ORIGINAL PAPER

Complete plastid genome sequence of *Vaccinium macrocarpon*: structure, gene content, and rearrangements revealed by next generation sequencing

Diego Fajardo · Douglas Senalik · Mercedes Ames · Huayu Zhu · Shawn A. Steffan · Rebecca Harbut · James Polashock · Nicholi Vorsa · Emily Gillespie · Kathy Kron · Juan E. Zalapa

Abstract The complete plastid genome sequence of the American cranberry (*Vaccinium macrocarpon* Ait.) was reconstructed using next-generation sequencing data by in silico procedures. We used Roche 454 shotgun sequence data to isolate cranberry plastid-specific sequences of "HyRed" via homology comparisons with complete sequences from several species available at the National Center for Biotechnology Information database. Eleven cranberry plastid contigs were selected for the construction of the plastid genome-based homologies and on raw reads flowing through contigs and connection information. We assembled and annotated a cranberry plastid genome (82,284 reads; 185x coverage) with a length of 176 kb and the typical structure found in plants, but with several structural rearrangements in the large single-copy region when compared to other plastid asterid genomes. To evaluate the reliability of the sequence data, phylogenetic analysis of 30 species outside the order Ericales (with 54 genes) showed *Vaccinium* inside the clade *Asteridae*, as reported in other studies using single genes. The cranberry plastid genome sequence will allow the accumulation of critical data useful for breeding and a suite of other genetic studies.

Keywords Chloroplast · 454 Sequencing · Pyrosequencing · Ericales · *Vaccinium* · American cranberry

D. Senalik · S. A. Steffan · J. E. Zalapa (✉)
USDA-ARS, Vegetable Crops Research Unit, USDA,
1575 Linden Drive,
Madison, WI 53706, USA
e-mail: jezalapa@wisc.edu

D. Fajardo · M. Ames · H. Zhu · R. Harbut
Department of Horticulture, University of Wisconsin,
1575 Linden Drive,
Madison, WI 53706, USA

J. Polashock
Genetic Improvement of Fruits and Vegetables Laboratory,
USDA-ARS,
Chatsworth, NJ, USA

N. Vorsa
Blueberry and Cranberry Research and Extension Center,
Rutgers University,
Chatsworth, NJ, USA

E. Gillespie
Department of Biology, Marshall University,
Huntington, WV, USA

K. Kron
Department of Biology, Wake Forest University,
Winston-Salem, NC, USA

Introduction

Plastids are organelles of prokaryotic origin that contain the biochemical machinery necessary to replicate and transcribe their own genomes. The plastid genome of higher plants is a circular molecule of double-stranded DNA ranging from 72 to 217 kb in size and containing about 130 genes, depending on the plant species (Sugiura 1992, 1995; Tangphatsornruang et al. 2010). Gene structure, content, and order are often conserved, but as more complete plastid genomes are sequenced, it is clear that initial assumptions of the conserved nature of the genome need to be reevaluated. Most plastid genomes possess two identical inverted repeat regions (IR) separating two unequally sized single-copy regions termed the large single-copy (LSC) and small single-copy (SSC) regions (Jansen et al. 2011). Substantial variation in intergenic spacer regions and regulatory sequences has been reported across the angiosperms providing new insights into the diversity and phylogenetic relationships of flowering plants (Jansen et al. 2011). Plastids also have immediate practical uses in agriculture and horticulture because they can be engineered for specific gene function (Maliga and Svab 2011) and can be used for DNA barcoding (Fazekas et al. 2008). Approximately 265 complete plastid genome sequences of green plants are currently deposited at the National Center for Biotechnology Information (NCBI) including many agricultural species, but no plastid genome has been reported for important fruit crops in the genus *Vaccinium* which include species such as cranberry (*Vaccinium macrocarpon* Ait.) and blueberry (e.g., *Vaccinium angustifolium* Ait. and *Vaccinium corymbosum* L.).

The cranberry is a native North American diploid ($2x=2n=24$), perennial, woody plant species in the Ericaceae family (Kron et al. 2002). *V. macrocarpon* is adapted to temperate climates and largely confined to acidic and peat soils at latitudes about 39° in the Northern Hemisphere (Vander Kloet 1988). Cranberries are very popular for their tart flavor and nutritional and medicinal attributes (Eck 1990). Cranberries have been popular as a source of vitamin C, dietary fiber, and the essential dietary mineral manganese (Eck 1990). More recently, cranberries have been consumed as a supplement due the potential benefits of inherent flavonoid antioxidants (e.g., flavonols, anthocyanins, and proanthocyanidins) and other phytochemicals to the cardiovascular system, immune system, urinary tract, digestive system, teeth, and as general, anticancer (Wang and Stretch 2001; Kalt 2002). In 2011, cranberry production contributed over $500 million to the US economy and over 75 % of the worldwide cranberry production (USDA-NASS 2011). Although cranberry is a commercially important fruit crop, molecular genetics and genomic information available for this species are extremely limited (Zalapa et al. 2012; Zhu et al. 2012). Of the 39,000 sequences available for Ericaceae at the NCBI, approximately 23,000 are from 73 species of *Vaccinium*. Of these, *V. macrocarpon* has only 69 reported sequences.

The sequencing of the cranberry plastid genome will provide the necessary genetic information to begin exploring adaptive and acclimation processes in cranberry. In the field, cranberry photosynthesis exhibits midday depression which may limit biomass accumulation (Kumudini 2004). The cause of midday depression is unknown, but the intensity of solar radiation and temperature likely play a major role (Kumudini 2004). Interspecific variation in the plastid has been reported for Rubisco in species with persistent leaves and high photosynthetic ability such as cranberry (Galmes et al. 2005; Hikosaka and Shigeno 2009). Additionally, plant cold acclimation, freezing tolerance, and cold regulatory mechanisms related to photosynthesis may be located in the plastid (Clarke and Daniell 2011).

Because of its economic, agricultural, and medical importance, we investigated and assembled the complete plastid genome sequence of cranberry ("HyRed;" McCown and Zeldin 2003) using 454 sequencing data and in silico approaches. We present the organization of the cranberry plastid and genome comparisons and phylogenetic relationship with other angiosperms.

