



Article

Biodiversity of *rolB/C*-like Natural Transgene in the Genus *Vaccinium* L. and Its Application for Phylogenetic Studies

Roman Zhidkin ¹, Peter Zhurbenko ^{1,2}, Olesya Bogomaz ³, Elizaveta Gorodilova ⁴, Ivan Katsapov ¹, Dmitry Antropov ¹ and Tatiana Matveeva ^{1,*}

¹ Department of Genetic and Breeding, St. Petersburg State University, Saint Petersburg 199034, Russia; st085586@student.spbu.ru (R.Z.)

² Komarov Botanical Institute of the Russian Academy of Sciences, Saint Petersburg 197022, Russia

³ Faculty of Bioengineering and Bioinformatics, Moscow State University, Moscow 119991, Russia

⁴ Research Park, St. Petersburg State University, Saint Petersburg 199034, Russia

* Correspondence: t.v.matveeva@spbu.ru; Tel.: +7-904-515-4005

Abstract: A variety of plant species found in nature contain agrobacterial T-DNAs in their genomes which they transmit in a series of sexual generations. Such T-DNAs are called cellular T-DNAs (cT-DNAs). cT-DNAs have been discovered in dozens of plant genera, and are suggested to be used in phylogenetic studies, since they are well-defined and unrelated to other plant sequences. Their integration into a particular chromosomal site indicates a founder event and a clear start of a new clade. cT-DNA inserts do not disseminate in the genome after insertion. They can be large and old enough to generate a range of variants, thereby allowing the construction of detailed trees. Unusual cT-DNAs (containing the *rolB/C*-like gene) were found in our previous study in the genome data of two *Vaccinium* L. species. Here, we present a deeper study of these sequences in *Vaccinium* L. Molecular-genetic and bioinformatics methods were applied for sequencing, assembly, and analysis of the *rolB/C*-like gene. The *rolB/C*-like gene was discovered in 26 new *Vaccinium* species and *Agapetes serpens* (Wight) Sleumer. Most samples were found to contain full-size genes. It allowed us to develop approaches for the phasing of cT-DNA alleles and reconstruct a *Vaccinium* phylogenetic relationship. Intra- and interspecific polymorphism found in cT-DNA makes it possible to use it for phylogenetic and phylogeographic studies of the *Vaccinium* genus.

Keywords: *Vaccinium* L.; cellular T-DNA; allele phasing; phylogenetics



Citation: Zhidkin, R.; Zhurbenko, P.; Bogomaz, O.; Gorodilova, E.; Katsapov, I.; Antropov, D.; Matveeva, T. Biodiversity of *rolB/C*-like Natural Transgene in the Genus *Vaccinium* L. and Its Application for Phylogenetic Studies. *Int. J. Mol. Sci.* **2023**, *24*, 6932. <https://doi.org/10.3390/ijms24086932>

Academic Editors: Wenpan Dong and Shiliang Zhou

Received: 25 March 2023

Revised: 6 April 2023

Accepted: 6 April 2023

Published: 8 April 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

“*Agrobacterium*” is a soil bacterium with a unique mechanism to transfer well-defined DNA fragments (T-DNAs) into the chromosomes of a large variety of dicotyledonous plant species [1–6]. Normally, the transfers result in crown gall disease (in the case of *Agrobacterium tumefaciens* Conn or *A. vitis* Ophel and Kerr) or hairy roots formation (in the case of *A. rhizogenes*, also called *Rhizobium rhizogenes* Young). It has been experimentally demonstrated that hairy roots can regenerate into fertile plants [7–9]. Moreover, a variety of plant species, found in nature, contain T-DNA in their genomes, which they transmit in a series of sexual generations. Such T-DNA is called cellular (cT-DNA), and the plants containing it are called naturally transgenic or genetically-modified organisms (nGMO). Initially, such sequences were discovered in *Nicotiana glauca* Graham and other *Nicotiana* L. species [10–12], and further nGMOs were found inside genera *Linaria* Mill. [13], *Ipomoea* L. [14], *Cuscuta* L. [15], and dozens of other species [6,16–18]. These discoveries enabled making a number of generalizations regarding the distribution, structures, and possible functions of cT-DNA, summarized in our previous paper [19]. Sequences of cT-DNAs have been used in phylogenetic studies of *Linaria* [20], *Ipomoea* [21], *Nicotiana* [22], and *Camellia* [23]. They present several advantages for exploring the origin and evolution of nGMO species. To begin with, cT-DNAs are well-defined, highly specific, and well

recognizable DNA fragments, unrelated to plant sequences. They do not occur in non-transformed ancestors, and their integration at a particular chromosomal site constitutes a founder event which creates a clear start for a new clade. Moreover, unlike transposable elements, cT-DNA inserts do not amplify and disseminate in the genome after insertion. This prevents the generation of multiple paralogs, which is a significant advantage over classical nuclear markers. Additionally, cT-DNAs can be large and old enough to generate a range of variants, thereby allowing the construction of detailed trees [23]. Phasing of T-DNA alleles enables the reconstruction of more precise phylogenetic relationships and identification of hybridization events in the evolution of plant species [24]. These approaches have been successfully applied in the study of the genus *Camellia* [23].

We also found unusual sequences of the *rolB/C*-like gene in the genome data of *Vaccinium corymbosum* L. and *Vaccinium macrocarpon* Ait. [6,16]. This gene belongs to the *plast*-gene family. It is more similar to *plast*-genes present in fungus *Laccaria bicolor* (Maire) P.D.Orton (and absent in most basidiomycetes), plant *Nyssa sinensis* Oliv., and bacteria *Ensifer* sp. Casida, than to ones from known strains of *Agrobacterium* (*Rhizobium*) *rhizogenes*, which indicates the transfer of this sequence from an unknown «*Agrobacterium*» strain [16].

Based on the results of previous phylogenetic studies, we can try to reconstruct phylogenetic relationships between species of *Vaccinium* genus, using the *rolB/C*-like gene sequence. This task is quite relevant since the phylogeny of *Vaccinium* genus is highly controversial.

The *Vaccinium* genus was first described by Carl Linnaeus in 1753 [25]. After that, there were no large-scale phylogenetic studies of this genus until the beginning of the 20th century, when active breeding of cranberries and blueberries started in the United States [26,27]. Therefore, the first studies of *Vaccinium* phylogeny and taxonomy mainly involved species distributed in North America. The main difficulties in such studies were caused by the absence of fertility barriers in morphologically different organisms, which leads to the formation of a large number of hybrids, along with polyploidy distribution throughout the genus [27].

Camp was one of the first to introduce the system of *Vaccinium* genus in 1945 [28]. He conducted his classification according to morphological features. As a result, the genus was divided into several sections. The North American blueberry section *Cyanococcus* included 9 diploid, 12 tetraploid, and 3 hexaploid species [29]. Later, some of the species were merged into *V. corymbosum* with three levels of ploidy; therefore, in the new Kloet's classification, the *Cyanococcus* section contained 6 diploid, 5 tetraploid, and 1 hexaploid blueberry species [29], while dividing into sections did not change itself [30].

Thereby, *Vaccinium* species differentiation is complicated by polyploidy, similar morphology, and introgressions during hybridization [29]. Additionally, the application of morphological traits for phylogenetic studies of this genus does not always enable an ambiguous assessment of the evolutionary relationships between the studied species. This problem can be partially solved by the application of modern methods of phylogeny.

Methods such as DNA barcoding, fingerprinting, phylogenomics, and chemosystematics have been used recently in the study of *Vaccinium* genus. They allowed us to reconsider the taxonomy proposed by K. Kloet. An example of such study is the application of *matK* and *ITS* markers to determine phylogenetic relationships of various representatives of the entire tribe *Vaccinieae* [31]. The resulting dendrograms did not confirm traditional generic boundaries, but several well-supported clades were found on the tree: Andean; Mesoamerican/Caribbean; East Malaysian; *Agapetes* (consisting of some Asiatic *Vaccinium* and *Agapetes*); *Bracteata-Oarianthe* (represented by the respective sections); *Orthaea/Notopora* (containing the genera *Orthaea* Klotzsch and *Notopora* Hook); *Myrtillus*; and *Vaccinium* clades. Moreover, most of the discovered clades combined representatives of various genera, while the clades *Vaccinium* and *Myrtillus* contained species from different *Vaccinium* sections according to the previous classifications. Based on the obtained results, the authors concluded that it is necessary to reassess the taxonomy of the genus *Vaccinium*, since it is not monophyletic. Although work on this reassessment began in 2003 [32], a number of researchers pointed at too radical differences from the accepted structure of the genus.

