SYMBIOGENETICS

New naturally transgenic plants: 2020 update

Tatiana Matveeva

Department of Genetics and Biotechnology, Saint Petersburg State University, Universitetskaya nab., 7–9, Saint Petersburg, 199034, Russian Federation

Address correspondence and requests for materials to Tatiana Matveeva, t.v.matveeva@spbu.ru

Abstract

Agrobacterium-mediated gene transfer leads to crown gall or hairy roots disease, due to expression of transferred T-DNA genes. Spontaneous plant regeneration from the transformed tissues can produce natural transformants carrying cellular T-DNA (cT-DNA) sequences of agrobacterial origin. In 2019, based on genomic sequencing data, cT-DNA horizontally transferred from *Agrobacterium* were found in two dozen species of angiosperms. This made it possible to evaluate the spread of this phenomenon, as well as make some generalizations regarding the diversity of horizontally transferred genes. The presented research is a continuation of work in this field. It resulted in the description of new naturally occurring transgenic species *Aeschynomene evenia* C. Wright*, Eperua falcata* Aubl.*, Eucalyptus cloeziana* F. Muell.*, Boswellia sacra* Flueck.*, Kewa caespitosa* (Friedrich) Christenh.*, Pharnaceum exiguum* Adamson*, Silene noctiflora* L., *Nyssa sinensis* Oliv.*, Vaccinium corymbosum* L., *Populus alba* L. × *Populus glandulosa* Moench*.* The previously identified patterns regarding the frequency of the occurrence of natural transformants and the general properties of the cT-DNAs were confirmed in this study.

Keywords: cT-DNA, horizontal gene transfer, naturally-transgenic plants

Introduction

Agrobacterium-mediated transformation is the most common method for obtaining genetically modified plants. It is based on the ability of these soil bacteria to transfer a fragment of their plasmid (T-DNA, transferred DNA) and integrate it into the chromosome of the host plant. In nature, such a transfer leads to the development of two types of diseases: crown gall and hairy root diseases. These neoplasms are transgenic tissues on a non-transgenic plant. Scientists have managed to replace T-DNA genes with the sequences they need, transfer them using agrobacterial vectors into plant cells, and regenerate whole plants from such transgenic cells (Nester, 2014). It turned out that similar processes occur in nature, since plants were found to contain sequences homologous to the T-DNA of *Agrobacterium* in their genomes (Chen and Otten, 2017; Matveeva, 2018). This T-DNA was named cellular T-DNA (cT-DNA). The first such plants were found within the genus *Nicotiana* (White et al., 1983), and more than 20 years later in the genomes of *Linaria* and *Ipomoea* (Matveeva et al., 2012; Kyndt et al., 2015). Until 2019, the list of naturally transgenic plants was limited to these three genera. Digressing slightly from the main topic, we want to note that we are aware that the phylogeny of the genus *Agrobacterium* has been revised since the first discovery of T-DNA in wild plants (Young et al., 2001, 2003; Farrand et al., 2003); however, in the text of the manuscript we will use the collective term *Agrobacterium* as a tribute to tradition, and also because of the impossibility of accurately identifying the type of bacteria that participated in the transformation of the plant millions of years ago. The small fragments of T-DNA present in plant genomes are not sufficient for this. At the same time, further in the text of the manuscript, when indicating the closest of the modern strains, we will provide their modern name.

Citation: Matveeva, T. 2021. New naturally transgenic plants: 2020 update. Bio. Comm. 66(1): 36–46. https://doi.org/10.21638/spbu03.2021.105

Authors' information: Tatiana Matveeva, Dr. of Sci. in Biology, Professor, orcid.org/0000-0001-8569-6665

Manuscript Editor: Kirill Antonets, Department of Cytology and Histology, Faculty of Biology, Saint Petersburg State University, Saint Petersburg, Russia

Received: September 24, 2020;

Revised: November 8, 2020;

Accepted: November 25, 2020.

Funding: The article was made with support of the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement № 075-15-2020-922 dated 16.11.2020 on providing a grant in the form of subsidies from the federal budget of the Russian Federation. The grant was provided for state support for the creation and development of a world-class scientific center, Agrotechnologies for the Future.

Competing interests: The authors have declared that no competing interests exist.

The development of genomic sequencing and bioinformatics methods have opened up new opportunities for the search for new natural GMOs. Such a search was crowned with success in 2019 (Matveeva and Otten, 2019): another two dozen species, the ancestors of which underwent *Agrobacterium*-mediated transformation during their evolution, were described within the genera *Eutrema, Arachis, Nissolia, Quillaja, Euphorbia, Parasponia, Trema, Humulus, Psidium, Eugenia, Juglans, Azadirachta, Silene, Dianthus, Vaccinium, Camellia* and *Cuscuta*. Analysis of transcriptome data revealed an additional list of natural transformants. However, the degree of confidence in natural transgenicity based on transcriptomic data is lower than that based on results of genome sequencing and assembly. This is due to the lack of information about the localization site of the sequences, which leads to the possibility that the sequences result from *Agrobacterium* DNA contamination. The most interesting results of transcriptome assembly were several T-DNA-like sequences of the representatives of the genus *Diospyros*, containing a combination of opine and *plast*-genes. Matveeva and Otten's (2019) study was done exclusively using bioinformatic analysis of published sequences of plant genomes. A few months later, an article was published in which molecular methods confirmed the presence of T-DNA in plants of the genus *Cuscuta*, previously identified by bioinformatics means (Zhang et al., 2020). Numerous new examples of natural transformants show that at least 7 % of the dicotyledonous species are naturally transformed plants, and provide valuable material for studying the role of horizontal gene transfer in plant evolution (Matveeva and Otten, 2019). These results also serve as an important argument in support of GMOs.