Materials and methods

Plastid genome sequence isolation and assembly

The DNA isolation, library preparation, emulsion PCR, and pyrosequencing on the Roche 454 platform were performed as described by Zhu et al. (2012). Four hundred fifty-four raw reads were assembled using gsAssembler (Newbler version 2.6; Roche 454 Life Sciences, Branford, CT) and were used to create FASTA and FASTQ files for all contigs of "HyRed". For the assembly, we used the default settings for the minimum overlap identity, minimum overlap length, and seed length while selecting large or complex genome and heterozygotic mode settings and an expected depth of 180x. The expected depth was determined by the number of reads, average read length, and expected genome size within the asterids group (Supplemental Fig. 1). The minimum overlap identity and Scripts bb.454contignet and bb.454contiginfo (Iorizzo et al. 2012; http://www.vcru.wisc.edu/simonlab/sdata/software/) were applied and developed to visualize and determine the contig connections derived from the gsAssembler output files. We used MUMmer3.0 (Kurtz et al. 2004) to compare the homology of the assembled "HyRed" cranberry plastid genome with the most closely related complete plastid genome sequences available at NCBI. Genome annotation was performed using the ORF Finder tool (http://www.ncbi.nlm.nih.gov/gorf/gorf. html) and Dual Organellar GenoMe Annotator

(DOGMA; Wyman et al. 2004). To detect possible plastid genome rearrangements and further explore genome architecture, the data were analyzed through comparisons with other complete plastid published genomes using MAUVE v 2.0 genome alignment software (Darling et al. 2010). The number of rearrangements was calculated using GRIMM web server (Tesler 2002). Simple sequence repeats (SSRs) were identified using WebSat web software (Martins et al. 2009) with thresholds of at least ten repeat units for mononucleotides, five for dinucleotides, four for trinucleotides, and three for tetra-, penta- and hexanucleotides. Because of the 454 pyrosequencing limitation in determining the exact number of nucleotides in regions with homopolymers, ABI-Sanger sequencing was performed for selected coding regions to correct possible misidentification of genes as pseudogenes and misreading of homopolymers. Primers used for the Sanger sequence of homopolymers in these regions can be found in the Supplemental Table 1. In order to verify the order of the genes in the assembled plastid, we conducted long-range PCR. Primers were designed in coding regions and targeted to amplify from 4 to 12 kb. In total, 24 primer pairs were designed (Supplemental Table 2) and allowed verification of the gene order in the assembled plastid DNA (Supplemental Fig. 2).

Phylogenetic analysis

Complete sequences of specific genes from the cranberry plastid genome were compared to those from complete published plastid genomes using Basic Local Alignment Search Tool (BLAST) (Altschul et al. 1990). A dataset of 54 genes from 30 complete published plastid genomes was used to investigate the phylogenetic placement of the cranberry genome relative to other members of the eudicots. Sequences were aligned using MUSCLE software (Edgar 2004). The GTR + Γ + i model was determined with Modeltest v3.7 (Posada and Crandall 1998) and used maximum likelihood analyses conducted with the RAxML v. 7.0.4 (Stamatakis et al. 2008) server in the Swiss Institute of Bioinformatics (http://phylobench.vital-it.ch/raxml-bb) with 1,000 bootstrap runs. The accession number JQ_ 757046 was assigned to the complete cranberry plastid genome sequence after being submitted to NCBI.

Results

The general flow chart for the in silico construction of the plastid genome is presented in Fig. 1. A total of 1,670,156 cleaned reads, with adaptors and low quality reads removed, corresponding to 620 Mb of cranberry sequence were obtained from a single run of Roche 454 pyrosequencing. The average read length was 372 nucleotides long, and the

total sequence obtained covered an estimated 266.3 Mb. The assembly of the nucleotide sequence reads was performed by using Newbler gsAssembler v. 2.6 to obtain nonredundant contigs and singletons. Contigs related to the cranberry plastid sequence were selected using MUMmer 3.0 by aligning the assembled contigs to the complete plastid genome sequence from *Daucus carota* (GenBank number DQ_898156), one of the closest available published complete plastid genomes from the asterid clade, as a reference genome. Forty contigs were found to be homologous to the *Daucus* plastid. The longest contig (104,544 bases long) with the highest BLAST hits was selected as a starting point to reconstruct the cranberry plastid based on contig connections and flow through reads. The cranberry plastid consisted of 11 contigs derived from 82,284 reads (5 % of the total number of reads) with a 185x coverage. The large and small single-copy regions were represented by single contigs, while the inverted repeat regions were inferred by connecting nine small contigs. The reconstructed plastid genome was 176,037 bp long with the LSC region of 104,544 bp, two IR of 34,232 bp each, and the SSC region of 3,029 bp.

Gene annotation for protein-coding sequences, rRNAs and tRNAs, was performed manually and by using DOGMA web-based software. A total of 110 genes were annotated and classified based on their putative function and relative location in the plastid genome (i.e., if present in the inverted repeated, large or small single-copy regions; Table 1). The annotated sequences were classified as follows: 73 protein-coding genes, five hypothetical coding regions, 28 tRNAs, and four rRNAs (Table 1).

Initially, 20 genes possessed incomplete protein sequence or premature stop codons (13 were detected in the LSC region, and seven, in the IR region). After Sanger resequencing, four pseudogenes in the LSC region (atpB, cemA, rps4, and rps19) were corrected (homopolymer regions) yielding functional genes. Also, three LSC pseudogenes (ndhK, ycf15, and infA) possessed confirmed missing start or premature stop codons. Out of the six remaining LSC pseudogenes, three pseudogenes (accD, ycf2, and ndhI) possessed BLAST hits for less than 25 % of the expected protein length. The three remaining pseudogenes in the LSC region were either located in different fragments (rps12, two fragments 70 kb apart and atpF, three fragments covering 71 % of the gene) or had a functional and nonfunctional copy in the LSC (rps18).