Inconsistencies can be explained by the difficult interpretation of the results of phylogenetic analysis based on ITS and *matK*, in taxon, where hybridization and polyploidization played a significant evolutionary role [33].

Since the use of DNA barcoding does not enable unambiguous reconstruction of the phylogeny of *Vaccinium* genus, more time-consuming and expensive methods of molecular phylogenetics were used in genetic studies of economically important species. Recent assemblies of the genomes of *V. macrocarpon*, *V. microcarpum* (Turcz. ex Rupr.) Schmalh., *V. oxycoccus* L., and *V. corymbosum* made it possible to conduct their comparative genomics [34,35]. Molecular dating has shown that *V. macrocarpon* diverged from *V. oxycoccus* 2 mya, and from *V. microcarpum* 4.5 mya. Divergence time estimated for *V. macrocarpon* and *V. corymbosum* differs from 5 to 10.4 mya [34,35].

An analysis of the intra- and interspecific variability of American blueberry species by double digest restriction-site-associated DNA sequencing [36], showed that the rabbit eye blueberry and northern highbush blueberry are relatively homogeneous, but the southern highbush blueberry contains a significantly more mixed genetic background. Considering the pedigree of blueberries, the most optimal solution was to separate the entire dataset into nine hypothetical genomes, which correspond to the number of species actively used in breeding, *V. darrowii* Camp, *V. elliotii* Chapm., *V. tenellum* Ait., *V. angustifolium* Ait., *V. corymbosum*, *V. constablaei* A. Gray, *V. virgatum* Ait., *V. myrtilloides* Michx., and *V. pallidum* Ait. The trends identified in this way are consistent with the history of blueberry breeding [36].

Thereby, attempts to reconstruct the phylogenetic relationships of species of the genus *Vaccinium* by various methods have led to contradictory results. The use of phylogenetic markers ITS and *matK* leads to the construction of a phylogenetic tree with significant differences from the classical system. The use of more expensive and time-consuming methods of genome studies gives a more plausible picture, that is more consistent with the classical system. However, expensive methods can currently be applied to a narrow list of economically important species. For more extensive studies of various representatives of the genus, easy-to-use and cheap markers are required. They can be developed based on genome regions newly acquired in the course of evolution with a known structure and localization. An example of such sequences would be cellular T-DNA. Therefore, our study aims to characterize cellular T-DNA polymorphism in the genomes of plants of *Vaccinium* genus, and to evaluate the possibility of its application for phylogenetic studies of the genus.

2. Results

2.1. *Vaccinium rolB/C-like Gene: General Description*

Until recently, the *rolB/C*-like gene has been described only in *V. corymbosum* and *V. macrocarpon*. In the framework of this study, using BLAST algorithm, this gene was described in the WGS data for *V. myrtillus* and *V. darrowii*. Their ORFs correspond to the amino acid sequences from 290 (in *V. myrtillus*) to 292 amino acids (in *V. macrocarpon*).

In blueberries, the gene is located on chromosome 3 (in the cranberry genome, this region corresponds to chromosome 4). Located on both sides of it, at a distance of 5 kb, are genes of plant origin with functions that have not been described yet. No other agrobacterial genes were found in the genome of plants of *Vaccinium* genus.

Our search for homologues of the *rolB/C*-like gene in SRA database revealed new species of naturally transgenic plants in the *Vaccinium* genus, summarized in Table 1.

Natural GMOs were found in the sections *Oxycoccooides*, *Oxycoccus*, *Cyanococcus*, *Hemimyrtillus*, *Myrtillus*, *Polycodium*, *Bracteata*, *Vaccinium*, and *Vitis-idaea*, indicating that their common ancestor was transformed before the sections diverged.

Table 1. Natural transgene of *Vaccinium* species, based on NGS data.

Section	Species	NGS Data	SRR Accession Number(s)
<i>Oxycoccoides</i> (Hooker f.) Sleumer	<i>V. japonicum</i> Miq.	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR13349629
<i>Oxycoccus</i> (Hill) Koch	<i>V. macrocarpon</i> Aiton	<i>rolB/C</i> -like gene described earlier, in new SRA data coverage is sufficient to assemble a full-length gene	SRR13376387 SRR13376383 SRR13376388
	<i>V. oxycoccus</i> L.	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR14876344
<i>Cyanococcus</i> A. Gray	<i>V. virgatum</i> Aiton/ <i>V. ashei</i> Reade	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR12686860 SRR12686864
	<i>V. darrowii</i> Camp	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR13865261 SRR15673719
	<i>V. corymbosum</i> L.	<i>rolB/C</i> -like gene was described earlier, in new SRA data coverage is sufficient to assemble a full-length gene	SRR8298002 SRR8298006
	<i>V. stenophyllum</i> Steud.	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19401505
	<i>V. tenellum</i> Aiton	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19908712
<i>Hemimyrtillus</i> Sleumer	<i>V. smallii</i> A. Gray	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19400793
<i>Myrtillus</i> Dumortier	<i>V. myrtillus</i> L.	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR7523927 SRR7523930 SRR7523926
	<i>V. scoparium</i> Leiberg ex Coville	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19405468
	<i>V. yatabe</i> Makino	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19400792
<i>Polycodium</i> (Rafinesque) Rehder	<i>V. stamineum</i> L.	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19401493
<i>Pseudocephalanthos</i> C.Y.Wu and R.C.Fang.	<i>V. dunalianum</i> Wight	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR7768528 SRR2057022
<i>Bracteata</i> J.J. Smith	<i>V. bracteatum</i> Thunb.	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR17459400 SRR17459403 SRR17459408
	<i>V. wrightii</i> A.Gray	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19401672
<i>Vaccinium</i> L.	<i>V. uliginosum</i> L.	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR7686610 SRR7686609
<i>Vitis-idaea</i> (Moench) Koch	<i>V. vitis-idaea</i> L.	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR5799277 SRR5799278 SRR5799279

Table 1. Cont.

Section	Species	NGS Data	SRR Accession Number(s)
Relation to any section is not defined	<i>V. talamancense</i> (Wilbur and Luteyn) Luteyn	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19401435
	<i>V. schoddei</i> Sleumer	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19401400
	<i>V. stanleyi</i> Schweinf	<i>rolB/C</i> -like gene is present in the genome, and coverage is sufficient to assemble a full-length gene	SRR19354672
	<i>V. wollastonii</i> Wernham	<i>rolB/C</i> -like gene is present in the genome, but coverage is not sufficient to assemble a full-length gene	SRR19401324

Genomic data for different species varies in coverage quality. In some cases, the presence of a transgene can be inferred from single reads. In other cases (in *V. selerianum*, *V. shastense*, *V. striicaule*, *V. symplocifolium*, *V. vidalii*, and *V. whitmeei*), the transgene was not found at all, but this does not mean its absence from the genome, it indicates, however, that additional studies are required.

Sequences with good coverage were used further to assemble full-length sequences. For a deeper understanding of the gene variability, field material was collected, which was included in further variability analysis of the *rolB/C*-like gene. In all studied species, except for *V. oxycoccus* (see below), only full-length sequences of the *rolB/C*-like gene were found.

When studying SNPs within full-length sequences, one can see that there are nucleotide differences characteristic of the species, distinguishing each of them from other representatives of the genus *Vaccinium*. They are represented by substitutions and indels. Moreover, indels are equal or multiples of three nucleotides. This is evidence in favor of the gene product functionality. Examples of such polymorphism can be SNPs in *V. uliginosum* and *V. vitis-idaea* shown in Figure S1, as well as indels in different *Vaccinium* species shown in Figure S2.

Modeling of protein structures based on the predicted amino acid sequences (Table S1) shows their similarity in representatives of the genus *Vaccinium*, as well as in *Laccaria*, *Nyssa*, and *Ensifer*. The closest of the described agrobacterial proteins is the RolB protein (Figure 1).

Collectively, these data allow us to outline the directions for future research in terms of these sequences' functionality.

2.2. Intraspecific Variability of *rolB/C*-like Gene

To assess intraspecific variability of the *rolB/C*-like gene, we used sequences of several independent samples presented in the SRA for *V. bracteatum*, *V. corymbosum*, *V. darrowii*, *V. dunalianum*, *V. macrocarpon*, *V. myrtillus*, *V. oxycoccus*, *V. uliginosum*, *V. virgatum/ashei*, and *V. vitis-idaea*. In addition, samples of *V. myrtillus*, *V. oxycoccus*, *V. uliginosum*, and *V. vitis-idaea* collected from different geographic locations were characterized for polymorphism of the studied gene. For each genotype, alleles were reconstructed, based on Sanger sequencing data, or SRA reads.