A year has passed since the publication of Matveeva and Otten (2019). During this time, new plant genomes were sequenced and deposited in the NCBI database (O'Leary et al. 2016). The aim of this work was to update the list of naturally transgenic plants taking into account new NGS data, and generalize all the results obtained.

Material and methods

The search for T-DNA-like sequences was done based on National Center for Biotechnology Information (NCBI) Whole-Genome Shotgun (WGS) contigs of all plant genomes sequenced since April 2019 to date, using the TBLASTN algorithm with default settings. In the second step, Vir protein sequences were used to search for possible *Agrobacterium* contaminations in those genomes. In the third step, contigs that potentially encoded T-DNA-like protein sequences with identity levels 30 % or higher were analyzed further. They were used as queries in BLASTX with default settings to detect the closest protein homologs and to identify proteins encoded by plant genes surrounding the cT-DNA. All query sequences are detailed in our previous paper (Matveeva and Otten, 2019). The Vector NTI AdvanceTM software was used to build the cT-DNA maps.

Phylogenetic analysis of *rolB/C* homologs was done in MEGA 7.0 (Kumar et al., 2016) by using the Maximum Likelihood method based on the JTT matrix-based model (Jones et al., 1992) (In addition, the Dayhoff matrix based model (Schwarz and Dayhoff, 1979), Poisson correction model (Zuckerkandl and Pauling, 1965) and Equal Input model (Tajima and Nei, 1984) were used for more reliable conclusions). The bootstrap consensus tree inferred from 500 replicates was taken to represent the evolutionary history of the taxa analyzed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates were collapsed. 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 a JTT model, and then selecting the topology with superior log likelihood value. The analysis involved 19 amino acid sequences. All positions with less than 95 % site coverage were eliminated. That is, fewer than 5 % alignment gaps, missing data, and ambiguous bases were allowed at any position. There were a total of 140 positions in the final dataset.

The supplementary materials present a similar analysis performed by UPGMA method (Sneath and Sokal, 1973) and neighbor-joining method (Saitou Nei, 1987).

Results and discussion

Since April 2019 (Matveeva and Otten, 2019), the genomes of another 206 angiosperm species have been sequenced. New examples of natural GMOs were identified in 10 species (about 5 %) from 10 genera, 9 families and 7 orders, according to the previously described methodology (Matveeva and Otten, 2019). They are listed in Table 1. Schemes of extended cT-DNAs are shown in Figure 1.

For representatives of two genera, the cT-DNA structure was specified. At the same time, their transgenic nature was described earlier.

Until recently, two variants of cT-DNA have been characterized in plants of the genus *Ipomoea* (Kyndt et al., 2015; Quispe-Huamanquispe et al., 2019). In our study, based on the genome sequences of *I. trifida* (Kunth) G. Don and *I. batatas* (L.) Lam., a new cT-DNA variant was discovered. It contains *mas2'*-like and *mas1'*-like sequences. The fragment that we found in *I. trifida* was named *It*-TDNA3. A similar (86 %) fragment was also found in *I. batatas*. At the same time, the boundary sequences of plant origin are 97 % similar, showing that they result from the same transformation event. The database also contains short contigs containing *mas2'* homologues. However, it is not possible to attribute them to any extended sequence. Further research is required