The seven pseudogenes in the IR region were not Sanger resequenced due the difficulty in designing unique primers. However, three (rpl22, psbA, and rps3) out of the seven pseudogenes in the IR region had a functional copy of the gene in the LSC. The remaining IR pseudogenes were too short (ccsA, ndhG, and ycf68) to expect recovery of the whole protein, and rps16 pseudogene had a premature stop codon. The presence of homopolymers in the cranberry

Fig. 1 Scheme for in silico assembly, annotation, and construction of the cranberry (*V. macrocarpon*) plastid genome using next-generation sequencing data
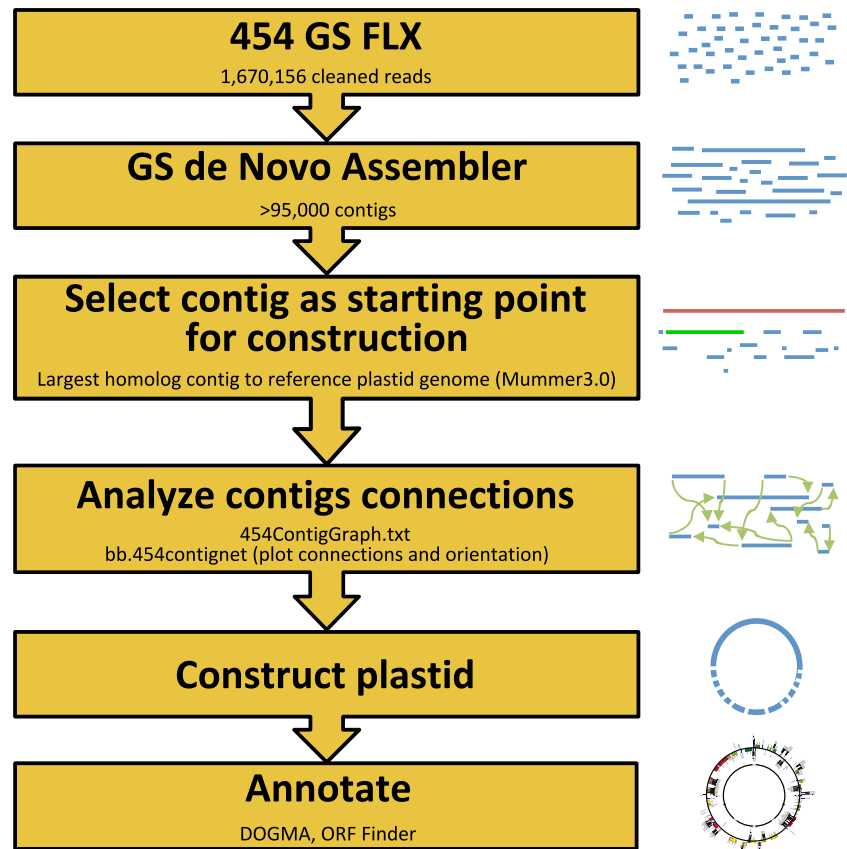
plastid genome sequence suggests that the size of the genome could be slightly larger due to the technical limitation of the 454 platform that tends to underestimate homopolymers. The final length of the plastid after combining the Sanger sequences with the 454 assembled data was 176,045 bp (Fig. 2). The raw reads used for assembly were mapped against the plastid cranberry genome to check if the premature stop codons were supported by multiple reads (Supplemental Fig. 1). The regions with the lowest coverage were located in intergenic regions. The coding regions showed high coverage discarding the fact that the premature stop codons were due to low coverage.

A total of 62 SSRs were identified which were distributed along the plastid genome (Table 2). Most of the SSRs were located in the LSC ($n=39$), followed by the IR ($n=22$) and the SSC ($n=1$). Ten SSRs were located within genes ($n=6$) or pseudogene ($n=4$) sequences. The remaining SSRs were located in intergenic regions.

The comparison of common genes from the cranberry plastid genome to those from complete published plastid genomes using BLAST (Altschul et al. 1990) showed "hits" with taxa from the asterid and rosid groups. Complete gene sequences were selected and aligned for further phylogenetic analyses. Data were subjected to phylogenetic analysis in RAxML with 54 protein coding genes from 31 taxa, including representatives from the asterid and rosid groups, plus

Zea mays, which was selected as outgroup. The sequence data comprised 53,965 aligned nucleotides with 14,383 parsimony-informative characters. The maximum likelihood results showed 100 % bootstrap support for all recovered nodes (Fig. 3), with cranberry placed in the asterid clade along with carrot (*D. carota*), sunflower (*Helianthus annuus*), blue throatwort (*Trachelium caeruleum*), and spinach (*Spinacia oleracea*).

The comparison of BLAST hits (minimum match length = 50; minimum blast percent similarity ≥ 70) and syntenic regions inside the LSC region, between the cranberry plastid against *D. carota* (GenBank accession NC_008325), *S. oleracea* (NC_002202), *H. annuus* (NC_007977), and *T. caeruleum* (NC_010442) showed multiple rearrangements in cranberry (Figs. 4 and 5), while other species from the asterid group showed a high degree of conservation except for *T. caeruleum* and *H. annuus*, which also showed a high number of rearrangements. The cranberry plastid genome was characterized by numerous inversions and other rearrangements not seen in other members of the asterids, but the position and size of the inverted repeats and LSC and SSC regions are conserved relative to most angiosperms (Rivas 2002). When specifically comparing these with *V. macrocarpon*, the observed rearrangements were considerably different (Figs. 4 and 5). We are confident that these rearrangements are not related to the assembly and construction of the plastid genome