PCR with primers VaccR and VaccF, followed by sequencing of the products have shown that 8 out of 15 samples of *V. oxycoccus* were homozygous for a deletion spanning from nucleotides 198 to 800 of CDS (Figure 2). Three samples had deletions other than the most common. They spanned nucleotides 354–798, 390–548, and 159 to the end of the gene, respectively. Four samples contained full size genes. All samples of other species contained only full-size intact genes.

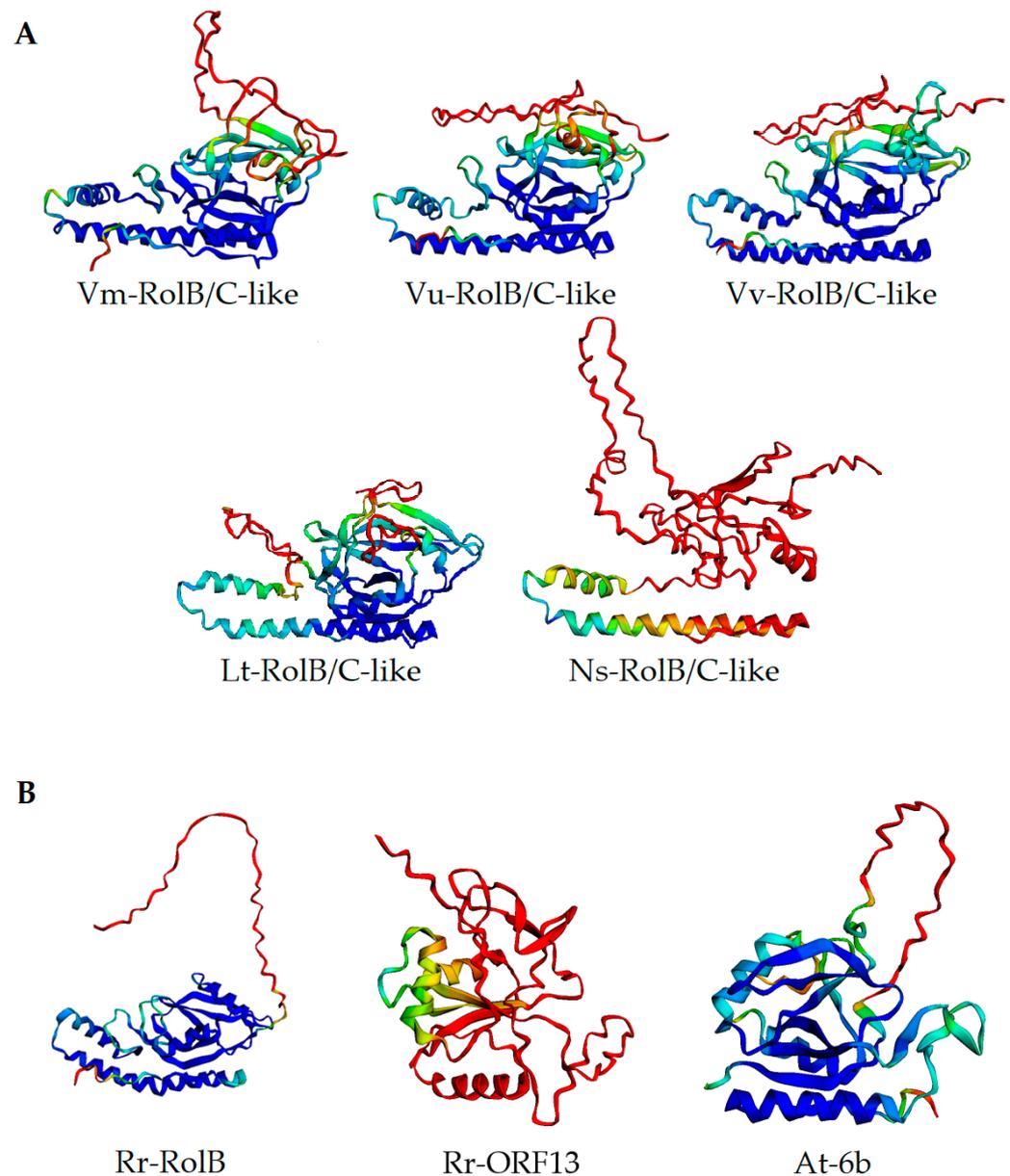


Figure 1. Predicted structures of various Plast-proteins. (A) RolB/C-protein from plants and fungus; (B) various plast-proteins from *Agrobacterium*. Abbreviations: Vm—*V. macrocarpon*, Vu—*V. uliginosum*, Vv—*V. vitis-idaea*, Lt—*L. trichodermophora*, Ns—*N. sinensis*, Rr—*R. rhizogenes*, and At—*A. tumefaciens*.

Within each species, there are SNPs represented by nucleotide substitutions that distinguish different alleles of a gene from each other. Based on these data, we can assess the diversity of alleles within a species, the frequency of occurrence of homo- and heterozygotes in populations, and the geographical distribution of alleles for some species. Data on the number of investigated alleles are presented in Table 2. Figure 3 shows allele frequency diagrams for *Vaccinium* species. This diagram does not claim to be an accurate estimate of allele frequencies, however, it stimulates the consideration that there are more common and rarer alleles. For example, A in *V. vitis-idaea* was found in most of the analyzed samples. The remaining alleles were found in a smaller number of samples and sometimes specific for certain locations. The proportion of homo- and heterozygotes and allelic diversity differ in the studied species, which can be explained by different reproduction biology. For example, blueberry species are more prone to cross-pollination, compared to cranberries.

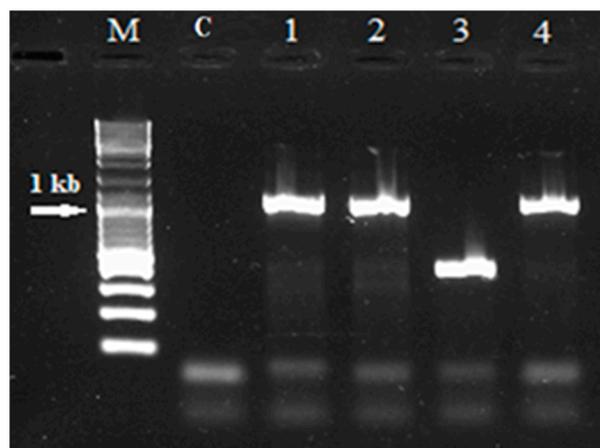


Figure 2. PCR products containing the *rolB/C*-like gene in cranberry samples. (M) Molecular Weight Marker GeneRuler™ 1 kb DNA Ladder (Thermo Fisher Scientific, USA); (C) negative control; (1,2,3)—*V. oxycoccus* (samples 1,2—full size, 3—the most common deletion); and (4) *V. macrocarpon*.

Table 2. Studied genotypes of *Vaccinium* and their allelic status.

Species	Number of Genotypes	Number of Unique Alleles	Number of Homozygotes	Number of Heterozygotes	Ploidy of Studied Species	Max. Alleles per Plant
<i>V. myrtillus</i>	13	10	10	3	Diploids	2
<i>V. corymbosum</i>	2	6	0	2	Tetraploids	4
<i>V. macrocarpon</i>	3	1	3	0	Diploids	1
<i>V. oxycoccus</i>	4	4	3	1	Diploids, tetraploids and hexaploids	2
<i>V. virgatum/ashei</i>	5	9	3	2	Hexaploids	3
<i>V. japonicum</i>	2	2	2	0	Diploids	1
<i>V. vitis-idaea</i>	18	9	5	13	Diploids	2
<i>V. uliginosum</i>	14	18	0	14	Diploids, tetraploids and hexaploids	3
<i>V. dunalianum</i>	2	1	2	0	Diploids	1
<i>V. bracteatum</i>	3	5	1	2	Diploids	2
<i>V. darrowii</i>	2	3	0	2	Diploids	2

Reconstruction of the phylogenetic relationships, based on molecular genetic studies of taxonomically significant regions of DNA, often relies on the evaluation of single specimens from the studied species. To avoid possible mistakes, especially when working with new markers, it is logical to test them out on a wider material. Since we used samples from collection points that are geographically distant from each other, it was decided to compare interspecific and intraspecific differences between them. Independent alleles were used for phylogenetic tree construction. Only full-length sequences were included in the analysis (Figure 4).