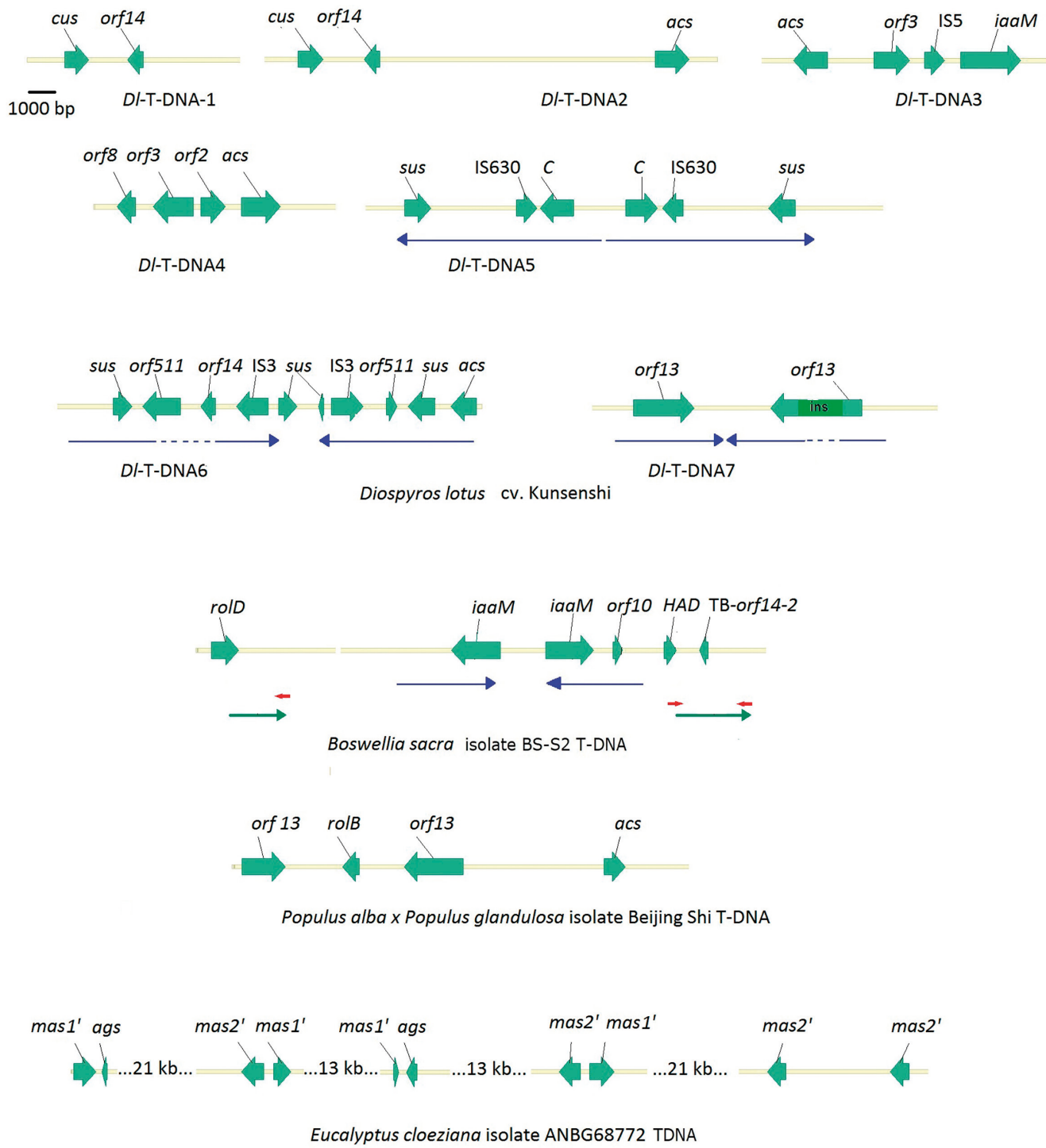


Fig. 1. Structure of cT-DNA plant species. (Wide green arrows show sequences similar to *Agrobacterium* T-DNA genes, blue arrows show inverted repeats, green thin arrows show direct repeats. Red arrows show short repeating sequences).

to clarify the nature of these sequences. Therefore, they are not currently listed in the results table.

We predicted a cT-DNA in *Diospyros lotus* L. (dateplum) based on the analysis of the TSA database (Matveeva and Otten, 2019). Analysis of the results of genome assembly made it possible to describe seven variants of cT-DNA in this species, representing footprints of several independent transformation events in the evolution of this species (Fig. 1). *Dl*-TDNA1 and 2 are located close to the boundaries of the assembled sequences. They share 99 % similarity and may be part of the same cT-DNA. If so, then this is the youngest cT-DNA in the genome of this species, which can be dated by the repeat structure. It is followed by *Dl*-TDNA5, 7 and 6. *Dl*-TDNA6 is the oldest one. Other traces of multiple acts of agrobacterial transformation in the evolution of ancestral forms of

Table 1. New cT-DNAs detected by analysis of WGS database

order	family	Species, cultivar, line, isolate	Accession #	Gene homolog	In-tact*	position	Identity level to proteins from NCBI		Similarity level between 2 arms of the cT-DNA
							% of identity	Organism and protein ID	
Malpighiales	Salicaceae	*Populus alba* × *Populus glandulosa* isolate Beijing Shi	SMNX01000141.1	*orf13*-like	-	98346 -99995	49	BAB16132.1 *Rhizobium rhizogenes* (Riker et al. 1930) Young et al. 2001	92
				rolB -like	-	102808 -102180**	60	CAA82552.1 *R. rhizogenes*	
				orf13-like	-	106748 -104516	49	BAB16132.1 *R. rhizogenes*	
				acs-like	-	112142 -112940	86	WP_174054263.1 *R. rhizogenes*	
Fabales	Fabaceae	*Aeschynomene evenia* isolate CIAT22838	RYYW01000009.1	*mis*-like	+	25927086 — 25926118	62	WP_174054202.1 *R. rhizogenes*	n/a
		Eperua falcata	CWNJ01257379.1	*nos*-like	+	1339-149	54	WP_167693616.1 *Sinorhizobium meliloti* (Dangeard 1926) De Lajudie et al. 1994	n/a
Myrtales	Myrtaceae	*Eucalyptus cloeziana* isolate ANBG68772	JABKBO010000005.1	*mas1'*-like	-	19003305 — 19004414	70	P27874.2 *R. rhizogenes*	95
				ags-like	-	19005239 -19004732	69	WP_172690594.1 *Agrobacterium* sp.	
				mas2'-like	-	19029872 -19028766	61	AIM40180.1 *R. rhizogenes*	
				mas1'-like	-	19030390 -19031235 ..... 19046200 -19046439	71...47	P27874.2 *R. rhizogenes*	
				ags-like	-	19047360 -19046844	77	WP_032488587.1 *Agrobacterium* sp.	
				mas2'-like	-	19064603 — 19063528	61	AIM40180.1 *R. rhizogenes*	
				mas1'-like	-	19065068 — 19066309	66	WP_034520976.1 *Agrobacterium* sp.	
				mas2'-like	-	19091525 — 19090591	44	WP_172690593.1 *Agrobacterium* sp.	
				mas2'-like	-	19097721 -19096760	56	(WP_172690593.1 *Agrobacterium* sp.	

