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.
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 *Eu- terna*, *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 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-Huanquispe 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.
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. (date-plum) 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). *Di*-T-DNA1 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 *Di*-T-DNA5, 7 and 6. *Di*-T-DNA6 is the oldest one. Other traces of multiple acts of agrobacterial transformation in the evolution of ancestral forms of *Diospyros lotus*. 

![Diagram of cT-DNA structures](image-url)
Table 1. New cT-DNAs detected by analysis of WGS database

<table>
<thead>
<tr>
<th>order</th>
<th>family</th>
<th>Species, cultivar, line, isolate</th>
<th>Accession #</th>
<th>Gene homolog</th>
<th>In-tact*</th>
<th>position</th>
<th>Identity level to proteins from NCBI</th>
<th>Similarity level between 2 arms of the cT-DNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>Malpighiales</td>
<td>Salicaceae</td>
<td><em>Populus alba × Populus glandulosa</em> isolate Beijing Shi</td>
<td>SMNX01000141.1</td>
<td>orf13-like</td>
<td>-</td>
<td>98346 - 99995</td>
<td>49 BAB16132.1 <em>Rhizobium rhizogenes</em> (Riker et al. 1930) Young et al. 2001</td>
<td>92</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>rolB-like</td>
<td>-</td>
<td>102808 - 102180**</td>
<td>60 CAA82552.1 <em>R. rhizogenes</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>orf13-like</td>
<td>-</td>
<td>106748 - 104516</td>
<td>49 BAB16132.1 <em>R. rhizogenes</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>acs-like</td>
<td>-</td>
<td>112142 - 112940</td>
<td>86 WP_174054263.1 <em>R. rhizogenes</em></td>
<td></td>
</tr>
<tr>
<td>Fabales</td>
<td>Fabaceae</td>
<td><em>Aeschynomene evenia</em> isolate CIAT22838</td>
<td>RYYW01000009.1</td>
<td>mis-like</td>
<td>+</td>
<td>25927086 - 25926118</td>
<td>62 WP_174054202.1 <em>R. rhizogenes</em></td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td><em>Eperua falcata</em> isolate CIAT22838</td>
<td></td>
<td>mis-like</td>
<td>+</td>
<td>1339 - 149</td>
<td>54 WP_167693616.1 <em>Sinorhizobium melloti</em> (Dangeard 1926) De Lajudie et al. 1994</td>
<td>n/a</td>
</tr>
<tr>
<td>Myrtáles</td>
<td>Myrtáceae</td>
<td><em>Eucalyptus cloeziana</em> isolate ANBG68772</td>
<td>JABKBO010000005.1</td>
<td>mas1'-like</td>
<td>-</td>
<td>19003305 - 19004414</td>
<td>70 P27874.2 <em>R. rhizogenes</em></td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ags-like</td>
<td>-</td>
<td>19005239 - 19004732</td>
<td>69 WP_172690594.1 <em>Agrobacterium sp.</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas2'-like</td>
<td>-</td>
<td>19029872 - 19028766</td>
<td>61 AIM40180.1 <em>R. rhizogenes</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas1'-like</td>
<td>-</td>
<td>19030390 - 19031235 - 19046200 - 19046439</td>
<td>71...47 P27874.2 <em>R. rhizogenes</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ags-like</td>
<td>-</td>
<td>19047360 - 19046844</td>
<td>77 WP_032488587.1 <em>Agrobacterium sp.</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas2'-like</td>
<td>-</td>
<td>19064603 - 19063528</td>
<td>61 AIM40180.1 <em>R. rhizogenes</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas1'-like</td>
<td>-</td>
<td>19065068 - 19066309</td>
<td>66 WP_034520976.1 <em>Agrobacterium sp.</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas2'-like</td>
<td>-</td>
<td>19091525 - 19090591</td>
<td>44 WP_172690593.1 <em>Agrobacterium sp.</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas2'-like</td>
<td>-</td>
<td>19097721 - 19096760</td>
<td>56 WP_172690593.1 <em>Agrobacterium sp.</em></td>
<td></td>
</tr>
<tr>
<td>order</td>
<td>family</td>