The resulting phylogenetic tree, in general, was consistent with our expectations. A total of 8 clades, that mostly correspond to analyzed species, can be distinguished on this tree. Interestingly, the clades represented by wild species (A, B, E, F) are homogeneous, despite the wide geography of sampling. At the same time, clades D and H are mosaic. Clade H is represented by alleles from *V. virgatum* and *V. ashei* Reade. Species name given by the authors of sequences was kept, but at the same time, we understand them as synonyms and refer them to the same species. Two *V. corymbosum* alleles C and E are noted in this clade. Both of them belong to the variety Draper, in the breeding of which *V. virgatum* and *V. darrowii* were involved [37]. Clade G consists of the related species *V. corymbosum* and *V. darrowii*, with an admixture of two alleles of *V. virgatum*. Their presence may be explained by hybridization during development of cultivars of North American Rabbit-eye blueberries.

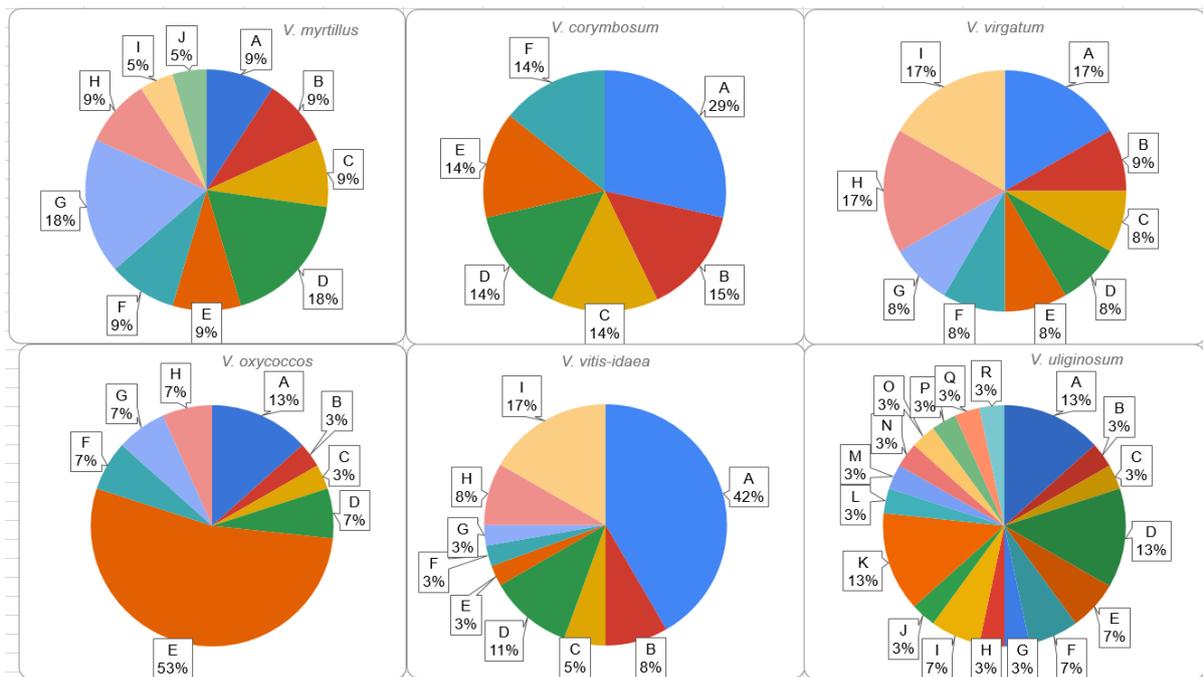


Figure 3. Frequencies of alleles of *rolB/C*-like gene in *Vaccinium* species.

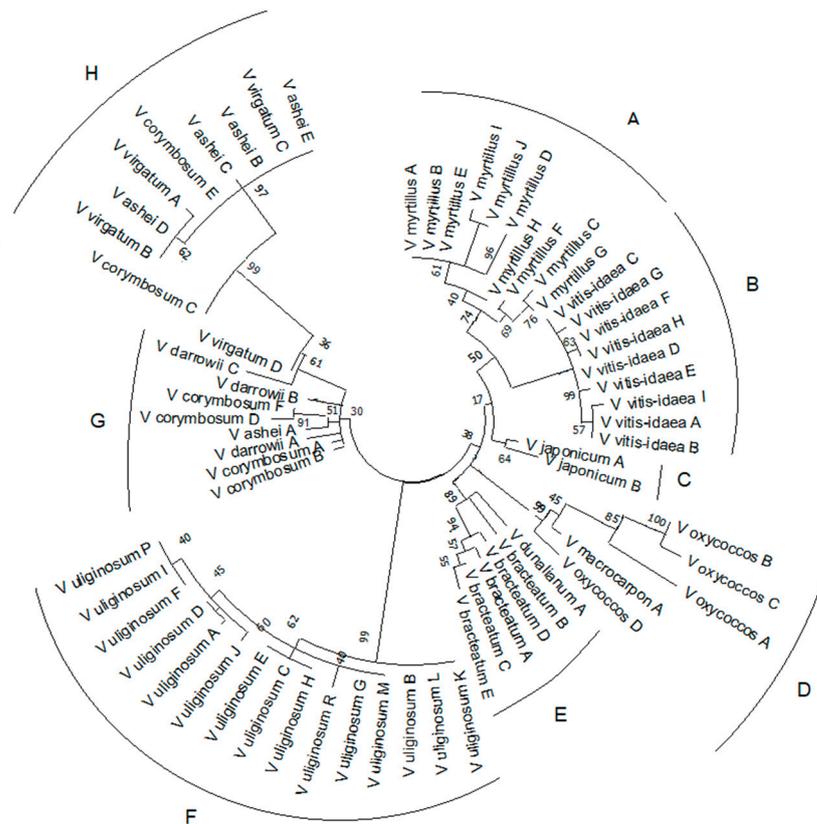


Figure 4. Phylogenetic relationships of alleles of *Vaccinium rolB/C*-like gene based on Maximum Likelihood method and General Time Reversible model (There were a total of 657 positions in the final dataset). (A) *V. myrtillus*, (B) *V. vitis-idaea*, (C) *V. japonicum*, (D) section *Oxycoccus*, (E) *V. bracteatum*, (F) *V. uliginosum*, (G,H) section *Cyanococcus*.

V. macrocarpon and *V. oxycoccus* are related species that form clade D. These species require more detailed study in the future, since the American cranberry is a cultivated species, and, therefore, hybridization was used in the early stages of its breeding. In turn, swamp cranberries are the only species known today where the transgene was extensively mutated in many of the studied samples.

In general, our studies have shown that, at least in the analyzed species, the intraspecific variability of transgene is lower than interspecific, and the mosaic distribution of alleles on the tree in American blueberries can be explained in terms of interspecific hybridization documented during the creation of studied varieties.

Therefore, we compiled consensus sequences, as a characteristic of each of the species analyzed at the previous stage, and supplemented this list with sequences obtained on the basis of gene sequences we obtained from single samples of the collection of the Komarov Botanical Institute.

2.3. Reconstruction of Phylogenetic Relationship of *Vaccinium* Species Based on *rolB/C*-like Sequence

In the final part of this study, several types of data were used. First of which, were the consensus sequences of the *rolB/C*-like gene of species studied in the previous stage. Alleles of presumably hybrid origin were excluded at the stage of constructing consensus sequences.

In addition, sequences of species sufficient for gene assembly but not sufficient for allele phasing were included in our analysis.

Finally, we obtained sequences of the *rolB/C*-like gene from additional species, represented by single samples in the collection of the Komarov Botanical Institute.

The phylogenetic tree presented in Figure 5 included, in addition to studied above, such species as *V. myrtilloides* (Sec. *Cyanococcus*), *V. ovalifolium* Sm. (Sec. *Myrtillus*), *V. praestans* Lamb. (Sec. *Praestantia*), *V. smallii* A. Gray (Sec. *Hemimyrtillus*), *V. vulcanorum* Kom. (Sec. *Vaccinium*), *V. conchophyllum* Rehder, *V. emarginatum* Hayata (Sec. *Conchophyllum*), and *Agapetes serpens* (Wight) Sleumer (based on previous studies, *Agapetes* can be considered a closely related genus in relation to *Vaccinium*, so its involvement in the work seems logical) [31].

As can be seen from Figure 5, species assigned to the same section in the classical system of *Vaccinium* genus are clustered together when constructing a tree based on the proposed DNA marker. Marked with ovals in Figure 5.