Continuation of the Table 1

order	family	Species, cultivar, line, isolate	Accession #	Gene homolog	In-tact*	position	Identity level to proteins from NCBI		Similarity level between 2 arms of the cT-DNA
							% of identity	Organism and protein ID	
Sapindales	Burseraceae	*Boswellia sacra* isolate BS-S2	SNVD01001790.1	*rolD*	-	4554 — 5602	69	WP_034521016.1 *Agrobacterium* sp.	n/a
				iaaM-like	-	15557 -13686	47	WP_174054196.1 *R. rhizogenes*	87
				iaaM-like	-	17274 — 19107	45	WP_174054196.1 *R. rhizogenes*	
				orf10-like	-	19840 -20193	40	AAA22094.1 *R. rhizogenes*	n/a
				HAD hydrolase family	?	21778 — 22257	40	WP_149743959.1 *Rhizobium* sp.	
				TB-orf14-2-like	-	23456 — 23151	31	AIM40179.1 *R. rhizogenes*	
				TB-orf14- 1-like	-	23561 -23151	29	AIM40178.1 *Nicotiana tomentosiformis*	
	Kewaceae	*Kewa caespitosa*	ONZA01007622.1	*mis*-like	+	509 — 1510	65	BAB85949.1 *Nicotiana glauca*	n/a
			ONZA01013126.1	*mis*-like	-	1503 — 2552	62	WP_176453671.1 *R. rhizogenes*	n/a
			ONZA01022047.1	*mis*-like	-	1924 — 2417	54	WP_176453671.1 *R. rhizogenes*	n/a
			ONZA01003306.1	*mis*-like	-	4964 — 5356	51	BAB85949.1 *Nicotiana glauca*	n/a
	Molluginaceae	*Pharnaceum exiguum*	ONZK01002102.1	*mis*-like	+	26894 — 27826	66	BAB85949.1 *Nicotiana glauca*	n/a
			ONZK01003836.1	*mis*-like	+	7318 — 8232	62	BAB85949.1 *Nicotiana glauca*	73
				mis-like	-	14903 — 15822	59	BAB85949.1 *Nicotiana glauca*	
	Caryophyllaceae	*Silene noctiflora* isolate OPL-1.1	VHZZ01000004.1	*cus*-like	-	1170610 — 1176104	60	WP_156551602.1 *Allorhizobium vitis* (Ophel and Kerr 1990)	n/a
				cus-like	+	1182125 — 1183057	64	WP_156551602.1 *A. vitis*	n/a
				cus-like	+	1210326 — 1209418	64	WP_156551602.1 *A. vitis*	n/a
			VHZZ01016788.1	*cus*-like	+	37722 — 36790	63	WP_174084799.1 *A. vitis*	n/a
			VHZZ01000003.1	*cus*-like	-	5037 — 5411	54	WP_174084799.1 *A. vitis*	n/a
				cus-like	-	25613 -24951	68	WP_174084799.1 *A. vitis*	n/a
				cus-like	+	28697 -27765	63	WP_174084799.1 *A. vitis*	n/a
				cus-like	+	35431 -34499	62	WP_174084799.1 *A. vitis*	n/a
			VHZZ01056725.1	*cus*-like	-	6031 — 5510	72	WP_071208191.1 *A. vitis*	n/a

order	family	Species, cultivar, line, isolate	Accession #	Gene homolog	In-tact*	position	Identity level to proteins from NCBI		Similarity level between 2 arms of the cT-DNA
							% of identity	Organism and protein ID	
Cornales	Nyssaceae	*Nyssa sinensis* isolate J267	VIRR01000271.1	*rolB/C*- like	+	5364628 -5365320	52	XP_001881215.1 *Laccaria bicolor*	n/a
Ericāles		*Diospyros lotus* cv. Kunsenshi	*Dl*-T-DNA1 BEWH01006414.1	*cus*-like	-	41404 — 42306	73	WP_156551602.1 *A. vitis*	n/a
				orf14-like	-	44357 -43779	56	WP_174054201.1 *R. rhizogenes*	
			Dl-T-DNA2 BEWH01000237.1	*cus*-like	+	334877 — 335779	73	WP_156551602.1 *A. vitis*	n/a
				orf14-like	-	337830 -337252	56	WP_174054201.1 *R. rhizogenes*	
				acs-like	+	347771 — 348979	81	GAJ95539.1 *R. rhizogenes*	
			Dl-T-DNA3 BEWH01006419.1	*acs*-like	+	8509 — 7220	80	GAJ95539.1 *R. rhizogenes*	n/a -
				orf3-like	-	10262 — 11610	70	KEA04445.1 *R. rhizogenes*	
				IS5 transposase	-	12185 -12934	83	WP_184141638.1 *Shinella fusca* Vaz-Moreira et al. 2010	
				iaaM-like	-	13558 -15830	45	WP_034521028.1 *Agrobacterium* sp.	
			Dl-T-DNA4 BEWH01000029.1	*orf8*-like	+	2523369 — 2522746	68	WP_116979321.1 *Agrobacterium salinitolerans* Yan et al. 2017	n/a
				orf3-like	-	2525305 -2523957	72	KEA04445.1 *R. rhizogenes*	
				orf2-like	-	2525577 -2526392		WP_174054193.1 *R. rhizogenes*	
				acs-like	-	2526938 — 2528244	73	KEA04447.1 *Agrobacterium* sp.	
			Dl-T-DNA5 BEWH01004217.1	*sus*-like	+	21622 — 22629	77	WP_174080856.1 *R. rhizogenes*	97
				IS630	-	25887 -26667	83	WP_165826447.1 *Rhizobium wuzhouense* Yuan et al. 2018	
				C -like	-	28060 -26786	76	WP_174054195.1 *R. rhizogenes*	
				C -like	-	30052 — 31252	76	WP_174054195.1 *R. rhizogenes*	
				IS630	-	32218 -31443	83	WP_165826447.1 *R. wuzhouense*	
				sus-like	-	36499 -35486	76	WP_174080856.1 *R. rhizogenes*	