<td>Species, cultivar, line, isolate</td>
<td>Accession #</td>
<td>Gene homolog</td>
<td>Intact*</td>
<td>position</td>
<td>Identity level to proteins from NCBI</td>
<td>Similarity level between 2 arms of the cT-DNA</td>
</tr>
<tr>
<td>---------</td>
<td>----------------</td>
<td>-------------------------------</td>
<td>---------------</td>
<td>----------------</td>
<td>--------</td>
<td>-------------------</td>
<td>--------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Boswellia sacra isolate BS-S2</td>
<td>SNVD01001790.1</td>
<td>rolD</td>
<td>-</td>
<td>4554 — 5602</td>
<td>69 WP_034521016.1 Agrobacterium sp.</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>iaaM-like</td>
<td>-</td>
<td>15557-13686</td>
<td>47 WP_174054196.1 R. rhizogenes</td>
<td>87</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>iaaM-like</td>
<td>-</td>
<td>17274 — 19107</td>
<td>45 WP_174054196.1 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>orf10-like</td>
<td>-</td>
<td>19840-20193</td>
<td>40 AAA22094.1 R. rhizogenes</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>HAD hydrolase</td>
<td>?</td>
<td>21778 – 22257</td>
<td>40 WP_149743959.1 Rhizobium sp.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TB-orf14-2-like</td>
<td>-</td>
<td>23456 – 23151</td>
<td>31 AIM40179.1 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>TB-orf14-1-like</td>
<td>-</td>
<td>23561-23151</td>
<td>29 AIM40178.1 Nicotiana tomentosaformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Kewa caespitosa</td>
<td>ONZA01007622.1</td>
<td>mis-like</td>
<td>+</td>
<td>509 — 1510</td>
<td>65 BAB85949.1 Nicotiana glauca</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ONZA01013126.1</td>
<td>mis-like</td>
<td>-</td>
<td>1503 — 2552</td>
<td>62 WP_176453671.1 R. rhizogenes</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ONZA01022047.1</td>
<td>mis-like</td>
<td>-</td>
<td>1924 — 2417</td>
<td>54 WP_176453671.1 R. rhizogenes</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ONZA01003306.1</td>
<td>mis-like</td>
<td>-</td>
<td>4964 — 5356</td>
<td>51 BAB85949.1 Nicotiana glauca</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Pharnaceum exiguum</td>
<td>ONZK01002102.1</td>
<td>mis-like</td>
<td>+</td>
<td>26894 — 27826</td>
<td>66 BAB85949.1 Nicotiana glauca</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>ONZK01003836.1</td>
<td>mis-like</td>
<td>+</td>
<td>7318 — 8232</td>
<td>62 BAB85949.1 Nicotiana glauca</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mis-like</td>
<td>-</td>
<td>14903 — 15822</td>
<td>59 BAB85949.1 Nicotiana glauca</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>Silene noctiflora isolate OPL-1.1</td>
<td>VHZZ01 000004.1</td>
<td>cus-like</td>
<td>-</td>
<td>1170610 – 1176104</td>
<td>60 WP_156551602.1 Allorhizobium vits (Ophe and Kerr 1990)</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VHZZ01 000003.1</td>
<td>cus-like</td>
<td>+</td>
<td>1182125 — 1183057</td>
<td>64 WP_156551602.1 A. vits</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VHZZ01 016788.1</td>
<td>cus-like</td>
<td>+</td>
<td>1210326 — 1209418</td>
<td>64 WP_156551602.1 A. vits</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>VHZZ01 056725.1</td>
<td>cus-like</td>
<td>-</td>
<td>6031 — 5510</td>
<td>72 WP_071208191.1 A. vits</td>
<td>n/a</td>
</tr>
<tr>
<td>Carrot order</td>
<td>Family</td>
<td>Species, cultivar, line, isolate</td>
<td>Accession #</td>
<td>Gene homolog</td>
<td>Intact*</td>
<td>position</td>
<td>Identity level to proteins from NCBI</td>
<td>Similarity level between 2 arms of the cT-DNA</td>
</tr>
<tr>
<td>-------------</td>
<td>--------</td>
<td>---------------------------------</td>
<td>------------</td>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td>------------------------------------</td>