Another phylogenetic tree was constructed based on ITS sequences of samples studied in this work, combined with sequences from Kron et al. research [31]. This tree is presented in Figure 6. We can distinguish Andean, Meso-American/Caribbean, East Malaysian, *Agapetes*, *Bracteata-Oarianthe*, *Orthaea/Notopora*, *Myrtillus*, and *Vaccinium* clades, which were previously identified by Kron et al., composition of these clades differs significantly from the traditional taxonomy. Thereby, species *V. macrocarpon* and *V. oxycoccus* belong to the section *Oxycoccus*, but on ITS tree these species are separated by species *V. vitis-idaea*, which belongs to the section *Vitis-idaea*, Clade *Myrtillus* unites species of sections *Myrtillus* and *Macropelma*. Clade *Bracteata-Oarianthe* also includes species from sections *Hemimyrtillus*, *Oarianthe*, *Bracteata*, and *Baccula-Nigra*. Clade *Agapetes* includes representatives of genus *Agapetes* and some representatives of genus *Vaccinium* from sections *Rigiolepis*, *Conchophyllum*, and *Galeopetalum*, while some other species from these sections are spread over the tree. *Orthaea/Notopora* clade includes representatives of different genera, *Notopora*, *Orthaea*, and *Vaccinium*. The combination of *Vaccinium* representatives with other genera is also characteristic of Meso-American/Caribbean clades, which includes a representative of section *Oreades*.

There are much fewer of such contradictions in the phylogenetic tree based on the *rolB/C*-like gene.

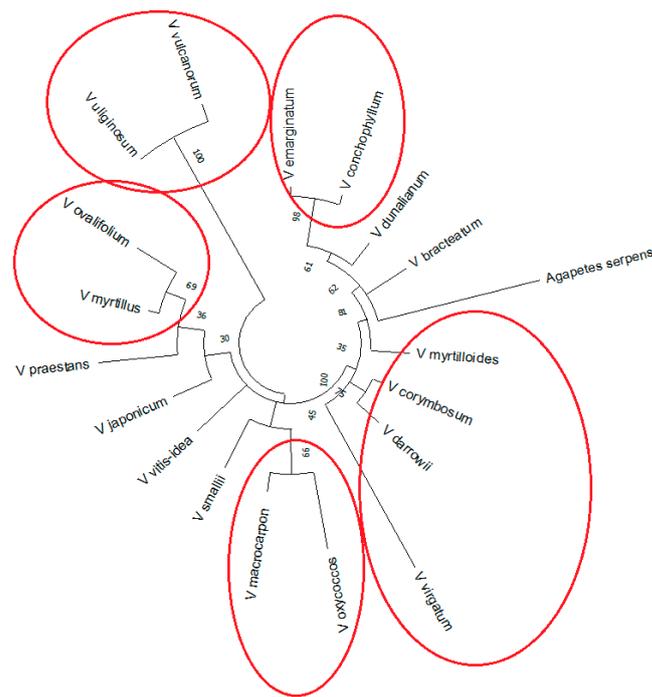


Figure 5. Molecular phylogenetic analysis of *Vaccinium* based on the *rolB/C*-like gene using Maximum Likelihood method and General Time Reversible model (There were a total of 667 positions in the final dataset). Representatives of the same section are circled in ovals.

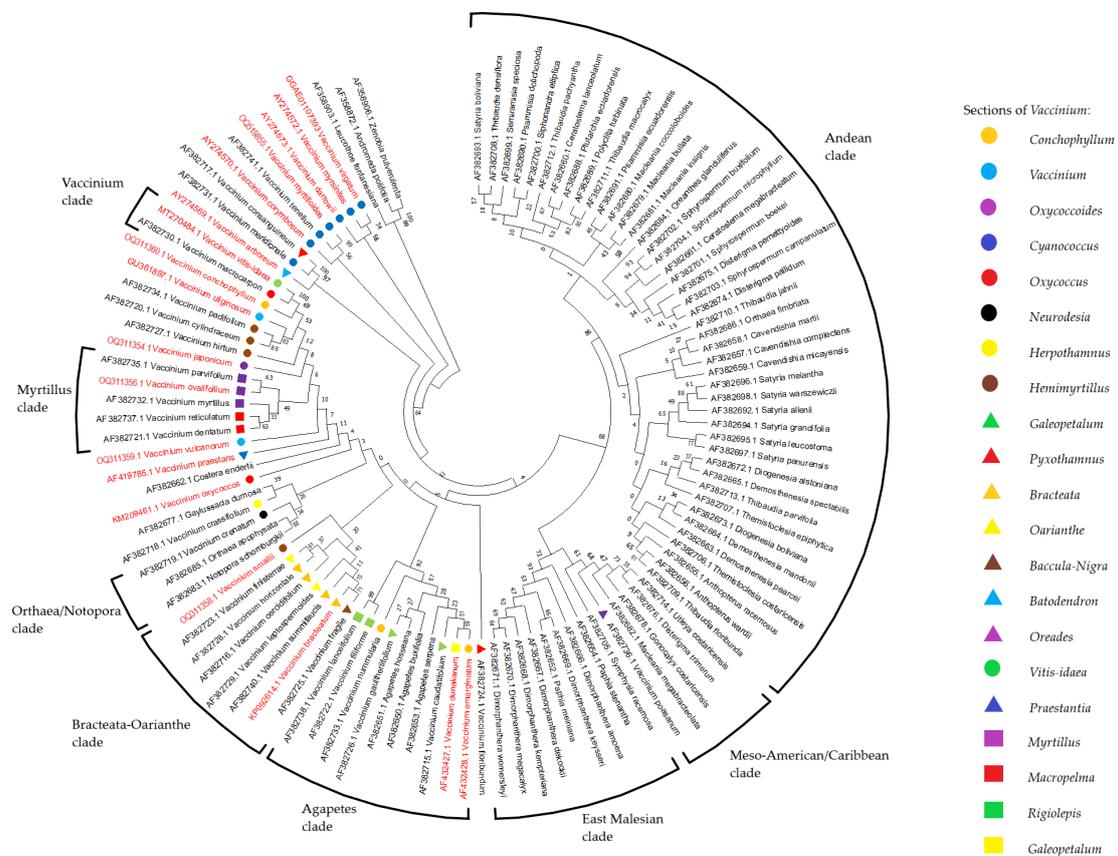


Figure 6. Molecular phylogenetic analysis of tribe Vaccinieae based on ITS sequences using Maximum Likelihood method and General Time Reversible model (There were a total of 651 positions in the final dataset). Species with sequences obtained by us are highlighted in red.

3. Discussion

Vaccinium is the fifth genus of natural GMOs after *Nicotiana*, *Linaria*, *Ipomoea*, and *Camellia*, where the cT-DNA structure was characterized in greater detail in many species. [11,13,14,23].

In addition to the unusual *plast*-gene found in plant genomes, a distinguishing feature of the *Vaccinium* genus is also the fact that the gene has been preserved intact in most of the studied species. The *plast*-genes family includes most of the oncogenes of *R. rhizogenes* (coding for RolB, RolC, Orf13, and Orf14 proteins), and some T-DNA genes of *A. tumefaciens* (coding for p5, p7, 6a, 6b) [38]. They have also been found in naturally transgenic plants, fungi, and some other bacteria [10,13,14,38]. *Plast* genes were described in detail in the review by Leon Otten [38], where he defined them on the basis of weak, but significant, protein similarity, and highlighted their capacity to change plant development, being introduced into plant genomes during the genetic transformation process. Levesque et al. [39] suggested that *plast* genes could have similar functions because of their common ancestry, and their diversification could be an adaptation to different plant species. Describing the biodiversity of *plast* genes can be very important in order to find their basic function and some common features of all these genes. In this regard, the discovery of *plast* genes, that are not quite similar to previously known ones (as in our case), is very valuable for future research.

In accordance with the analysis of previously assembled genomes and sequences, the representatives of two *Vaccinium* subgenera and *Agapetes* have a common localization site of the *rolB/C*-like sequence in the genome. This indicates a common origin of the sequence, as a result of a single transformation event of the ancestral form, and, accordingly, the monophyletic origin of all species containing transgene, included in our analysis. During evolution, the newly acquired sequence seems to have accumulated mutations. This variability can be assessed and used to estimate species divergence.

This is very important for the *Vaccinium* genus since there is still no consensus on the genus system.

Phylogenetic studies of the *Vaccinium* genus have been conducted since the middle of last century [26,27]. The traditional system of the genus involves its division into 2 subgenera and 33 sections. This division is based on morphological features [30]. The use of DNA barcoding methods showed the polyphyletic nature of the genus. At the same time, the construction of phylogenetic trees, based on ITS and *matK*, revealed contradictions with the traditional system. The trees constructed using ITS sequences by Kron et al. [31], and reconstructed in our study, based on an extended list of species, have shown a similar topology. They contained several well-maintained clades, such as Andean; Mesoamerican/Caribbean; East Malaysian; *Agapetes*, consisting of some Asiatic *Vaccinium* and *Agapetes*; *Bracteata-Orianthe*, uniting representatives of the respective sections; *Orthaea/Notopora*, which includes the genera *Orthaea* and *Notopora*; *Myrtillus*; and *Vaccinium*. These clades contained species of the genus *Vaccinium*, which, in previous classifications, belonged to different sections. Moreover, representatives of the *oxycoccus* subgenus were separated on the tree by *V. vitis-idea*. Representatives of the *Cyanococcus* section also fell into different clades, clustering with species of other sections.