End of the Table 1

order	family	Species, cultivar, line, isolate	Accession #	Gene homolog	Intact*	position	Identity level to proteins from NCBI		Similarity level between 2 arms of the cT-DNA
							% of identity	Organism and protein ID	
			DI-T-DNA5a BEWH01000041.1 (2128496 — 214 3594)	99 % identical to BEWH01004217.1					
			DI-T-DNA6 BEWH01000056.1	*sus*-like	-	1177021 -1177736	58	WP_034521016.1 *Agrobacterium* sp	91
				orf511-like	-	1179590 -1178145	60	AIM40183.1 *Nicotiana tomentosiformis*	
				orf14-like	-	1180901 — 1180354	68	AIM40184.1 *Nicotiana tomentosiformis*	
				IS3 family transposase	-	1182902 — 1181706	70	WP_143239454.1 *Agrobacterium rosae*	
				sus-like	-	1183310 — 1184002... 1185015 -1184824	47	WP_174080856.1 *R. rhizogenes*	
				IS3 family transposase	-	1185301 -1186508	73	WP_143239454.1 *Agrobacterium rosae*	
				orf511-like	-	1187393 -1187790	60	AIM40183.1 *Nicotiana tomentosiformis*	
				sus-like	-	1189225 -1188221	55	WP_174080856.1 *R. rhizogenes*	
				acs-like	-	1190796 — 1189840	38	KIJ92238.1 *Laccaria amethystina*	
	Ebenaceae		*DI*-T-DNA7 BEWH01000037.1	*orf13*-like	-	262687 -264989	50	WP_174075801.1 *R. rhizogenes*	92
				orf13-like	-	271357 -270518... 269030 — 267893	48	WP_174075801.1 *R. rhizogenes*	
	Vaccinioideae	*Vaccinium corymbosum* cultivar W8520	JACAOB010009726.1	*rolB/C*-like	+	1 — 729	41	XP_001884861.1 *Laccaria bicolor*	n/a

order	family	Species, cultivar, line, isolate	Accession #	Gene homolog	In-tact*	position	Identity level to proteins from NCBI		Similarity level between 2 arms of the cT-DNA
							% of identity	Organism and protein ID	
(is this similar to the earlier Silene T-DNA?)Solanales	Convolvuláceae	*Ipomoea trifida* cultivar Y22	SMMV01000602.1	Contains *Ib*-TDNA2					
			It-TDNA3 SMMV01000003.1	*mas2'*-like	-	13842355 -13841152	71	WP_032488585.1 *Agrobacterium* sp.	n/a
				mas2-like'	-	13842876 -13842715	72	AIM40180.1 *R. rhizogenes*	
				mas1'-like	-	13845955 -13846602	77	P27874.2 *R. rhizogenes*	
				mas1'-like	-	13847831 — 13848160	75	P27874.2 *R. rhizogenes*	
		Ipomoea batatas cultivar Taizhong6	NXFB01008336.1	Contains *Ib*-TDNA1, described by Kyndt et al. (2015)					
			FLTB01041015.1						
			NXFB01000007.1	Contains *Ib*-TDNA2, described by Kyndt et al. (2015)					
			Ib-TDNA3 NXFB01000244.1	*mas1'*-like	-	54319 -53674	77	WP_034520976.1 *Agrobacterium* sp.	n/a
				mas2'-like	-	56951 — 58017	68	WP_032488585.1 *Agrobacterium* sp.	
			NXFB01000002.1	*mas2'*-like	-	27628510 — 27627275	63	AIM40180.1 *R. rhizogenes*	n/a

