during natural senescence of *Arabidopsis*, *Brachypodium*, and switchgrass and in *Arabidopsis* Beta-oxidation mutants. *Plant Physiol*. 2009;150(4):1981–9.
68. Moellering ER, Muthan B, Benning C. Freezing tolerance in plants requires lipid remodeling at the outer chloroplast membrane. *Science*. 2010;330(6001):226–8.
69. Gasulla F, vom Dorp K, Dombink I, Zaehring U, Gisch N, Doermann P, Bartels D. The role of lipid metabolism in the acquisition of desiccation tolerance in *Craterostigma plantagineum*: a comparative approach. *Plant J*. 2013;75(5):726–41.
70. Remigi P, Zhu J, Young JPW, Masson-Boivin C. Symbiosis within symbiosis: evolving nitrogen-fixing legume symbionts. *Trends Microbiol*. 2016;24(1):63–75.
71. Bourras S, Rouxel T, Meyer M. *Agrobacterium tumefaciens* gene transfer: how a plant pathogen hacks the nuclei of plant and nonplant organisms. *Phytopathology*. 2015;105(10):1288–301.
72. Ceccarelli EA, Arakaki AK, Cortez N, Carrillo N. Functional plasticity and catalytic efficiency in plant and bacterial ferredoxin-NADP(H) reductases. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics*. 2004;1698(2):155–65.
73. Rohrich R, Englert N, Troschke K, Reichenberg A, Hintz M, Seeber F, Balconi E, Aliverti A, Zanetti G, Kohler U, Pfeiffer M, Beck E, Jomaa H, Wiesner J. Reconstitution of an apicoplast-localised electron transfer pathway involved in the isoprenoid biosynthesis of plasmodium falciparum. *FEBS Lett*. 2005;579(28):6433–8.
74. Rodriguez RE, Lodeyro A, Poli HO, Zurbruggen M, Peisker M, Palatnik JF, Tognetti VB, Tschiersch H, Hajirezaei M, Valle EM, Carrillo N. Transgenic tobacco plants overexpressing chloroplastic ferredoxin-NADP(H) reductase display normal rates of photosynthesis and increased tolerance to oxidative stress. *Plant Physiol*. 2007;143(2):639–49.
75. Nikolaeva MK, Maevskaya SN, Shugaev AG, Bukhov NG. Effect of drought on chlorophyll content and antioxidant enzyme activities in leaves of three wheat cultivars varying in productivity. *Russ J Plant Physiol*. 2010;57(1):87–95.
76. Linder P, Lasko P, Ashburner M, Leroy P, Nielsen P, Nishi K, Schnier J, Slonimski P. Birth of the D-E-A-D box. *Nature*. 1989;337(6203):121–2.
77. Aubourg S, Kreis M, Lecharny A. The DEAD box RNA helicase family in *Arabidopsis thaliana*. *Nucleic Acids Res*. 1999;27(2):628–36.
78. Linder P, Jankowsky E. From unwinding to clamping - the DEAD box RNA helicase family. *Nat Rev Mol Cell Biol*. 2011;12(8):505–16.
79. Tuteja N, Tarique M, Banu MSA, Ahmad M, Tuteja R. *Pisum sativum* p68 DEAD-box protein is ATP-dependent RNA helicase and unique bipolar DNA helicase. *Plant Mol Biol*. 2014;85(6):639–51.

80. Li D, Liu H, Zhang H, Wang X, Song F. OsBIRH1, a DEAD-box RNA helicase with functions in modulating defence responses against pathogen infection and oxidative stress. *J Exp Bot.* 2008;59(8):2133–46.
81. Zhang X, Zhao X, Feng C, Liu N, Feng H, Wang X, Mu X, Huang L, Kang Z. The cloning and characterization of a DEAD-box RNA helicase from stress-responsive wheat. *Physiol Mol Plant Pathol.* 2014;88:36–42.
82. Nakamura T, Muramoto Y, Yokota S, Ueda A, Takabe T. Structural and transcriptional characterization of a salt-responsive gene encoding putative ATP-dependent RNA helicase in barley. *Plant Sci.* 2004;167(1):63–70.