This contradiction could be related to the manifestation peculiarities of used DNA markers in species of hybrid origin. The ITS of hybrid species retain the features of only one parent, and chloroplast markers are inherited mainly through the maternal line [40]. At the same time, the phylogeny based on genomic data is in better agreement with the traditional system [34,35,41,42]. Collectively, these data indicate that additional research is required, including the search for new cheap and reliable markers for the molecular phylogeny of *Vaccinium*.

In this article, we propose such a marker. An important feature of the work carried out by us was the use of a large number of forms of cranberries, lingonberries, and blueberries of different geographical origin, as well as the phasing of natural transgene alleles using SRA data in order to assess intraspecific variability. Such an approach was first used to study cellular T-DNA of representatives of the *Camellia* genus [23]. The analysis of individual alleles in phylogenetic studies makes it possible to identify acts of hybridization,

including interspecific hybridization. This is important to understand when assessing the degree of speciation. Alleles identification makes it possible to better describe intraspecific variability in order to assess the occurrence of alleles and their geography. In the future, this marker can be used to describe the allelic state of a gene in various populations in order to mark them and control unauthorized picking of wild berries. In addition to the general assessment of variability, the description of alleles makes it possible to identify the facts of hybridization in the pedigree of specific isolates, varieties, lines of cultivated *Vaccinium* plants, as well as to identify homo- and heterozygotes for this marker. This may be useful in interpreting controversial results of the origin of hybrid forms, as well as illustrating the incomplete isolation of species, as previously demonstrated in case of the genus *Camellia* [23].

Based on the phylogenetic analysis of allelic variants, obtained in our research, it can be concluded that intraspecific variability of marker in *Vaccinium* is less than the interspecific one. Therefore, it can be used to study the phylogenetic relationships of *Vaccinium* species.

A phylogenetic analysis of the genus *Vaccinium* was performed, using sequences of the *rolB/C*-like gene, as well as ITS sequences from previous studies [31], supplemented by sequences of our samples. The addition of new species to phylogenetic analysis, based on ITS, did not significantly change the topology of the tree.

The comparison of trees built based on ITS and *rolB/C*-like gene shows that they have similarities and differences. Species of *Myrtillus* section (*V. myrtillus* and *V. ovalifolium*) cluster together in both trees. Species of *Cyanococcus* section are clustered on both trees in a similar way. No other similarities in the topology of these trees were found. Species *V. conchophyllum* and *V. emarginatum* belonging to section *Conchophyllum* cluster together on the *rolB/C*-like tree, but not on ITS tree. Species *V. macrocarpon* and *V. oxycoccus* from section *Oxycoccus*, *V. vulcanorum* and *V. uliginosum* from section *Vaccinium* form separate clades on the *rolB/C*-like tree, but are scattered across different parts of the ITS tree.

Bracteata, *Hemimyrtillus*, *Vitis-idaea*, *Oxycoccoides*, and *Praestantia* sections are represented by single species on the *rolB/C*-tree. These sections require additional research involving new species. Reconstruction of phylogeny using the *rolB/C*-like gene is consistent with data obtained by other researchers using more expensive and time-consuming methods [34,35,42].

Therefore, a phylogeny, built on the basis of the *rolB/C*-like gene, has fewer contradictions with classical ideas, compared to a tree built based on ITS. The *rolB/C*-like gene, as a molecular marker, is cheap, easy to use, and is not time consuming. Given minimal amounts of biological material, new species can easily be included in further analyses.

4. Materials and Methods

4.1. Plant Material

Plant material was represented by samples of *Vaccinium myrtillus* L., *V. oxycoccus*, *V. vitis-idaea* L., and *V. uliginosum* L., collected from various habitats in Russia and Belgium. Samples of 9 species of the studied genus were taken from the collection of the Komarov Botanical Institute (St. Petersburg, Russia). A detailed description of the samples is presented in Table S2.

4.2. SRA Data

Sequences of *Vaccinium* species from SRA database, used in this study are presented in Table S3.

4.3. DNA Isolation

Leaves and stems of plants were grounded in liquid nitrogen. Plant DNA was isolated by a CTAB method described by Draper and co-authors [43].

4.4. PCR

Type I. Classic PCR was performed using the primers shown in Table 3, in the Tercyc gene cyclor (DNA Technology, Moscow, Russia). Primers were selected on the basis of cranberry genome fragment (JOTO01169953.1). The 40 μ L reaction mixture contained DreamTaq Green PCR Master Mix (2X) buffer (Thermo Fisher, Waltham, MA, USA), 100 ng DNA, and 10 μ M primers. The following program was used for PCR, 5 min at 93 $^{\circ}$ C; 40 cycles of 17 s at 93 $^{\circ}$ C, 30 s at X $^{\circ}$ C and 90 s at 72 $^{\circ}$ C; then 5 min at 72 $^{\circ}$ C, where X was the annealing temperature, depending on the primers used.

Table 3. Primers used in the work.

Primers		Tm	Amplicon Size	Purpose of Use
Name	Sequence			
VaccFvn	tctaacctaaccctgacc	50	1503	Amplification of the entire gene and flanking sequences
VaccRvn	aactcgtgattgtacctcgtt			
VaccR	ttactggccggctcctcatca			
VaccF	cacgtgtaaagcccgtgatgtt	55	1002	Amplification and sequencing of the entire gene
Vacc_ts_F	taaagcctgccactgcgatt	55	838	Efficient amplification of 838 out of 876 bp of coding sequence in any of studied <i>Vaccinium</i> genotype
Vacc_ts_pR	ccaatgcaccagtaactaac			
Vacc_reR	ccgaagtccgcgtcctg	n/a *	n/a	Primers for sequencing of gene fragments
Vacc_reF	tccaagccatcgacctactc			
pJET1.2 Forward Sequencing Primer	cgactcactatagggagagcggc	58	n/a	Real-time PCR from colonies and sequencing
pJET1.2 Reverse Sequencing Primer	aagaacatcgatttccatggcag			

* n/a—not applicable.

Type II. Real-time PCR from colonies was performed for additional control of early threshold values to search for clones containing the target insert. Real-time PCR was carried out in ANK-32 cyclor (Sintol, Moscow, Russia). Primers for plasmid pJET1.2 from the CloneJET PCR Cloning Kit (Thermo Fisher, Waltham, MA, USA) were used in the reaction. The 20 μ L of the reaction mixture contained SsoAdvanced Universal SYBR Green Supermix buffer (Bio-Rad, Hercules, CA, USA), 10 μ M primers, and a suspension of *E. coli* bacterial cells taken from a colony of transformants. The following program was used for real-time PCR: 60 s at 50 $^{\circ}$ C, 40 cycles of 5 min at 95 $^{\circ}$ C, 30 s at 58 $^{\circ}$ C, 60 s 72 $^{\circ}$ C, and 18 s at 95 $^{\circ}$ C.

PCR products were separated on agarose gel in 1x TBE buffer and visualized using GelDoc Go (BioRad, Hercules, CA, USA).

4.5. Molecular Cloning

PCR products were cloned into pJET1.2 vector, using CloneJET PCR Cloning Kit (Thermo Fisher, Waltham, MA, USA) according to the manufacturer instructions, and transformed into DH5 alpha chemically competent cells according to Inoue and co-authors [44].

4.6. DNA Sequencing

PCR fragments were sequenced using Sanger method, and BrilliantDyeTM Terminator (v3.1) Cycle Sequencing Kit (NimaGen, Nijmegen, The Netherlands). Then, sequencing mixtures were separated at the Resource Center of St. Petersburg University “Development of Molecular and Cellular Technologies” using an ABI Prism 3500 xl sequencer (Applied Biosystems, Waltham, MA, USA).

4.7. Allele Reconstruction from Sanger Sequencing Data

In order to determine the sequences of alleles, we represented in binary form each sample as a set of polymorphic positions, where “1” indicates the most frequent nucleotide in a given position in the species, and “0” the least frequent. Then, for each sample, all

possible combinations of values were written for polymorphic positions. Since homozygotes and samples with one SNP were found among the samples, we were able to obtain a primary pool of alleles, and these alleles were subsequently found in the remaining samples. Since each of these alleles, in a diploid, must match to an allele with alternative values in polymorphic positions, a homologous pair can be found for the primary allele [45].