Table 1 Genes found in the cranberry (*V. macrocarpon*) plastid genome

Type of gene	Gene names
Acetyl-CoA carboxylase	accD[a, b]
ATP synthase subunits	atpA[a], atpB[a], atpE[a], atpF[a, b], atpH[a], atpI[a]
Cytochrome b/f complex subunit	petA[a], petB[a], petD[a], petG[a], petL[a], petN[a]
Cytochrome c protein	ccsA[b, c]
Envelope membrane protein	cemA[a]
Maturase K	matK[a]
NADH-dehydrogenase subunits	ndhA[c], ndhB[a], ndhC[a], ndhD[c], ndhE[c], ndhF[a], ndhG[b, c], ndhH[c], ndhI[a, c, d], ndhJ[a], ndhK[a, b]
Photosystem I subunits	psaA[a, c, d], psaB[a], psaC[c], psaI[a], psaJ[a]
Photosystem II subunits	psbA[a, c], psbB[a], psbC[a], psbD[a], psbE[a], psbF[a], psbH[a], psbI[a], psbJ[a], psbK[a], psbL[a], psbM[a], psbN[a], psbT[a]
Ribosomal RNAs	rrn16[c], rrn23[c], rrn4.5[c], rrn5[c]
RNA polymerase subunits	rpoA[a], rpoB[a], rpoC1[a], rpoC2[a]
Rubisco large subunit	rbcL[a]
Small ribosomal protein units	rps11[a], rps12[a, b], rps14[a], rps15[c], rps16[b, c], rps18[a, d], rps19[a], rps2[a], rps3[a, c, d], rps4[a], rps7[a], rps8[a]
Large ribosomal protein units	rpl14[a], rpl16[a], rpl2[a], rpl22[a, c, d], rpl23[a], rpl32[c], rpl33[a], rpl36[a]
Hypothetical genes	ycf15[a, b], ycf2[a, b], ycf3[a], ycf4[a], ycf68[b, c]
Transfer RNAs	trnA-UGC[c], trnC-GCA[a], trnD-GUC[a], trnE-UUC[a], trnF-GAA[a], trnG-UCC[a], trnH-GUG[c], trnI-CAU[a], trnI-GAU[c], trnL-CAA[a], trnL-UAA[a], trnL-UAG[c], trnM-CAU[a], trnN-GUU[c], trnP-GGG[a], trnP-UGG[a], trnQ-UUG[a], trnR-ACG[c], trnR-UCU[a], trnS-GCU[a], trnS-GGA[a], trnS-UGA[a], trnT-GGU[a], trnT-UGU[a], trnV-GAC[a], trnV-UAC[a], trnW-CCA[a], trnY-GUA[a]
Other genes (translational initiation factor A and photosystem II protein Z, respectively)	infA[a, b], lhbA[a]

[a] Genes located in single-copy regions

[b] Pseudogene

[c] Genes located in the inverted repeats

[d] Two copies, one functional and one pseudogene

because all the reported rearrangements in Fig. 4 corresponded to the LSC and SSC regions which were obtained from a single contig.

Discussion

The availability of plastid genome sequences is increasing rapidly due to the use of next-generation sequencing approaches. Currently, the complete genomes of 265 green plants are publicly available at NCBI with most major angiosperm lineages represented. However, this is the first report of a complete genome in the genus *Vaccinium*. We used shotgun 454 sequence data and in silico approaches in a rapid procedure (Fig. 1), allowing the assembly of the cranberry plastid genome (Fig. 2). The use of next-generation sequencing technologies for the assembly of plastid genomes has been successfully applied in several plant species (Nock et al. 2011; Tangphatsornruang et al. 2010; Wang et al. 2011). The present work further demonstrates the use of emerging shotgun sequencing technologies for the rapid and cost-effective sequencing of complete plastid genomes of understudied but commercially important crop species such as cranberry without in-species or in-genus reference sequences, without the need of plastid DNA isolation, and using a minimal amount of 454 sequence data.

Due to the long and high-quality reads obtained and the minor effort required relative to other sequencing approaches, whole genome shotgun sequencing using the 454 technology is an efficient way to sequence whole plastid genomes in plants. Even though the 454 platform presents intrinsic limitations in identifying the exact number of bases in homopolymer regions, it was possible to reconstruct the complete plastid genome of cranberry (Fig. 2). Analysis of the connections between cranberry plastid contigs gave a single possible configuration with the LSC and SSC regions represented by single contigs (104,544 and 3,029 bp, respectively) and the IR regions represented by nine contigs. The order and orientation of contigs from the LSC region with contigs from the IR were verified by PCR and Sanger sequencing, and the order of the genes was verified by long-range PCR (Supplemental Table 2; Supplemental Fig. 2). Four pseudogenes (atpB, cemA, rps4, and rps19 located in the LSC region) were corrected