* does not contain premature stop codons and / or frame shift

** gene location on the negative strand

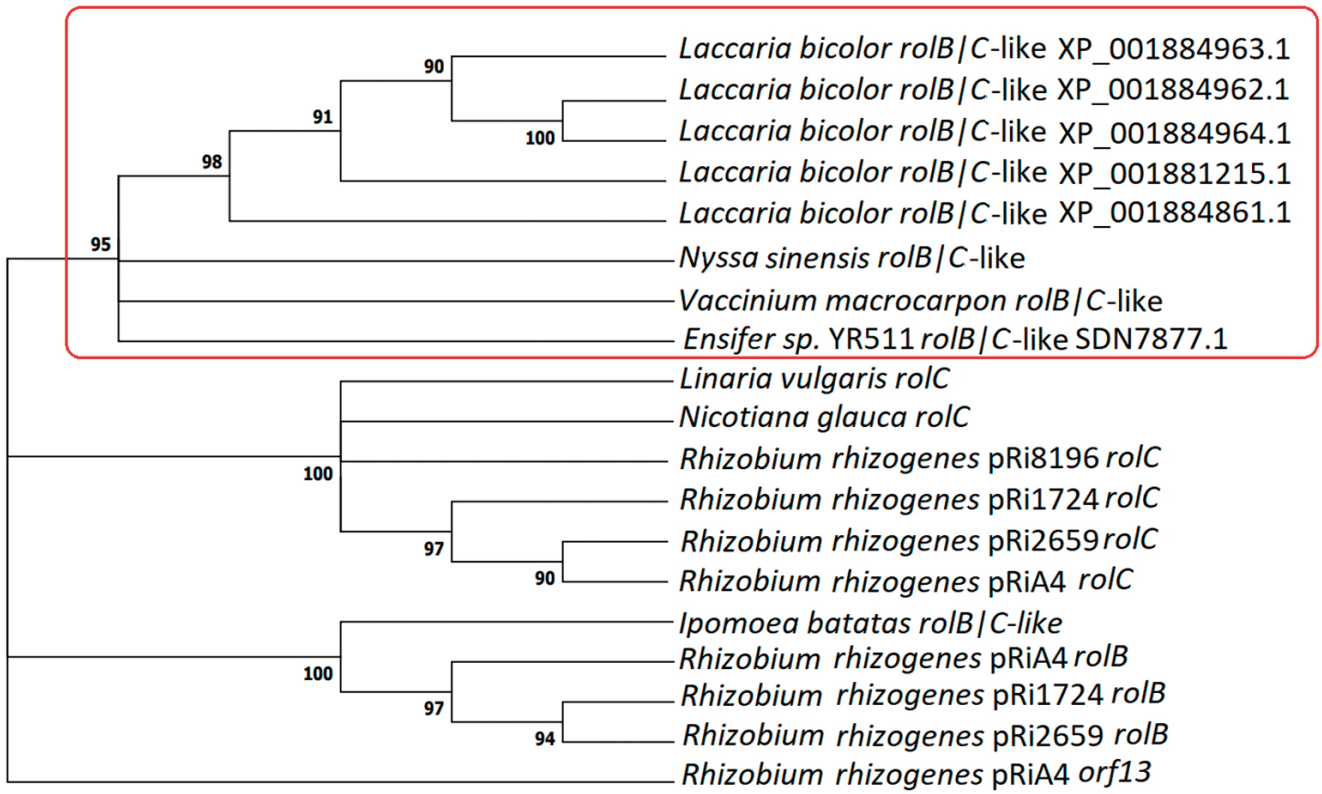


Fig. 2. Molecular phylogenetic analysis of *rolB/C* homologs from *Rhizobium, Ensifer*, *Laccaria*, *Ipomoea*, *Vaccinium* and *Nyssa* species by Maximum Likelihood method based on the JTT matrix-based model. (Dayhoff matrix based model, Poisson correction model and Equal Input model resulted to the same topology of the tree). The cluster containing new *rolB/C*-like gene is outlined in red.

modern species have been previously described within the genera *Nicotiana* and *Parasponia* (Chen et al., 2014; Matveeva and Otten, 2019)

All new species of naturally transgenic plants belong to the same orders where natural GMOs were previously described. *Vaccinium corymbosum* L. and *Silene noctiflora* L. belong to genera in which natural GMOs were previously found. They contain sequences similar to those described earlier, which can be further used for phylogenetic studies based on the T-DNA structure. Our study also confirms the prevalence of opine genes in natural transformants. As before, we observe extended cT-DNAs organized as repeats. Inverted repeats may be generated during the process of T-DNA transfer and integration into plant chromosomes. Direct repeats may possibly be explained by DNA rearrangements associated with transposons found around the repeated cT-DNA regions. An interesting feature of eucalyptus T-DNA is that relatively short fragments of agrobacterial origin with similar opine genes are interspersed with extended DNA fragments of plant origin. A large number of repeats of the same opine genes, that are found in *Silene* species, *Kewa caespitosa* (Friedrich) Christenh. and *Pharnaceum exiguum* Adamson is another feature that requires further study; it may result from the insertion of multiple copies during the initial transformation event, or from amplification of integrated copies at a later stage.

The data on the fine structure of cT-DNA in representatives of different taxa obtained earlier and in the present work can be further used to search for patterns of host specificity of modern agrobacterial strains. This issue can be investigated both from a phylogenetic and from an ecological point of view, since the idea of coevolution of symbionts is gaining in importance (Matveeva et al., 2018). We can already illustrate this thesis with the case of an unusual *plast* gene, which we described for the first time in the genomic sequence of *Vaccinium macrocarpon* Aiton. This fragment attracted our interest because it was closer to fungal *plast*-genes than agrobacterial ones. In the present work, a similar sequence was found in *Nyssa sinensis* Oliv. Figure 2 shows that *Nyssa, Vaccinium* and *Laccaria* sequences cluster together with *rolB/C*-like gene of *Ensifer* sp. from the Rhizobiaceae family. Phylogenetic trees constructed by other methods (Supp. Fig. 1) have a similar topology, which confirms the reliability of this cluster. The genera *Nyssa* and *Vaccinium* are not related, but these plants share similar habitats, characterized by increased moisture (https://www.hortweek.com; Song and Hancock, 2011). Perhaps the search for an *Agrobacterium* strain similar to those that transformed these species will lead to the discovery

of bacterial determinants that are important for the survival of such strains in wet habitats.

Conclusion

Thus, in this study, new natural GMOs were described in 10 species (*Aeschynomene evenia, Eperua falcate, Eucalyptus cloeziana, Boswellia sacra, Kewa caespitosa, Pharnaceum exiguum, Silene noctiflora, Nyssa sinensis, Vaccinium corymbosum, Populus alba* × *Populus glandulosa*) belonging to 10 genera, 9 families and 7 orders. The new type of cT-DNA was described in *Ipomoea trifida* and *Ipomoea batatas*, and the structure of cT-DNAs of *Diospyros lotus* cv. Kunsenshi was clarified. The previously identified patterns regarding the frequency of the occurrence of naturally transgenic plants and the general properties of the cT-DNAs were confirmed. The data obtained can be used further for genetic engineering, plant phylogeny and evolutionary research.

Acknowledgments

The author expresses her deep gratitude to Prof. L. Otten (IBMP, France) for critical reading of the manuscript, advice and comments.