<td>-----------------------------------------------</td>
</tr>
<tr>
<td>Carrot</td>
<td>Nyssa</td>
<td>Nyssa sinensis isolate J267</td>
<td>VIR01000271.1</td>
<td>rolB/C-like</td>
<td>+</td>
<td>5364628 -5365320</td>
<td>52 XP_001881215.1 Laccaria bicolor</td>
<td>n/a</td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01006414.1</td>
<td>cus-like</td>
<td>-</td>
<td>41404 — 42306</td>
<td>73 WP_156551602.1 A. vitis</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>orf14-like</td>
<td>-</td>
<td>44357 -43779</td>
<td>56 WP_174054201.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>orf14-like</td>
<td>+</td>
<td>33878 -337252</td>
<td>56 WP_174054201.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>orf3-like</td>
<td>-</td>
<td>10262 — 11610</td>
<td>70 KEA04445.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>iss transposase</td>
<td>-</td>
<td>12185-12934</td>
<td>83 WP_184141638.1 Shinella fusca Vaz-Moreira et al. 2010</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>issM-like</td>
<td>-</td>
<td>13558-15830</td>
<td>45 WP_034521028.1 Agrobacterium sp.</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>orf8-like</td>
<td>+</td>
<td>2523369 — 2522746</td>
<td>68 WP_116979321.1 Agrobacterium salinitolerans Yan et al. 2017</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>orf3-like</td>
<td>-</td>
<td>2525305-2523957</td>
<td>72 KEA04445.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>orf2-like</td>
<td>-</td>
<td>2525577 -2526392</td>
<td>WP_174054193.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>orcs-like</td>
<td>-</td>
<td>2526938 -2528244</td>
<td>73 KEA04447.1 Agrobacterium sp.</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>sus-like</td>
<td>+</td>
<td>21622 — 22629</td>
<td>77 WP_174080856.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>IS630</td>
<td>-</td>
<td>25887-26667</td>
<td>83 WP_165826447.1 Rhizobium wuzhouense Yuan et al. 2018</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>C-like</td>
<td>-</td>
<td>28060 -26786</td>
<td>76 WP_174054195.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>C-like</td>
<td>-</td>
<td>30052 -31252</td>
<td>76 WP_174054195.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>IS630</td>
<td>-</td>
<td>32218-31443</td>
<td>73 WP_156551602.1 A. vitis</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Ebenaceae</td>
<td>Diospyros lotus cv. Kun-senshi</td>
<td>BEWH01000237.1</td>
<td>sus-like</td>
<td>-</td>
<td>36499-35486</td>
<td>76 WP_174080856.1 R. rhizogenes</td>
<td>n/a</td>
<td></td>
</tr>
<tr>
<td>Order</td>
<td>Family</td>
<td>Species, cultivar, line, isolate</td>
<td>Accession #</td>
<td>Gene homolog</td>
<td>In-tact*</td>
<td>position</td>
<td>Identity level to proteins from NCBI</td>
<td>Organism and protein ID</td>
</tr>
<tr>
<td>-------</td>
<td>--------------</td>
<td>----------------------------------</td>
<td>--------------------------------------------</td>
<td>--------------</td>
<td>----------</td>
<td>------------------------</td>
<td>--------------------------------------</td>
<td>------------------------</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DI-T-DNA5a BEWH01000041.1 (2128496 — 2143594)</td>
<td>99 % identical to BEWH01004217.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>DI-T-DNA6 BEWH01000056.1</td>
<td>sus-like</td>
<td>-</td>
<td>11 77021 -11 77736</td>
<td>58 WP_034521016.1 Agrobacterium sp</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>orf511-like</td>
<td>-</td>
<td>11 79590 -11 78145</td>
<td>60 AIM40183.1 Nicotiana tomentosiformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>orf14-like</td>
<td>-</td>
<td>11 80901 — 11 80354</td>
<td>68 AIM40184.1 Nicotiana tomentosiformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IS3 family transposase</td>
<td>-</td>
<td>11 82902 — 11 81706</td>