Fig. 2 Gene map of the plastid genome of cranberry (*V. macrocarpon*) cultivar "HyRed" (complete sequence, GenBank accession JQ_757046). The *inner circle* indicates the large single-copy region, inverted repeat regions, and the small single-copy region. Genes *inside and outside the outer circle* are transcribed in a counter-clockwise and clockwise direction, respectively. The genes are color coded based on their function
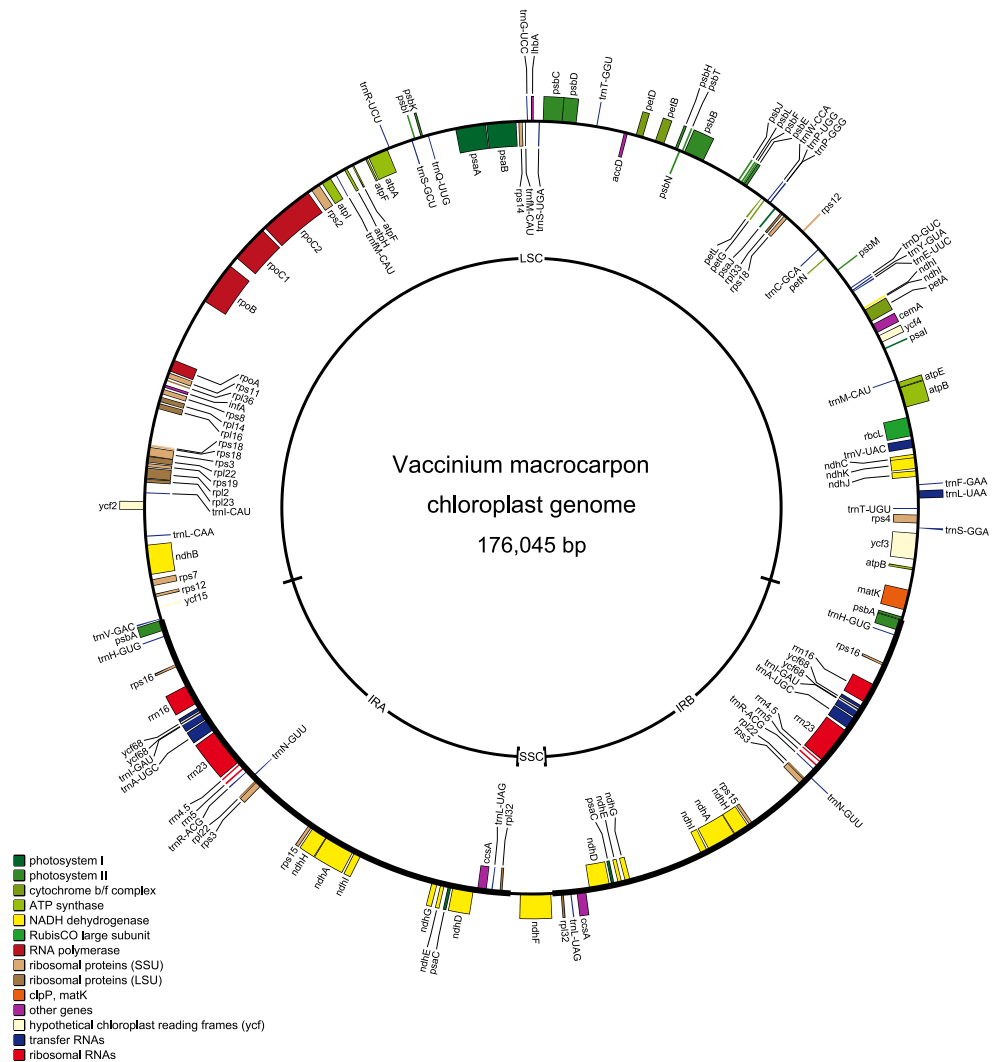
to functional genes using Sanger sequencing by adding missing nucleotides in homopolymer regions. No other differences were found between 454 and Sanger sequence data.

Previous phylogenetic work using plastid sequence from the atpB, matK, ndhF, and rbcL and genes placed the order

Table 2 Location and length distribution of cranberry (*V. macrocarpon*) plastid simple sequence repeats found in the large single-copy region, inverted repeat regions, and small single-copy region

Type of repeat	Repeats found	Number of motif repeats	LSC/IR/ SSC	Genes/ pseudo
Mono-	23	10–14	19/4/0	1/3
Di-	8	5–7	7/1/0	2/0
Tri-	6	4–5	6/0/0	1/0
Tetra-	16	3–4	6/9/1	1/1
Penta-	5	3	0/5/0	0/0
Hexa-	4	3	1/3/0	1/0

Repeated sequences were classified by type, based on the number of repeated nucleotides (mono-, di-, tri-, tetra-, penta-, and hexanucleotides)
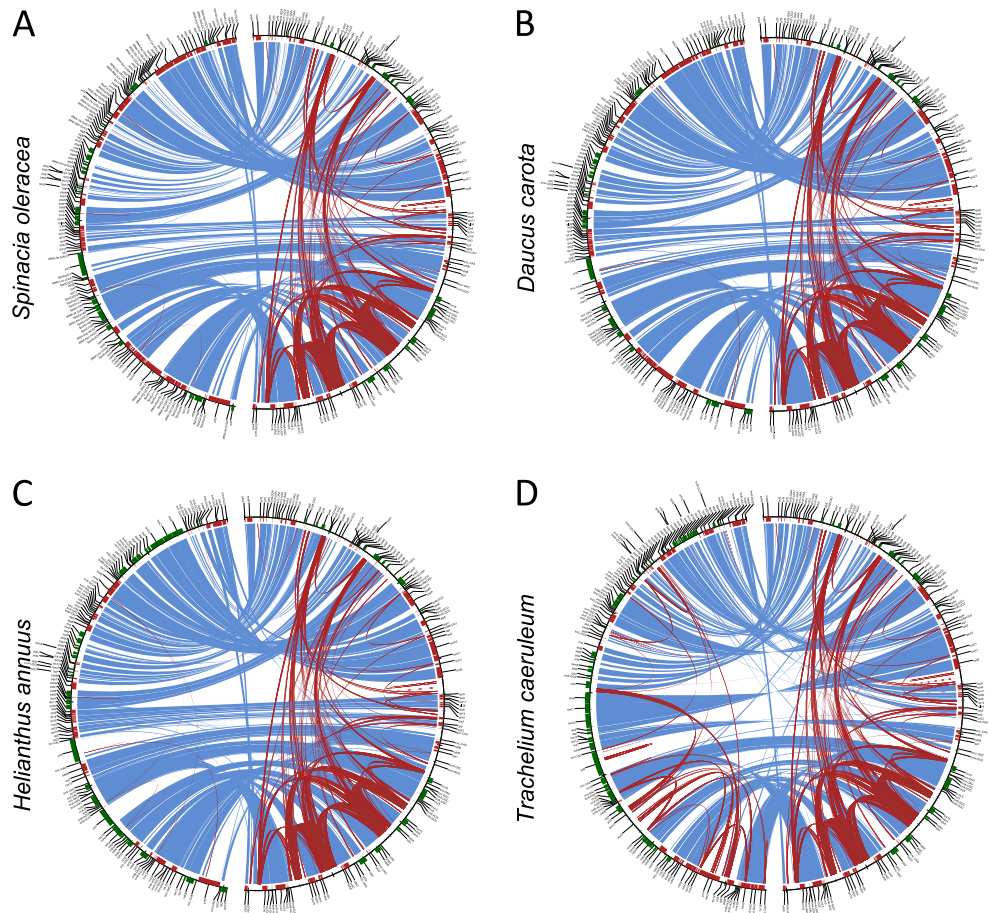
Ericales as a member of the clade *Asteridae* (Kron and Chase 1993; Kron and Judd 1997; Anderberg et al. 2002). Our study including 54 protein coding genes and 31 taxa (including representatives from the asterid and rosid groups) supports the hypothesis that *V. macrocarpon* belongs the *Asteridae* clade (Fig. 3). The phylograms supported monophyly for the order Ericales and resolved as a basal branch from the asterid group. No other taxon from the genus *Vaccinium* was included in the analysis since this is the first complete published plastid genome from subfamily Vaccinioidae. Additional sequencing of related species will have to be performed to corroborate the phylogenetic placement of the subfamily Vaccinioidae.