References

Chen, K. and L. Otten. 2017. Natural *Agrobacterium* transformants: recent results and some theoretical considerations. *Frontiers in Plant Science* 8:1600. https://doi.org/10.3389/fpls.2017.01600

Elenevsky A. G. 2006. Botany. Systematics of higher, or terrestrial, plants: textbook. Moscow, Academy. (In Russian)

Farrand S. K., Van Berkum P. B., and Oger P. 2003. *Agrobacterium* is a definable genus of the family Rhizobiaceae. *International Journal of Systematic and Evolutionary Microbiology* 53(5):1681–1687. https://doi.org/10.1099/ijs.0.02445-0

Felsenstein, J. 1985. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution* 39:783–791. https://doi.org/10.1111/j.1558-5646.1985.tb00420.x

Jones, D. T., Taylor, W. R., and Thornton, J. M. 1992. The rapid generation of mutation data matrices from protein sequences. *Bioinformatics* 8(3):275–282. https://doi.org/10.1093/bioinformatics/8.3.275

Kumar, S., Stecher, G., and Tamura, K. 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular Biology and Evolution* 33(7):1870–1874. https://doi.org/10.1093/molbev/msw054

Kyndt, T., Quispe, D., Zhai, H., Jarret, R., Ghislain, M., Liu, Q., Gheysen, G., and Kreuze, J. F. 2015. The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop. *Proceedings of the National Academy of Sciences USA* 112(18):5844–5849. https://doi.org/10.1073/pnas.1419685112

Matveeva, T. V. and Otten, L. 2019. Widespread occurrence of natural genetic transformation of plants by *Agrobacterium*. *Plant Molecular Biology* 101(4–5):415–437. https://doi.org/10.1007/s11103-019-00913-y

Matveeva, T. V., Bogomaz, D. I., Pavlova, O. A., Nester, E. W., and Lutova, L. A. 2012. Horizontal gene transfer from genus *Agrobacterium* to the plant *Linaria* in nature. *Molecular Plant-Microbe Interactions* 25(12):1542–1551. https://doi.org/10.1094/MPMI-07-12-0169-R

Matveeva, T., Provorov, N., and Valkonen, J. P. T. 2018. Editorial: Cooperative adaptation and evolution in plant-microbe systems. *Frontiers in Plant Science* 9:1090. https://doi.org/10.3389/fpls.2018.01090

Matveeva, T. V. 2018. *Agrobacterium*-mediated transformation in the evolution of plants; pp. 421–441 in S. Gelvin (ed.), *Agrobacterium* biology. Current topics in microbiology and immunology. Springer, Cham. https://doi.org/10.1007/82_2018_80

Nester, E. W. 2014. *Agrobacterium*: nature's genetic engineer. *Frontiers in Plant Science* 5:730. https://doi.org/10.3389/fpls.2014.00730

O'Leary, N. A., Wright, M. W., Brister, J. R., Ciufo, S., Haddad, D., McVeigh, R., Rajput, B., Robbertse, B., Smith-White, B., Ako-Adjei, D., Astashyn, A., Badretdin, A., Bao, Y., Blinkova, O., Brover, V., Chetvernin, V., Choi, J., Cox, E., Ermolaeva, O., Farrell, C. M., Goldfarb, T., Gupta, T., Haft, D., Hatcher, E., Hlavina, W., Joardar, V. S., Kodali, V. K., Li, W., Maglott, D., Masterson, P., McGarvey, K. M., Murphy, M. R., O'Neill, K., Pujar, S., Rangwala, S. H., Rausch, D., Riddick, L. D., Schoch, C., Shkeda, A., Storz, S. S., Sun, H., Thibaud-Nissen, F., Tolstoy, I., Tully, R. E., Vatsan, A. R., Wallin, C., Webb, D., Wu, W., Landrum, M. J., Kimchi, A., Tatusova, T., DiCuccio, M., Kitts, P., Murphy, T. D., and Pruitt, K. D. 2016. Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation. *Nucleic Acids Research* 44(D1):D733–745. https://doi.org/10.1093/nar/gkv1189

Quispe-Huamanquispe, D. G., Gheysen, G., Yang, J., Jarret, R., Rossel, G., and Kreuze, J. F. 2019. The horizontal gene transfer of *Agrobacterium* T-DNAs into the series Batatas (Genus *Ipomoea*) genome is not confined to hexaploid sweetpotato. *Scientific Reports* 9(1):12584. https://doi.org/10.1038/s41598-019-48691-3

Saitou, N. and Nei, M. 1987. The neighbor-joining method: A new method for reconstructing phylogenetic trees. *Molecular Biology and Evolution* 4:406–425.

Schwarz, R. and Dayhoff, M. 1979. Matrices for detecting distant relationships; pp. 353–358 in M. Dayhoff (ed.), Atlas of protein sequences. National Biomedical Research Foundation.

Sneath, P. H. A. and Sokal, R. R. 1973. Numerical taxonomy. Freeman, San Francisco.

Song, G. Q. and Hancock, J. F. 2011. *Vaccinium*. In C. Kole (ed.), Wild crop relatives: Genomic and breeding resources. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-16057-8_10

Tajima, F. and Nei, M. 1984. Estimation of evolutionary distance between nucleotide sequences. *Molecular Biology and Evolution* 1:269–285.

