Euphytica (2020) 216:186
https://doi.org/10.1007/s10681-020-02722-7

Ancient horizontal gene transfer from *Rhizobium rhizogenes* to European genera of the Figwort family (Scrophulariaceae)

Sophia Schiermacher Stavnstrup · Jaime Pérez Molina · Henrik Lütken · Renate Müller · Josefine Nymark Hegelund

Received: 14 January 2020 / Accepted: 25 October 2020 / Published online: 8 November 2020
© Springer Nature B.V. 2020

Abstract *Rhizobium rhizogenes* exploits horizontal gene transfer as part of its mechanism of pathogenesis. In this respect, genetic material, transfer DNA (T-DNA) from the bacterium, is transferred transiently to the host plant genome. However, plant species within *Nicotiana, Linaria* and *Ipomea* genera contain genomic traces of ancient bacterial T-DNA. To determine if ancient bacterial T-DNA is present in uncultivated plants of European genera within Scrophulariaceae, seeds of *Linaria, Antirrhinum*, *Digitalis*, and *Veronica* were analysed for the presence of one of the *root oncogenic loci* genes; *rolC* of *R. rhizog*enes. This study discloses remnants of ancient *rolC* haplotypes in several species of *Linaria*, *Antirrhinum*, *Digitalis* and *Veronica.* The distribution of plant species harbouring *rolC* sequences within each genus was not uniform. In total, 7 of the 16 investigated species were found to be positive for at least one of the *rolC* haplotypes, where 6 of these are new additions to the group of naturally transformed plants. Sequence alignment showed high interspecies homology of *rolC*. Five unique *rolC* haplotypes (*rolCa-rolCe*) were found in several plant species. The most abundant, *rolCa*, was identified in all the plant species holding ancient bacterial DNA. Transcripts of *rolC* were not detected in leaves, which indicates that *rolC* may not have a function in leaves under non-stress conditions. This study adds *Antirrhinum*, *Digitalis* and *Veronica* to the list of plant genera within Scrophulariaceae which have been subjected to ancient transformation events via horizontal gene transfer from bacteria to plants.

Keywords *Agrobacterium rhizogenes* · Horizontal gene transfer · Natural transformation · Plantaginaceae · *rol*-genes

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s10681-020-02722-7) contains supplementary material, which is available to authorized users.

S. S. Stavnstrup · J. P. Molina · H. Lütken · R. Müller · J. N. Hegelund (✉)
Department of Plant and Environmental Sciences, University of Copenhagen, Højbakkegård Allé 9-13, 2630 Taastrup, Denmark
e-mail: jnh@plen.ku.dk

Introduction

Horizontal gene transfer (HGT) allows beneficial genetic plasticity within microbial communities. Via HGT, microbes exchange genetic material among unrelated organisms; this allows an efficient adaptation to immediate changes in the environment independent of vertical gene transfer among parents and offspring. HGT confers selective advantages to prokaryotic ecosystems e.g. resistance to antibiotics, virulence, photosynthesis, or nitrogen fixation (Syvanen 1994). In comparison, HGT in eukaryotes is much

less predominant and occurs as intracellular gene transfer from organelles to the nucleus or with bacteria as donors (White et al. 1982; Bergthorsson et al. 2004; Richardson and Palmer 2006).

In plants, foreign DNA can be received by HGT from viruses, bacteria or even from other parasitic plants (Richardson and Palmer 2006). A well-known example of HGT from bacteria to plants happens when phytopathogenic *Rhizobium* (previously *Agrobacterium*) infects its host. The phytopathogenic subphylum of *Rhizobium* is a gram-negative genus of soil bacteria and its pathogenesis is dependent on the transfer of a plasmid-borne DNA fragment (T-DNA) which is integrated into the host genome by HGT. Some of the most predominant species within the genus are *R. radiobacter*, *R. rhizogenes*, *R. rubi* and *R. vitis* each of which causes distinct abnormal tissue growth such as crown gall tumours, hairy root disease, cane gall disease and neoplastic tumours, respectively (Riker 1930; Hildebrand 1940; Ophel and Kerr 1990). Upon transfer of T-DNA to the plant genome, opines are produced to supply the pathogen with a nitrogen and carbon source. Opines are derived from plant amino acids and their synthesis is controlled by genes located on the T-DNA of bacterial origin (Hong et al. 1997). Disease infected plant cells are not expected to transfer T-DNA to subsequent generations of plants as they are in a determined state. Therefore, they will not develop zygotes or gametes, which constitute the elements of plant embryogenesis and hence will not regenerate into whole plants (Goldberg et al. 1994; Mordhorst et al. 1997).