In cases where the number of sequences was insufficient to resolve the set of sequences per allele, they were cloned into pJET 1.2 according to Clone Jet PCR Cloning Kit (Thermo Fisher Scientific, Waltham, MA, USA) manufacturer's instructions. Then, inserts from individual clones, that represent individual alleles, were sequenced.

4.8. Allele Phasing

The NCBI SRA database was used to search for T-DNA in the *Vaccinium* genus. The search was carried out using BLAST against the reference (JOTO01169953.1). Resulting reads were aligned to the reference using BWA 0.7.17 [46]. The processing of SAM files was carried out in SAMtools 1.7 [47]. The alignment visualization for ploidy estimation was performed in IGV 2.12.3 [48]. Allele phasing was performed using variant calling with GATK 4.2 [49] followed by WhatsHap 1.0 [50]. Additionally, the SAMtools phase was used for alternative analysis. In the case of small coverage, alleles were phased manually, with some only partially successful. Upon the assumption of SNP being artificial, its presence was verified in other samples.

4.9. Phylogenetic Analysis

Sequences were aligned using the MAFFT online service for multiple sequence alignment [51]. The evolutionary history was inferred by using the Maximum Likelihood method and General Time Reversible model [52]. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. Trees were drawn to scale, with branch lengths measured in the number of substitutions per site. All positions with less than 95% site coverage were eliminated, i.e., fewer than 5% alignment gaps, missing data, and ambiguous bases were allowed at any position (partial deletion option). Evolutionary analyses were conducted in MEGA11 [53].

4.10. Prediction of Protein's Structure

AlphaFold prediction [54] of the structures of various Plast-proteins was made using the Colabfold web resource at <https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb> (accessed on 18 April 2022), using default settings.

5. Conclusions

In this study, new species of natural GMOs were described within the genera *Vaccinium* and *Agapetes*. The natural transgene of these plants can be used for phylogenetic studies of the genus. This nuclear marker is cheap, easy to use, and is not time consuming. It can also be recommended for further research. The allele phasing approach makes it possible to track hybridization events in the evolution of studied plants. Data, regarding intraspecific variability of the proposed marker, can be used to mark populations of wild berries of the *Vaccinium* genus, in order to prevent their unauthorized harvesting.

Supplementary Materials: The following supporting information can be downloaded at: <https://www.mdpi.com/article/10.3390/ijms24086932/s1>.

Author Contributions: Investigation, Writing—original draft, R.Z.; Software, Writing—original draft, P.Z.; Resources, Software, O.B.; Investigation, E.G.; Resources, I.K. and D.A.; Conceptualization, Investigation, Methodology, Project administration, Resources, Writing—original draft, Writing—review and editing, T.M. All authors have read and agreed to the published version of the manuscript.

Funding: The work was performed using the equipment of the Resource Center of Saint Petersburg State University “Development of Molecular and Cellular Technologies” with the support of the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement No. 075-15-2022-322 dated 22 April 2022 on the provision of a grant in the form of a subsidy from the Federal budget of the Russian Federation. The grant was provided as part of the state support for the creation and development of the world-class Scientific Center “Agrotechnologies for the Future”.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The sequences obtained during the work were deposited into the National Center for Biotechnology Information database (<https://www.ncbi.nlm.nih.gov/>) (accessed on 26 February 2023) with accession numbers OQ095269-OQ095321, OQ129379-OQ129399, OQ311354, OQ311356, OQ311358-OQ311360, and OQ519855.

Acknowledgments: We are grateful to Shevchuk S. V. for the kind provision of the vaccinium material from the collection of the Russian Academy of Sciences Vladimir Komarov Botanical Institute’s Botanical Garden of Peter the Great, which he supervises.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Chilton, M.-D.; Drummond, M.H.; Merlo, D.J.; Sciaky, D.; Montoya, A.L.; Gordon, M.P.; Nester, E.W. Stable incorporation of plasmid DNA into higher plant cells: The molecular basis of crown gall tumorigenesis. *Cell* **1977**, *11*, 263–271. [[CrossRef](#)] [[PubMed](#)]
2. Zhu, J.; Oger, P.M.; Schrammeijer, B.; Hooykaas, P.J.J.; Farrand, S.K.; Winans, S.C. The bases of crown gall tumorigenesis. *J. Bacteriol.* **2000**, *182*, 3885–3895. [[CrossRef](#)] [[PubMed](#)]
3. Nester, E.W. *Agrobacterium*: Nature’s genetic engineer. *Front. Plant Sci.* **2015**, *5*, 730. [[CrossRef](#)] [[PubMed](#)]
4. Gelvin, S.B. Integration of *Agrobacterium* T-DNA into the plant genome. *Annu. Rev. Genet.* **2017**, *51*, 195–217. [[CrossRef](#)] [[PubMed](#)]
5. Lacroix, B.; Citovsky, V. Pathways of DNA transfer to plants from *Agrobacterium tumefaciens* and related bacterial species. *Ann. Rev. Phytopath.* **2019**, *57*, 231–251. [[CrossRef](#)]
6. Matveeva, T.V.; Otten, L. Widespread occurrence of natural transformation of plants by *Agrobacterium*. *Plant Mol. Biol.* **2019**, *101*, 415–437. [[CrossRef](#)]
7. Tepfer, D. Genetic transformation using *Agrobacterium rhizogenes*. *Physiol. Plant* **1990**, *79*, 140–146. [[CrossRef](#)]
8. Christey, M.C. Use of Ri-mediated transformation for production of transgenic plants. *In Vitro Cell. Dev. Biol.-Plant* **2001**, *37*, 687–700. [[CrossRef](#)]
9. Lütken, H.; Clarke, J.L.; Müller, R. Genetic engineering and sustainable production of ornamentals: Current status and future directions. *Plant Cell Rep.* **2012**, *31*, 1141–1157. [[CrossRef](#)]
10. White, F.F.; Garfinkel, D.J.; Huffman, G.A.; Gordon, M.P.; Nester, E.W. Sequence homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature* **1983**, *301*, 348–350. [[CrossRef](#)]
11. Furner, I.J.; Huffman, G.A.; Amasino, R.M.; Garfinkel, D.J.; Gordon, M.P.; Nester, E.W. An *Agrobacterium* transformation in the evolution of the genus *Nicotiana*. *Nat.* **1986**, *319*, 422–427. [[CrossRef](#)]
12. Suzuki, K.; Yamashita, I.; Tanaka, N. Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution. *Plant J.* **2002**, *32*, 775–787. [[CrossRef](#)]
13. Matveeva, T.V.; Bogomaz, D.I.; Pavlova, O.A.; Nester, E.W.; Lutova, L.A. Horizontal gene transfer from genus *Agrobacterium* to the plant *Linaria* in nature. *Mol. Plant Microbe Interact.* **2012**, *25*, 1542–1551. [[CrossRef](#)]
14. Kyndt, T.; Quispe, D.; Zhai, H.; Jarret, R.; Ghislain, M.; Liu, Q.; Gheysen, G.; Kreuzer, J.F. The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop. *Proc. Natl. Acad. Sci. USA* **2015**, *112*, 5844–5849. [[CrossRef](#)]
15. Zhang, Y.; Wang, D.; Wang, Y.; Dong, H.; Yuan, Y.; Yang, W.; Lai, D.; Zhang, M.; Jiang, L.; Li, Z. Parasitic plant dodder (*Cuscuta* spp.): A new natural *Agrobacterium*-to-plant horizontal gene transfer species. *Sci. China Life Sci.* **2020**, *63*, 312–316. [[CrossRef](#)]
16. Matveeva, T.V. New naturally transgenic plants: 2020 up-date. *Biol. Commun.* **2021**, *66*, 36–46. [[CrossRef](#)]
17. Bogomaz, F.D.; Matveeva, T.V. Expression sequences of opine synthase genes in natural GMOs based on analysis of their transcriptomes. *Plant Biotechnol. Breed.* **2022**, *5*, 15–24. [[CrossRef](#)]
18. Lipatov, P.Y.; Bogomaz, F.D.; Gosudarev, K.D.; Kondrashova, S.A.; Kuchevsky, M.V.; Malyuga, N.L.; Myagkiy, E.V.; Sergeenkova, M.V.; Tverdokhlebova, V.R.; Shtina, A.D.; et al. New cellular T-DNAs in naturally transgenic plants. *Ecol. Genet.* **2022**, *20*, 40–41. [[CrossRef](#)]
19. Matveeva, T.V. Why do plants need agrobacterial genes? *Ecol. Genet.* **2021**, *19*, 365–375. [[CrossRef](#)]
20. Matveeva, T.V. *Agrobacterium*-mediated transformation in the evolution of plants. *Curr. Top. Microbiol. Immun.* **2018**, *418*, 421–441. [[CrossRef](#)]