<td>70 WP_143239454.1 Agrobacterium roseae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sus-like</td>
<td>-</td>
<td>11 83310 — 11 8402... 11 85015 -11 84824</td>
<td>47 WP_174080856.1 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>IS3 family transposase</td>
<td>-</td>
<td>11 85301 -11 86508</td>
<td>73 WP_143239454.1 Agrobacterium roseae</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>orf511-like</td>
<td>-</td>
<td>11 87393 -11 87790</td>
<td>60 AIM40183.1 Nicotiana tomentosiformis</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>sus-like</td>
<td>-</td>
<td>11 89225 -11 88221</td>
<td>55 WP_174080856.1 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>ocs-like</td>
<td>-</td>
<td>11 90796 — 11 89840</td>
<td>38 KJ92238.1 Laccaria amethystina</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>orf13-like</td>
<td>-</td>
<td>26 2687- 264989</td>
<td>50 WP_174075801.1 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>orf13-like</td>
<td>-</td>
<td>27 1357- 270518... 269030 — 267893</td>
<td>48 WP_174075801.1 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Ericaceae</td>
<td>Vaccinium corymbosum cultivar W6520</td>
<td>JACAOB0100009726.1</td>
<td>rolB/C-like</td>
<td>+</td>
<td>1 — 729</td>
<td>41 XP_001884861.1 Laccaria bicolor</td>
<td>r/a</td>
</tr>
<tr>
<td>order</td>
<td>family</td>
<td>Species, cultivar, line, isolate</td>
<td>Accession #</td>
<td>Gene homolog</td>
<td>Intact*</td>
<td>position</td>
<td>Identity level to proteins from NCBI</td>
<td>Similarity level between 2 arms of the cT-DNA</td>
</tr>
<tr>
<td>-------</td>
<td>--------</td>
<td>---------------------------------</td>
<td>-------------</td>
<td>--------------</td>
<td>---------</td>
<td>----------</td>
<td>------------------------------------</td>
<td>---------------------------------------------</td>
</tr>
<tr>
<td></td>
<td>Convolvulaceae</td>
<td>Ipomoea trifida cultivar Y22</td>
<td>SMMV010000602.1</td>
<td>Contains Ib-TDNA2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Jt-TDNA3 SMMV01000003.1</td>
<td>mas2'-like</td>
<td>-</td>
<td>13842355-13841152</td>
<td>71 WP_032488585.1 Agrobacterium sp.</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas2-like'</td>
<td>-</td>
<td>13842876-13842715</td>
<td>72 AIM40180.1 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas1'-like</td>
<td>-</td>
<td>13845955-13846602</td>
<td>77 P27874.2 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas1'-like</td>
<td>-</td>
<td>13847831-13848160</td>
<td>75 P27874.2 R. rhizogenes</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Convolvulaceae</td>
<td>Ipomoea batatas cultivar Taizhong6</td>
<td>NXFB01008336.1</td>
<td>Contains Ib-TDNA1, described by Kyndt et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>FLTBO1041015.1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>NXFB01000007.1</td>
<td>Contains Ib-TDNA2, described by Kyndt et al. (2015)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>Ib-TDNA3 NXFB01000244.1</td>
<td>mas1'-like</td>
<td>-</td>
<td>54319-53674</td>
<td>77 WP_034520976.1 Agrobacterium sp.</td>
<td>n/a</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas2'-like</td>
<td>-</td>
<td>56951-58017</td>
<td>68 WP_032488585.1 Agrobacterium sp.</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>mas2'-like</td>
<td>-</td>
<td>27628510-27627275</td>
<td>63 AIM40180.1 R. rhizogenes</td>
<td>n/a</td>
</tr>
</tbody>
</table>

* does not contain premature stop codons and/or frame shift
** gene location on the negative strand
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-genome 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 glandulosus* 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. Kunzenshi 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

Supplements

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