Reports of the occurrence of T-DNA fragments of bacterial origin in plant genomes have increased during the last decades. Initially, cellular T-DNA (cT-DNA) was detected in uninfected *Nicotiana glauca* (White et al. 1982). A later expanded analysis of multiple *Nicotiana* species revealed cT-DNA in *N. tabacum*, *N. tomentotiformis*, *N. tomentosa* and *N. otophora* (Fürner et al. 1986). Similarly, Russian species of *Linaria* namely *L. vulgaris*, *L. genistifolia* and *L. creticola* contain cT-DNA like sequences from *R. rhizogenes* (Matveeva et al. 2012, 2018; Pavlova et al. 2014). Interestingly, via the emergence of plant genomic and transcriptomic sequence data online, numerous additional natural transformants have been detected. A transcriptome study on cultivated sweet potato (*Ipomoea batatas*) identified several insertions of *R. rhizogenes* cT-DNA homologues in both *I. batatas* and in the wild relative *I. trifida* (Kyndt et al. 2015). Also, bioinformatic analysis of sequence data from more than 600 dicot species indicated that at least 49 domesticated plant species can be added to the list of plants containing variations of cT-DNA (Matveeva and Otten 2019).

Scrophulariaceae is a family with a worldwide geographical distribution. Albach et al. (2005) revised and restructured the taxonomy of the family using nuclear and plastid DNA sequence analysis, suggesting Scrophulariaceae as a part of Plantaginaceae. *Linaria* belongs to the Scrophulariaceae *sensu stricto* taxonomic rank but despite the ongoing studies on the evolution of Plantaginaceae, no phylogenetic certainty has been defined yet (Fernandez-Mazuecos et al. 2013; Vigalondo et al. 2015). According to Albach et al. (2005) the current version of Plantaginaceae *sensu lato*, it contains approximately 92 genera and 2000 species, including *Veronica* as the largest genus.

Previous studies on ancient bacterial *rolC* sequences primarily focused on the Scrophulariaceae *sensu stricto* clade and the presence and distribution of *R. rhizogenes* cT-DNA fragments were revealed in Russian accessions of *Linaria* (Matveeva et al. 2012, 2018; Pavlova et al. 2014). It is therefore of great interest to investigate vaster phylogenetic and geographical areas. This study was undertaken to determine the distribution of ancient bacterial *rol* genes within uncultivated plants of Scrophulariaceae in Europe. Plant materials were seeds from Scrophulariaceae collected in Denmark and seeds from botanical gardens in Germany and England. Here, we report the discovery of ancient bacterial *rolC* sequences in three additional genera: *Antirrhinum*, *Digitalis* and *Veronica* within the Figwort family (Scrophulariaceae) in addition to those disclosed previously in *Linaria*.

Materials and methods

Plant materials and growth conditions

Seeds of the *Antirrhinum*, *Digitalis*, *Linaria* and *Veronica* genera were obtained from Botanischer Garten der Universität Hamburg, Germany; The Royal Botanic Gardens, Kew, England; The University Gardens, Frederiksberg Campus, Faculty of Science, University of Copenhagen (UCPH-F) and the seed

collection available at Taastrup Campus, Faculty of Science (UCPH-T). Furthermore, uncultivated plants were collected in North-Zealand, Denmark at the following GPS positions; 56°04′ 14.4″ N 12°33′ 03.1″ E, 56°02′ 57.2″ N 12°26′ 27.4″ E and 55°40′ 06.3″ N 12°18′ 12.9″ E. Plant species and origin of seeds are presented in Table 1. Additionally, a range of other plant samples were assessed from *Plantago* species of Danish and Greenlandic origin as well as *Wulfenia carinthiaca* (data not shown).