White, F. F., Garfinkel, D. J., Huffman, G. A., Gordon, M. P., and Nester, E. W. 1983. Sequences homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature* 301(5898):348–350. https://doi.org/10.1038/301348a0

Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A., and Sawada, H. 2001. A revision of *Rhizobium* Frank 1889, with an emended description of the genus, and the inclusion of all species of *Agrobacterium* Conn 1942 and *Allorhizobium* undicola de Lajudie et al. 1998 as new combinations: *Rhizobium radiobacter*, *R. rhizogenes*, *R. rubi, R. undicola* and *R. vitis*. *International Journal of Systematic and Evolutionary Microbiology* 51:89–103. https://doi.org/10.1099/00207713-51-1-89

Young, J. M., Kuykendall, L. D., Martínez-Romero, E., Kerr, A., and Sawada, H. 2003. Classification and nomencla-

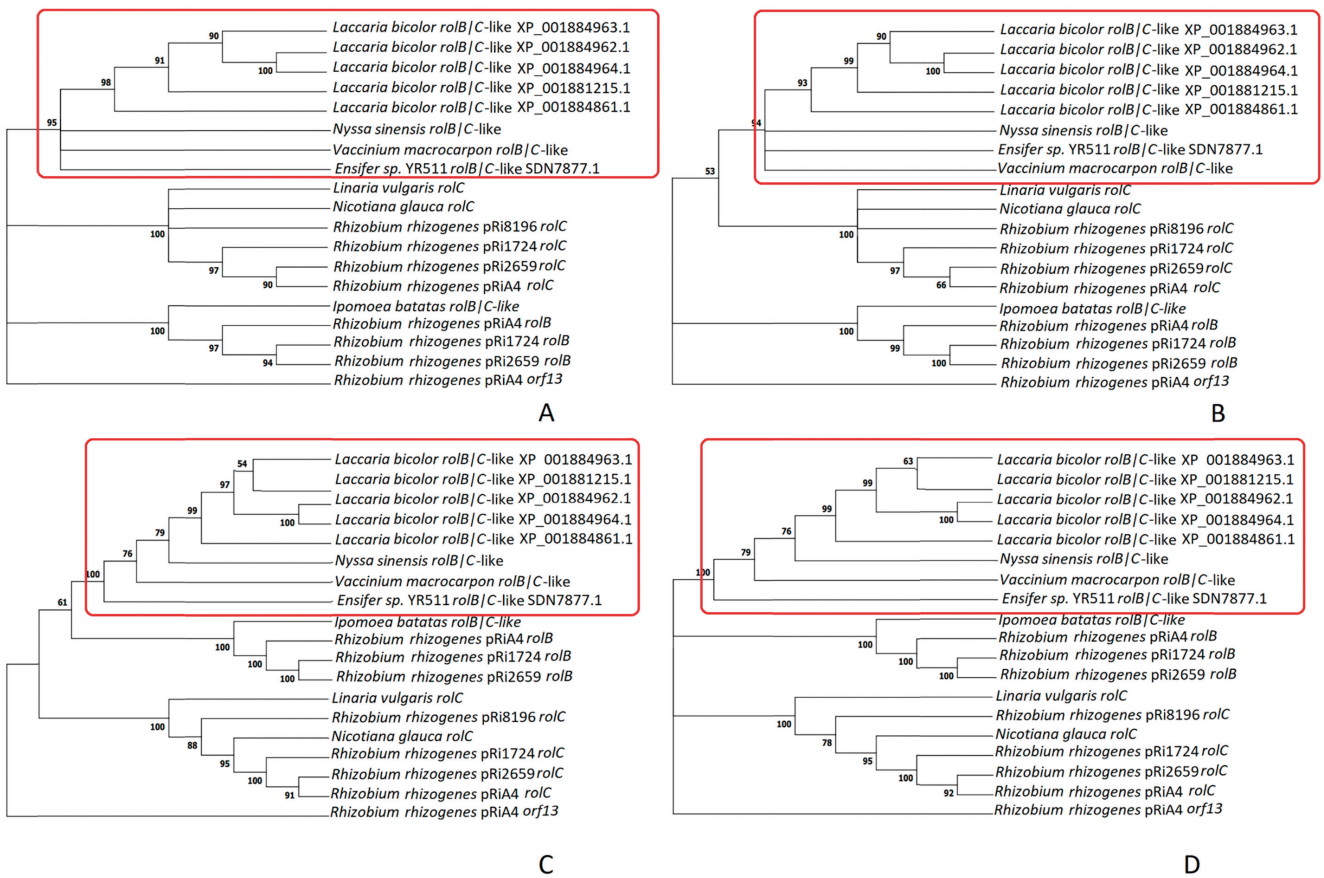
ture of *Agrobacterium* and *Rhizobium* — a reply to Farrand et al. (2003). *International Journal of Systematic and Evolutionary Microbiology* 53(5):1689–1695. https://doi.org/10.1099/ijs.0.02762-0

Zhang, Y., Wang, D., Wang, Y., Dong, H., Yuan, Y., Yang, W., Lai, D., Zhang, M., Jiang, L., and Li, Z. 2020. Parasitic plant dodder (*Cuscuta* spp.): A new natural *Agrobacterium*-to-plant horizontal gene transfer species. *Science China Life Sciences* 63(2):312–316. https://doi.org/10.1007/s11427-019-1588-x

Zuckerkandl, E. and Pauling, L. 1965. Evolutionary divergence and convergence in proteins; pp. 97–166 in V. Bryson and H. J. Vogel (eds.), Evolving genes and proteins. Academic Press, New York. https://doi.org/10.1016/B978-1-4832-2734-4.50017-6

SUPPLEMENTS

Supplementary



Comparison of the topology of phylogenetic trees of *rolB/C* homologs constructed by
A — Maximum Likelihood method based on the JTT matrix-based model (as in fig. 1)
B — Neighbor-joining method based on the JTT matrix-based model
C — UPGMA method based on the Poisson correction model
D — UPGMA method based on the JTT matrix-based model