83. Tillich M, Hardel SL, Kupsch C, Armbruster U, Delannoy E, Gualberto JM, Lehwark P, Leister D, Small ID, Schmitz-Linneweber C. Chloroplast ribonucleoprotein CP31A is required for editing and stability of specific chloroplast mRNAs. *Proc Natl Acad Sci U S A.* 2009;106(14):6002–7.
84. Ruwe H, Kupsch C, Teubner M, Schmitz-Linneweber C. The RNA-recognition motif in chloroplasts. *J Plant Physiol.* 2011;168(12):1361–71.
85. Kupsch C, Ruwe H, Gusewski S, Tillich M, Small I, Schmitz-Linneweber C. *Arabidopsis* chloroplast RNA binding proteins CP31A and CP29A associate with large transcript pools and confer cold stress tolerance by influencing multiple chloroplast RNA processing steps. *Plant Cell.* 2012;24(10):4266–80.
86. Liu J, Zhao J, Lu P, Chen M, Guo C, Xu Z, Ma Y. The E-subgroup pentatricopeptide repeat protein family in *Arabidopsis thaliana* and confirmation of the responsiveness PPR96 to abiotic stresses. *Front Plant Sci.* 2016;7:1825.
87. Koizuka N, Imai R, Fujimoto H, Hayakawa T, Kimura Y, Kohno-Murase J, Sakai T, Kawasaki S, Imamura J. Genetic characterization of a pentatricopeptide repeat protein gene, orf687, that restores fertility in the cytoplasmic male-sterile Kosena radish. *Plant J.* 2003;34(4):407–15.
88. Zsigmond L, Rigo G, Szarka A, Szekeley G, Oetvoes K, Darula Z, Medzihradsky KF, Koncz C, Koncz Z, Szabados L. *Arabidopsis* PPR40 connects abiotic stress responses to mitochondrial electron transport. *Plant Physiol.* 2008;146(4):1721–37.
89. Laluk K, AbuQamar S, Mengiste T. The *Arabidopsis* mitochondria-localized pentatricopeptide repeat protein PGN functions in defense against necrotrophic fungi and abiotic stress tolerance. *Plant Physiol.* 2011;156(4):2053–68.
90. Jiang S, Mei C, Liang S, Yu Y, Lu K, Wu Z, Wang X, Zhang D. Crucial roles of the pentatricopeptide repeat protein SOAR1 in *Arabidopsis* response to drought, salt and cold stresses. *Plant Mol Biol.* 2015;88(4–5):369–85.
91. Wei K, Han P. Pentatricopeptide repeat proteins in maize. *Mol Breed.* 2016;36(12):170.
92. Ohtani M, Demura T, Sugiyama M. *Arabidopsis* ROOT INITIATION DEFECTIVE1, a DEAH-box RNA Helicase involved in pre-mRNA splicing, is essential for plant development. *Plant Cell.* 2013;25(6):2056–69.
93. Schneider S, Campodonico E, Schwer B, Motifs IV. V in the DEAH box splicing factor Prp22 are important for RNA unwinding, and helicase-defective Prp22 mutants are suppressed by Prp8. *J Biol Chem.* 2004;279(10):8617–26.
94. Ambawat S, Sharma P, Yadav NR, Yadav RC. MYB transcription factor genes as regulators for plant responses: an overview. *Physiol Mol Biol Plants.* 2013;19(3):307–21.
95. Moyano E, MartinezGarcia J, Martin C. Apparent redundancy in Myb gene function provides gearing for the control of flavonoid biosynthesis in antirrhinum flowers. *Plant Cell.* 1996;8(9):1519–32.
96. Dubos C, Le Gourrierec J, Baudry A, Huet G, Lanet E, Debeaujon I, Routaboul J, Alboresi A, Weisshaar B, Lepiniec L. MYBL2 is a new regulator of flavonoid biosynthesis in *Arabidopsis thaliana*. *Plant J.* 2008;55(6):940–53.