The majority of seeds were germinated in pools of 5–15 seeds in 5 cm × 5 cm pots in a greenhouse. Fertilisers were added daily with every watering (Brun Komplet Garta A/S, Denmark). Plants were grown with a minimum temperature of 20 °C and a 16-h photoperiod of natural light, supplemented with artificial light (190–220 μmol $m^{-2}s^{-1}$). Plants were grown a full life cycle to ensure correct morphological identification (Fig. 1).

Seeds unable to germinate in soil were surface-sterilised and germinated in vitro. Following sterilisation (1 min 70% (v/v) EtOH, 3 min 2–5% (v/v) NaOCl and rinse 3 times in sterile water), seeds were placed in plant boxes containing 3% sucrose (w/v), 0.7% Plant agar (w/v), 4.4% (w/v) Murashige and Skoog including vitamins and 0.05% 2-(N-morpholino) ethanesulfonic acid (MES, w/v). In vitro boxes were placed in a growth chamber with 120–150 μmol $m^{-2}s^{-1}$ light intensity at 25 °C in an 8 h photoperiod. Seeds germinated in vitro were from Kew, UK; *L. repens*, *V. agrestis*, *V. arvensis*, and *V. chamaedrys*. Plantlets were transferred to soil 3–4 weeks after germination and grown to maturity.

DNA extraction

Leaf material for DNA extractions was obtained from the youngest fully developed leaf of approximately 3 months old plants. Harvested plant material was frozen in liquid N_2 and stored at − 80 °C. All DNA extractions were performed twice from independent plants using the Shorty method developed for the identification of T-DNA insertion mutants in Arabidopsis (Visscher et al. 2010) with modifications described in Hegelund et al. (2018). DNA concentrations were determined by a Nanodrop™ 1000 spectrophotometer (Thermo Fisher Scientific Inc., USA).

Primer design and detection of *rolC* sequences

In this study, primers specific for *rolC* were designed to align to regions conserved among *Linaria* and *R. rhizogenes rolC* sequences. Specifically, open reading frames of *rolC*, derived from the NCBI Genbank (Benson et al. 2013) and originating from *L. genistifolia* (KC309424), *L. vulgaris* (EU735069) and *R.*

Table 1 Suppliers of seeds of wild species of *Antirrhinum*, *Digitalis*, *Linaria* and *Veronica* genera

Supplier	Origin of seeds	Species
Botanischer Garten der Universität Hamburg, Germany	Germany (DE)	*A. braun-blanquetii* and *L. vulgaris*
Royal Botanic Gardens (KEW), England	England (UK)[a]	*D. purpurea*[a], *L. arenaria, L. purpurea, L. repens, L. supina, L. triornithophora, L. vulgaris, V. agrestis, V. arvensis, V. beccabunga, V. chamaedrys, V. persica* and *V. serpyllifolia*
The University Gardens, Frederiksberg Campus, Faculty of Science, University of Copenhagen (UCPH-F), Denmark	Denmark (DK)	*A. majus*
Seed collection, Taastrup Campus, Faculty of Science, University of Copenhagen (UCPH-T), Denmark	Denmark (DK)	*L. minor, V. agrestis, V. arvensis* and *V. persica*
Collected, Zealand, Denmark	Denmark (DK)[b, c, d]	*D. purpurea*[b,c], *L. vulgaris*[c], *V. agrestis*[c], *V. arvensis*[b,c], *V. chamaedrys*[d]