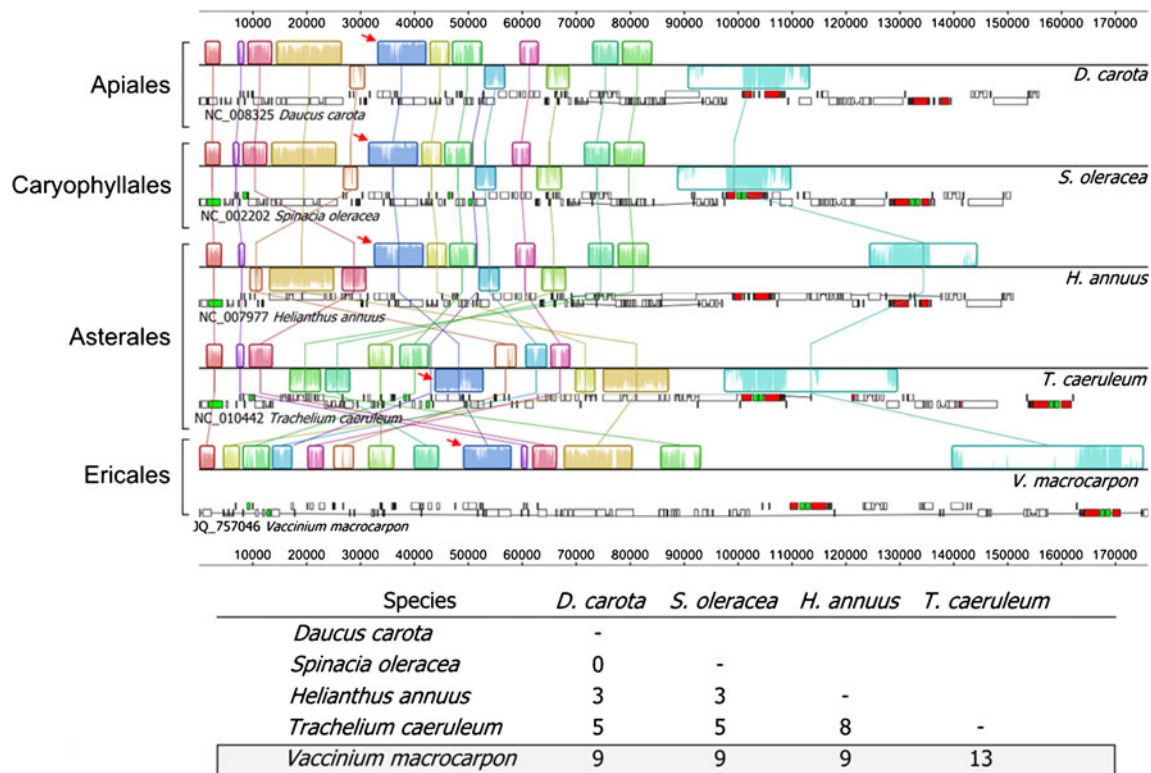Cranberry shared the same large-scale plastid architecture common among in angiosperms (Fig. 2). However, after alignment of syntenic regions, the cranberry plastid genome showed significant rearrangements when compared to the other complete asterid plastid genomes such as *D. carota*, *H. annuus*, and *T. caeruleum* as well as *S. oleracea* from the more distantly related Caryophyllales (Fig. 4). Based on the

Fig. 3 Phylogenetic analysis of the 54 genes of 31 complete plastid genome sequences including *V. macrocarpon*. Maximum likelihood bootstrap values were 100 for all nodes
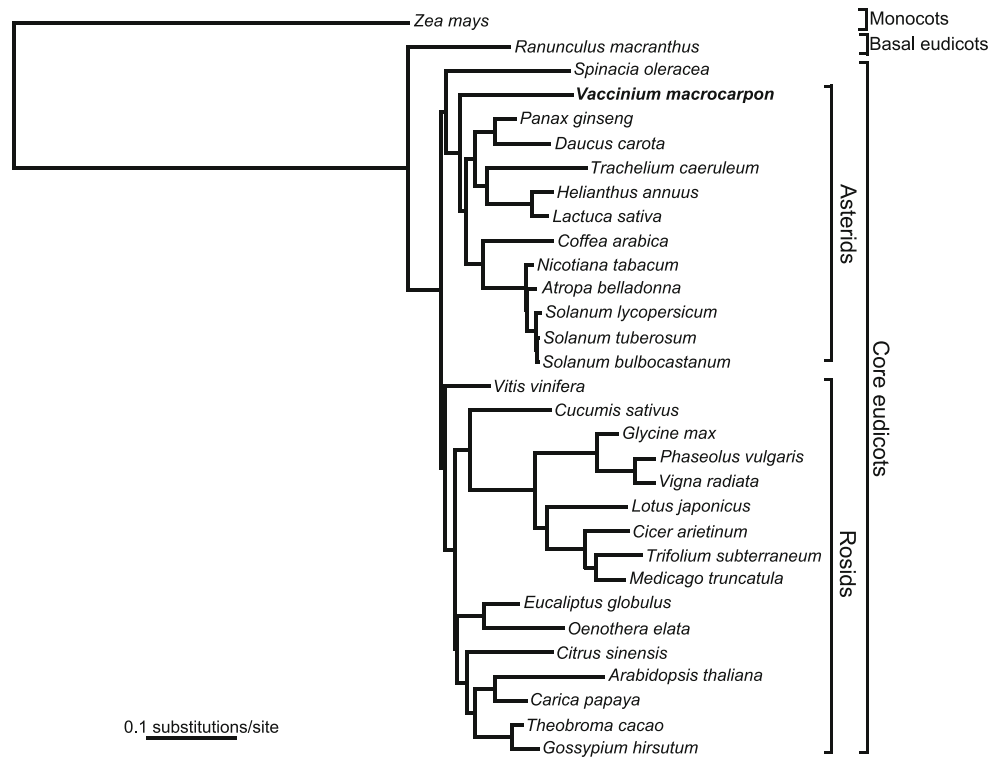
Fig. 4 Complete plastid genome comparisons of *V. macrocarpon* (GenBank accession JQ_757046) against **a** *S. oleracea* (NC_002202), **b** *D. carota* (NC_008325), **c** *H. annuus* (NC_007977), and **d** *T. caeruleum* (NC_010442). The *blue lines* represent BLAST hits between the two genomes, and the *red lines* represent BLAST hits within the same genome. The cranberry genome is drawn in the *right semicircle of each diagram*. Genes located in the plus strand are *green-colored blocks*, while genes in the minus strand are *red-colored blocks*