21. Quispe-Huamanquispe, D.G.; Gheysen, G.; Yang, J.; Jarret, R.; Rossel, G.; Kreuze, J.F. The horizontal gene transfer of *Agrobacterium* T-DNAs into the series batatas (Genus *Ipomoea*) genome is not confined to hexaploid sweetpotato. *Sci. Rep.* **2019**, *9*, 12584. [CrossRef]
22. Khafizova, G.V.; Matveeva, T.V. *Agrobacterium*-mediated transformation of *Nicotiana glauca* and *Nicotiana sylvestris*. *Vavilovskii Zhurnal Genetiki i Seleksii* **2022**, *26*, 697–703. [CrossRef]
23. Chen, K.; Zhurbenko, P.; Danilov, L.; Matveeva, T.; Otten, L. Conservation of an *Agrobacterium* cT-DNA insert in *Camellia* section *Thea* reveals the ancient origin of tea plants from a genetically modified ancestor. *Front. Plant Sci.* **2022**, *13*, 997762. [CrossRef]
24. Zhurbenko, P.M.; Klimenko, F.N. PhaseAll: A simple tool for read-based allele phasing. *Ecol. Genet.* **2022**, *20*, 32. [CrossRef]
25. Index Nominum Genericorum. Available online: <https://naturalhistory2.si.edu/botany/ing/> (accessed on 30 December 2022).
26. Vorsa, N.; Zalapa, J. Domestication, Genetics, and Genomics of the American Cranberry. *Plant Breed. Rev.* **2019**, *40*, 279–315. [CrossRef]
27. Camp, W.H. On the Structure of Populations in the Genus *Vaccinium*. *Brittonia* **1942**, *4*, 189–204. [CrossRef]
28. Camp, W. The North American blueberries with notes on other groups of Vacciniaceae. *Brittonia* **1945**, *5*, 203–275. [CrossRef]
29. Hancock, J.; Lyrene, P.; Finn, C.; Vorsa, N.; Lobos, G. Blueberries and Cranberries. *Temp. Fruit Crop Breed.* **2008**, 115–150. [CrossRef]
30. Vander Kloet, S.P. *The Genus Vaccinium in North America*; Agriculture Canada: Ottawa, ON, Canada, 1988.
31. Kron, K.; Powell, E.; Lutey, J. Phylogenetic relationships within the blueberry tribe (Vaccinieae, Ericaceae) based on sequence data from *matK* and nuclear ribosomal ITS regions, with comments on the placement of *Satyria*. *Am. J. Bot.* **2002**, *89*, 327–336. [CrossRef]
32. Powell, E.; Kron, K. Molecular Systematics of the Northern Andean Blueberries (Vaccinieae, Vaccinioideae, Ericaceae). *Int. J. Plant Sci.* **2003**, *164*, 987–995. [CrossRef]
33. Soltis, D.; Mavrodiev, E.; Doyle, J.; Rauscher, J.; Soltis, P. ITS and ETS Sequence Data and Phylogeny Reconstruction in Allopolyploids and Hybrids. *Syst. Bot.* **2008**, *33*, 7–20. [CrossRef]
34. Diaz-Garcia, L.; Garcia-Ortega, L.; González-Rodríguez, M.; Delaye, L.; Iorizzo, M.; Zalapa, J. Chromosome-Level Genome Assembly of the American Cranberry (*Vaccinium macrocarpon* Ait.) and Its Wild Relative *Vaccinium microcarpum*. *Front. Plant Sci.* **2021**, *12*, 137. [CrossRef] [PubMed]
35. Kawash, J.; Colt, K.; Hartwick, N.T.; Abramson, B.W.; Vorsa, N.; Polashock, J.J.; Michael, T.P. Contrasting a reference cranberry genome to a crop wild relative provides insights into adaptation, domestication, and breeding. *PLoS ONE* **2022**, *17*, e0264966. [CrossRef] [PubMed]
36. Nishiyama, S.; Fujikawa, M.; Yamane, H.; Shirasawa, K.; Babiker, E.; Tao, R. Genomic insight into the developmental history of southern highbush blueberry populations. *Heredity* **2020**, *126*, 194–205. [CrossRef] [PubMed]
37. Hancock, J. Blueberry Plant Denominated ‘Draper’. U.S. Patent No. USPP15103P3, 23 January 2003.
38. Otten, L. The *Agrobacterium* Phenotypic Plasticity (*Plast*) Genes. *Agrobacterium Biol.* **2018**, *418*, 375–419. [CrossRef]
39. Levesque, H.; Delepelaire, P.; Rouzé, P.; Slightom, J.; Tepfer, D. Common evolutionary origin of the central portions of the Ri TL-DNA of *Agrobacterium rhizogenes* and the Ti T-DNAs of *Agrobacterium tumefaciens*. *Plant Mol. Biol.* **1988**, *11*, 731–744. [CrossRef]
40. Rodionov, A.V.; Amosova, A.V.; Belyakov, E.A.; Zhurbenko, P.M.; Mikhailova, Y.V.; Punina, E.O.; Shneyer, V.S.; Loskutov, I.G. Genetic Consequences of Interspecific Hybridization, Its Role in Speciation and Phenotypic Diversity of Plants. *Russ. J. Genet.* **2019**, *55*, 278–294. [CrossRef]
41. Zhidkin, R.R.; Matveeva, T.V. Phylogeny problems of the genus *Vaccinium* L. and ways to solve them. *Ecol. Genet.* **2022**, *20*, 151–164. [CrossRef]
42. Schlautman, B.; Covarrubias-Pazarán, G.; Fajardo, D.; Steffan, S.; Zalapa, J. Discriminating power of microsatellites in cranberry organelles for taxonomic studies in *Vaccinium* and Ericaceae. *Genet. Resour. Crop Evol.* **2016**, *64*, 451–466. [CrossRef]
43. Draper, J.; Scott, R.; Armitage, P.; Walden, R. *Plant Genetic Transformation and Gene Expression—A Laboratory Manual*; Blackwell Science: Oxford, UK, 1988.
44. Inoue, H.; Nojima, H.; Okayama, H. High efficiency transformation of *Escherichia coli* with plasmids. *Gene* **1990**, *96*, 23–28. [CrossRef]
45. Zhidkin, R.R.; Zhurbenko, P.M.; Gorodilova, E.Y.; Matveeva, T.V. Molecular genetic and bioinformatic approaches for the allele reconstruction of the *rolB/C*-like gene in representatives of the genus *Vaccinium* L. *Ecol. Genet.* **2022**, *22*, 36–37. [CrossRef]
46. Li, H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* **2013**, arXiv:1303.3997. [CrossRef]
47. Danecek, P.; Bonfield, J.K.; Liddle, J.; Marshall, J.; Ohan, V.; Pollard, M.O.; Whitwham, A.; Keane, T.; McCarthy, S.A.; Davies, R.M.; et al. Twelve years of SAMtools and BCftools. *GigaScience* **2021**, *10*, 10–14. [CrossRef]
48. Robinson, J.T.; Thorvaldsdóttir, H.; Winckler, W.; Guttman, M.; Lander, E.S.; Getz, G.; Mesirov, J.P. Integrative genomics viewer. *Nat. Biotechnol.* **2011**, *29*, 24–26. [CrossRef]
49. Van der Auwera, G.A.; O’Connor, B.D. *Genomics in the Cloud: Using Docker, GATK, and WDL in Terra*, 1st ed.; O’Reilly Media: Sebastopol, CA, USA, 2020.
50. Martin, M.; Patterson, M.; Garg, S.; O’Fischer, S.; Pisanti, N.; Wklau, G.; Schöenhuth, A.; Marschall, T. WhatsHap: Fast and accurate read-based phasing. *bioRxiv* **2016**. [CrossRef]
51. Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*; Oxford University Press: New York, NY, USA, 2000.
52. Tamura, K.; Stecher, G.; Kumar, S. MEGA11: Molecular Evolutionary Genetics Analysis Version 11. *Mol. Biol. Evol.* **2021**, *38*, 3022–3027. [CrossRef]

53. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* **2019**, *20*, 1160–1166. [[CrossRef](#)]
54. Jumper, J.; Evans, R.; Pritzel, A.; Green, T.; Figurnov, M.; Ronneberger, O.; Tunyasuvunakool, K.; Bates, R.; Žídek, A.; Potapenko, A.; et al. Highly accurate protein structure prediction with AlphaFold. *Nature* **2021**, *596*, 583–589. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.