[a]*D. purpurea* supplied by Kew were of DE, ES, UK and DK origin

[b]GPS position 56° 04′ 14.4″ N 12° 33′ 03.1″ E

[c]GPS position 56°02′ 57.2″ N 12° 26′ 27.4″ E

[d]GPS position 55°40′ 06.3″ N 12° 18′ 12.9″ E

Fig. 1 Flowers of selected Scrophulariaceae species. **a** *Antirrhinum majus*. **b** *Linaria purpurea*. **c** *Linaria vulgaris*. **d** *Veronica agrestis*

rhizogenes (EF433766), were aligned using Clustal Ω (Madeira et al. 2019). Conserved regions were selected as primer annealing sites. Primers applicable as positive controls in PCRs and as reference genes in RT-PCR were designed similarly. Sequences of the Rubisco large subunit (RBC) were obtained via the NCBI Genbank from *Anthirrhinum majus* (L11688), *Brassica napus* (JF807908), *L. repens* (MG222642), *L. vulgaris* (KM360853), *Veronica dabneyi* (HM850449) and *V. officinalis* (AY034024). RBC specific primers suitable for use in *Anthirrhinum*, *Linaria* and *Veronica* genera were designed using Primer3Plus software (Untergasser et al. 2012). In *Digitalis*, primers designed to amplify *Digitalis purpurea ACT2* (HQ853642) functioned as positive control and reference gene.

For the initial detection of *rolC*, PCR reactions included 100–250 ng of genomic DNA, 2% (v/v) DMSO and Ex Taq polymerase as recommended by supplier (Takara Bio Inc., Japan). Cloning and subsequent sequencing reactions were based on PCR products amplified with the proof-reading polymerase LA Taq using 2% DMSO (v/v) and following supplier's instructions (Takara Bio Inc.). PCR reactions followed the program 94 °C for 4 min, 35 cycles of [30 s at 94 °C, 30 s at 60 °C, 30 s at 72 °C] and a last step of 72 °C for 7 min in a MyCycler (Bio-Rad) or a Master Cycler Gradient (Eppendorf, Germany). Primer sequences and PCR program exceptions due to primer specific annealing temperatures (Ta) are listed in Table 2.

In reactions amplifying genes or transcripts of bacterial origin, positive controls used purified *R. rhizogenes* plasmid *pRiA4* as template. Except in cases where *rolCexpr* primers were used, here gDNA of *L. vulgaris* was used as positive control. In all reactions,

Table 2 PCR and RT-PCR primers specific for *ACT*, *RBC* *rolC* and *virD2*. *RBC* served as positive control/reference gene in *Anthirrhinum*, *Linaria* and *Veronica* genera. *ACT* was used as positive control/reference in *Digitalis*. Temperature of annealing in PCR and RT-PCR (Ta) and product sizes are presented

Gene	Forward	Reverse	Ta (°C)	Size (bp)
ACT	5′-CGAAAAATAGTGCGGGACAT-3′	5′-TGGTTTCATGGATACCAGCA-3′	60	220
RBC	5′-GACAACTGTGTGGACCGATG-3′	5’-ATTCGCAGATCTTCCAGACG-3’	58	222
rolC	5’-CGCACTCCTCACCAACCT-3’	5’-TGCTGGCATAGAGGTCGAAT-3’	60	221
rolCexpr[a]	5’-GACGTGACATGCAGCGATGA-3’	5’-GTTAGCTCCATCTGCTCATTC-3’	55	385
virD2	5’-AGTCGTCATAGCAAGGAGAT-3’	5’-TGTCTCCAATGCAATCCGTA-3’	63	189

[a]Matveeva et al. 2018

H_2O was used as a negative control. To verify the absence of bacterial contamination in the extracted DNA, *virD2* detection was used as an additional negative control (data not shown).

Cloning and sequencing of *rolC*

PCR products amplified by LA taq were cloned into the PCR4 TOPO TA vector supplied via the TOPO TA cloning kit® (Invitrogen, USA) according to the manufacturer’s instructions. Plasmids were purified using GenElute Plasmid DNA Miniprep Kit (Sigma-Aldrich) as recommended by manufacturer. Inserts were verified by *EcoRI* digestion (New England BioLabs Inc., USA) and subsequent gel electrophoresis. Purified plasmids having inserts of the expected size (221 bp) were sequenced by Eurofins Genomics, Germany.