Species	D. carota	S. oleracea	H. annuus	T. caeruleum
Daucus carota	-			
Spinacia oleracea	0	-		
Helianthus annuus	3	3	-	
Trachelium caeruleum	5	5	8	-
Vaccinium macrocarpon	9	9	9	13

Fig. 5 Alignment of complete plastid genomes including *V. macrocarpon* (GenBank accession JQ_757046), *D. carota* (GenBank accession NC_008325), *S. oleracea* (NC_002202), *H. annuus* (NC_007977), and *T. caeruleum* (NC_010442). The *red arrows* indicate an example of a site with rearrangements and inversions. The table *below* shows the minimum number of rearrangements between the two plastid genomes

locally collinear blocks (LCB) detected by Mauve software, cranberry has a large number of rearrangements and inversions in the LSC region compared to the other species (Fig. 5). Despite plastid genomes possessing a conserved architecture, plastids with LSC rearrangements and conserved IRs have been found in different plant species, for example, *T. caeruleum* and *H. annuus* (Wu et al. 2011). For example, the relative position of a LSC gene block (flanked by psaA and psbD; Figs. 4 and 5) is similar in both sunflower (Asterales) and carrot (Apiales) and to some extent in *T. caeruleum* (although transcribed in the opposite direction), but this region is part of apparent inversions in cranberry with adjacent coding regions that are significantly different from other asterid species (Fig. 5). The relative position of the same region in the more distantly related spinach (Caryophyllales, non-Asterales) is more similar to that of sunflower and carrot than to that of cranberry, further indicating the unique gene order of the cranberry plastid genome. Furthermore, the size of the cranberry plastid genome is larger than the average in the *Asteridae* clade, likely due to repetitive sequences (Fig. 5). However, there was no obvious correlation between the location of such repeated sequences and the LCB determined by Mauve that could explain some of the rearrangements (Supplemental Fig. 3).

The American cranberry plastid possessed notable differences in genome order and direction of transcription when compared to other asterids. Cranberry is a member of the Ericales which comprises 25 families, 346 genera, and approximately 11,515 species. The current study has generated the first completed plastid genome for the genus *Vaccinium* (approximately 50 genera and 1,580 species), but more importantly, for the entire family Ericaceae and order Ericales. Approximately one third of these species is classified in the Ericaceae and represents approximately 6 % of the eudicots. The lineage is estimated to be originated from 124 to 90 mya, depending upon the analysis, with all families present by approximately 50 mya (Schönenberger et al. 2010). The Ericales lineage is sister to the core asterids (Stevens 2001), a phylogenetic position important for understanding the evolution of the asterids themselves. Furthermore, relatively few published plastid genomes occur within the *Asteridae sensu lato* compared to the rosid lineage. The completion of the first Ericales plastid genome will provide not only a reference genome for future work in the economically and horticulturally important members of the order but also a reference for researchers interested in broad evolutionary plastome questions within the Ericales and phylogenetically nearby lineages; many of whom also lack a reference genome to date.

As a starting point to study evolutionary history and determine when the cranberry rearrangements reported here occurred, we are currently sequencing close relatives of American cranberry, such as *Vaccinium oxycoccos* and *Vaccinium vitis-idea* as well as other members of the *Vaccinium* genus and the order Ericales. We are also sequencing representatives of *Vaccinium darrowii*. This species of blueberry is thought to be the progenitor of all the other *Vaccinium* spp. within the *Cyanococcus* section, including species with and without persistent leaves (Vander Kloet 1988; Bruederle and Vorsa 1994). Finally, in order to study more recent evolution and breeding history of the American cranberry, we are also sequencing other cranberry accessions from diverse regions and parentage. Assuming maternal inheritance of the plastid, the genome presented herein should originate from "McFarlin," a Massachusetts native selection ["HyRed" is derived from a "Stevens" × "Ben Lear" cross, and the "Stevens" pedigree is "McFarlin" × "Potter's favorite" (Eck 1990; McCown and Zeldin 2003)].

This genome will facilitate research in multiple arenas. This is the first published complete plastid genome sequence from the economically and ecologically important genus *Vaccinium*. It will be useful as a reference for further plastid genome sequencing efforts in *Vaccinium* and related species, for studies of plastid gene function, and for more intensive evolutionary and phylogenetic studies.

Data archiving statement The accession number JQ_757046 was assigned to the complete cranberry plastid genome sequence after being submitted to NCBI.