97. Hichri I, Barrieu F, Bogs J, Kappel C, Delrot S, Luviergeat V. Recent advances in the transcriptional regulation of the flavonoid biosynthetic pathway. *J Exp Bot.* 2011;62(8):2465–83.
98. Jaakola L, Poole M, Jones MO, Kamarainen-Karppinen T, Koskimäki JJ, Høhtola A, Haggman H, Fraser PD, Manning K, King GJ, Thomson H, Seymour GB. A SQUAMOSA MADS box gene involved in the regulation of anthocyanin accumulation in bilberry fruits. *Plant Physiol.* 2010;153(4):1619–29.
99. Li X, Sun H, Pei J, Dong Y, Wang F, Chen H, Sun Y, Wang N, Li H, Li Y. De novo sequencing and comparative analysis of the blueberry transcriptome to discover putative genes related to antioxidants. *Gene.* 2012;511(1):54–61.
100. Zifkin M, Jin A, Ozga JA, Zaharia L, Scherthner JP, Gesell A, Abrams SR, Kennedy JA, Constabel CP. Gene expression and metabolite profiling of developing highbush blueberry fruit indicates transcriptional regulation of flavonoid metabolism and activation of abscisic acid metabolism. *Plant Physiol.* 2012;158(1):200–24.
101. Gupta V, Estrada AD, Blakley I, Reid R, Patel K, Meyer MD, Andersen SU, Brown AF, Lila MA, Loraine AE. RNA-Seq analysis and annotation of a draft blueberry genome assembly identifies candidate genes involved in fruit ripening, biosynthesis of bioactive compounds, and stage-specific alternative splicing. *GigaScience.* 2015;4:5.
102. Primetta AK, Karppinen K, Riihinen KR, Jaakola L. Metabolic and molecular analyses of white mutant *Vaccinium* berries show down-regulation of MYBPA1-type R2R3 MYB regulatory factor. *Planta.* 2015;242(3):631–43.
103. Rowland LJ, Alkharouf N, Darwish O, Ogdan EL, Polashock JJ, Bassil NV, Main D. Generation and analysis of blueberry transcriptome sequences from leaves, developing fruit, and flower buds from cold acclimation through deacclimation. *BMC Plant Biol.* 2012;12:46.
104. Polashock J, Zelzion E, Fajardo D, Zalapa J, Georgi L, Bhattacharya D, Vorsa N. The American cranberry: first insights into the whole genome of a species adapted to bog habitat. *BMC Plant Biol.* 2014;14:165.
105. Sun H, Liu Y, Gai Y, Geng J, Chen L, Liu H, Kang L, Tian Y, Li Y. De novo sequencing and analysis of the cranberry fruit transcriptome to identify putative genes involved in flavonoid biosynthesis, transport and regulation. *BMC Genomics.* 2015;16:652.
106. Lee J, Finn CE. Lingonberry (*Vaccinium vitis-idaea* L.) grown in the Pacific northwest of North America: anthocyanin and free amino acid composition. *J Funct Food.* 2012;4(1):213–8.
107. Jiang G. Molecular markers and marker-assisted breeding in plants. InTech, Chapters: In Plant breeding from laboratories to fields. Edited by Andersen SB; 2013.
108. Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storer A. Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Mol Ecol Resour.* 2017;17(2):142–52.
109. McKinney GJ, Larson WA, Seeb LW, Seeb JE. RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on breaking RAD by Lowry et al. (2016). *Mol Ecol Resour.* 2017;17(3):356–61.
110. Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW. Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations. *Mol Ecol Resour.* 2017;17(3):362–5.

Submit your next manuscript to BioMed Central and we will help you at every step:

- We accept pre-submission inquiries
- Our selector tool helps you to find the most relevant journal
- We provide round the clock customer support
- Convenient online submission
- Thorough peer review
- Inclusion in PubMed and all major indexing services
- Maximum visibility for your research

Submit your manuscript at
www.biomedcentral.com/submit