References

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. J Mol Biol 215:403–410

Anderberg AA, Rydin C, Källerjö M (2002) Phylogenetic relationships in the order Ericales s.l.: analyses of molecular data from five genes from the plastid and mitochondrial genomes. American J Bot 89:677–687

Bruederle LP, Vorsa N (1994) Genetic differentiation of diploid blueberry, *Vaccinium* sect. *Cyanococcus* (Ericaceae). Syst Botany 19:337–349

Clarke JL, Daniell H (2011) Plastid biotechnology for crop production: present status and future perspectives. Plant Mol Biol 76:211–220

Darling AE, Mau B, Perna NT (2010) progressiveMauve: multiple genome alignment with gene gain, loss, and rearrangement. PLoS One 5(6):e11147

Eck P (1990) The American cranberry. Rutgers University Press, New Brunswick

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32:1792–1797

Fazekas AJ, Burgess KS, Kesanakurti PR, Graham SW, Newmaster SG, Husband BC, Percy DM, Hajibabaei M, Barrett SCH (2008) Multiple multilocus DNA barcodes from the plastid genome discriminate plant species equally well. PLoS One 3:e2802

Galmes J, Flexas J, Keys AJ, Cifre J, Mitchell RAC, Madgwick PJ, Haslam RP, Medrano H, Parry MAJ (2005) Rubisco specificity factor tends to be larger in plant species from drier habitats and in species with persistent leaves. Plant Cell Environ 28:571–579

Hikosaka K, Shigeno A (2009) The role of Rubisco and cell walls for the interspecific variation in photosynthetic capacity. Oecologia 160:443–451

Iorizzo M, Senalik D, Szklarczyk M, Grzebelus D, Spooner S, Simon P (2012) De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. BMC Plant Biol 12:61

Jansen RK, Saski C, Lee S, Hansen AK, Daniell H (2011) Complete plastid genome sequences of three rosids (*Castanea*, *Prunus*, *Theobroma*): evidence for at least two independent transfers of rpl22 to the nucleus. Mol Biol Evol 28:835–847

Kalt W (2002) Health functional phytochemicals of fruits. Hortic Rev 27:269–315

Kron KA, Chase MW (1993) Systematics of the Ericaceae, Empetraceae, Epacridaceae and related taxa based upon rbcL sequence data. Ann Mo Bot Gard 80:735–741

Kron KA, Judd WS (1997) Systematics of the Lyonia group (Andromedeae, Ericaceae) and the use of species as terminals in higher-level cladistic analyses. Syst Bot 22:479–492

Kron KA, Judd WS, Stevens PF, Crayn DM, Anderberg AA, Gadek PA, Quinn CJ, Luteyn JL (2002) Phylogenic classification of Ericaceae: molecular and morphological evidence. Bot Rev 68:335–423

Kumudini S (2004) Effects of radiation and temperature on cranberry photosynthesis and characterization of diurnal variation in photosynthesis. J Amer Soc Hort Sci 129:106–111

Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL (2004) Versatile and open software for comparing large genomes. Genome Biol 5:R12

Maliga P, Svab Z (2011) Engineering the plastid genome of *Nicotiana sylvestris*, a diploid model species for plastid genetics. Plant Chromosom Eng: Met Protoc 701:37–50

Martins WS, Lucas DCS, Neves KFS, Bertioli DJ (2009) WebSat—a web software for microsatellite marker development. Bioinformation 3:282–283

McCown BH, Zeldin EL (2003) 'HyRed' and early, high fruit color cranberry hybrid. Hortscience 38:304–305

Nock CJ, Waters DL, Edwards MA, Bowen SG, Rice N, Cordeiro GM, Henry RJ (2011) Chloroplast genome sequences from total DNA for plant identification. Plant Biotechnol J 9:328–333

Posada D, Crandall KA (1998) Modeltest: testing the model of DNA substitution. Bioinformatics 14:817–818

Rivas JDL (2002) Comparative analysis of chloroplast genomes: functional annotation, genome-based phylogeny, and deduced evolutionary patterns. Genome Res 12:567–583

Schönenberger J, von Balthazar M, Sytsma KJ (2010) Diversity and evolution of floral structure among early diverging lineages in the Ericales. Phil Trans R Soc B 365:437–448

Stamatakis A, Hoover P, Rougemont J (2008) A rapid bootstrap algorithm for the RAxML web-servers. Syst Biol 57:758–771

Stevens PF (2001) Angiosperm phylogeny website. Version 9 June 2008. http://www.mobot.org/MOBOT/research/APweb/

Sugiura M (1992) The chloroplast genome. Plant Mol Biol 19:149–168

Sugiura M (1995) The chloroplast genome. Essays Biochem 30:49–57

Tangphatsornruang S, Sangsrakru D, Chanprasert J, Uthaipaisanwong P, Yoocha T, Jomchai N, Tragoonrung S (2010) The chloroplast genome sequence of mungbean (*Vigna radiata*) determined by high-throughput pyrosequencing: structural organization and phylogenetic relationships. DNA Res 17:11–22

Tesler G (2002) GRIMM: genome rearrangements web server. Bioinformatics 18:492–493

USDA-NASS (2011) Crop values 2011 summary. http://www.nass. usda.gov USDA-NASS, Washington, DC

Vander Kloet SP (1988) The genus *Vaccinium* in North America. Agr Canada Publ 1828:201

Wang W, Messing J (2011) High-throughput sequencing of three Lemnoideae (duckweeds) chloroplast genomes from total DNA. PLoS One 6:e24670

Wang SY, Stretch AW (2001) Antioxidant capacity in cranberry is influenced by cultivar and storage temperature. J Agr Food Chem 49:969–974

Wu CS, Lin CP, Hsu CY, Wang RJ, Chaw SM (2011) Comparative chloroplast genomes of Pinaceae: insights into the mechanism of diversified genomic organizations. Genome Biol Evol 3:309–319

Wyman SK, Jansen RK, Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinformatics 20:3252–3255

Zalapa JE, Cuevas H, Steffan S, Zhu H, Senalik D, Zeldin E, McCown B, Harbut R, Simon P (2012) Using next generation sequencing approaches for the isolation of simple sequence repeats (SSR) in the plant sciences. Am J Bot 99:193–208

Zhu H, Senalik D, McCown BH, Zeldin EL, Speers J, Hyman J, Bassil N, Hummer K, Simon PW, Zalapa JE (2012) Mining and validation of pyrosequenced simple sequence repeats (SSRs) from American cranberry (*Vaccinium macrocarpon* Ait.). Theor Appl Genet 124:87–96