Expression analyses of *rolC*

A group of *rolC* positive species available from KEW were selected for expression analyses; *A. majus*, *D. purpurea*, *L. purpurea*, *L. vulgaris*, *V. agrestis* and *V. beccabunga*. RNA extractions were done from 80 to 100 mg tissue of the youngest fully expanded leaves of plants in the vegetative or the generative growth stage using RNeasy Plant Mini kit as recommended by the supplier (Qiagen). RNA yield and purity were estimated using a Nanodrop™ 1000 spectrophotometer. RNA integrity was verified on 1% bleach agarose gels as described by Aranda et al. (2012). Prior to expression analyses, 1 µg of RNA was treated with DNase I Amplification Grade (Sigma-Aldrich, USA) and cDNA synthesis were done via the iScript cDNA synthesis kit (Bio-Rad, USA) as recommended. cDNA was diluted 5-fold in RNase/DNase free Tris-EDTA pH 7.4 (Sigma-Aldrich) before use. Expression analyses were done by RT-PCR with primers specific for *rolC* produced in this study or the *rolC* expression primers (*rolCexpr*) designed by Matveeva et al. (2018). Positive controls were *RBC* for cDNA of *Linaria*, *Antirrhinum* and *Veronica* and *ACT* for cDNA of *Digitalis*. Primer details are presented in Table 2. RT-PCR reactions were conducted using Ex Taq polymerase as previously described.

Bioinformatics

Sequence identification, alignments and analyses were done using CLC Sequence viewer (Qiagen), BLAST and Clustal Ω (Coordinators 2013; Madeira et al. 2019). Phylogenetic analyses were conducted using the Maximum Likelihood method and Tamura-Nei model in MEGA-X (Tamura and Nei 1993; Kumar et al. 2018). Bootstrap values were inferred by tests of 1000 replicates. As references *rolC* homologs of *L. vulgaris* (*LvrolC* [Genbank: EU735069]), *L. creticola* (*LcrolC* [Genbank: MF997051]), *L. genistifolia* (*LgrolC* [Genbank: KC309424]), *N. tabacum* (*NtrolC* [Genbank: X91881]), *N. glauca* (*NgrolC* [Genbank: X03432]) and *R. rhizogenes* (*RrrolC* [Genbank: EF433766]) were used. Genbank accession numbers for sequences described here are presented in the Data Availability statement.

Results

rolC in European Scrophulariaceae species

This research was initiated to determine the distribution of ancient bacterial DNA within uncultivated plants of the Scrophulariaceae family in Europe. PCR primers were designed to anneal to conserved regions of available *rolC* sequences of *R. rhizogenes* and *Linaria*. Using PCR, *rolC* homologues were identified in plant species of the *Antirrhinum*, *Digitalis*, *Veronica* genera and in an additional species of the *Linaria* genus, *Linaria purpurea* (Table 3). Additionally, to further explore the presence of *rolC* homologs, samples of *Plantago* spp. and *Wulfenia carinthiaca* were assessed but did not contain *rolC* sequences (data not shown). To complement the studies of *rolC* sequences in Russian *Linaria* species (Matveeva et al. 2012, 2018; Pavlova et al. 2014), plants of this study were collected from European sources (Table 1). Of the 15 new species included here, 6 were positive for the presence of *rolC* sequences of bacterial origin. There was no correlation between the presence of *rolC* sequences and origin of the plant material (Table 3).

Within each genus, the presence of *rolC* sequences varied (Table 3). *A. majus* contained *rolC* homologues, whereas *A. braun-blanquetii* did not. In *Digitalis*, *D. purpurea* contained several *rolC* sequences. From the *Linaria* genus, *L. purpurea* and *L. vulgaris* were tested positive for *rolC*, whereas no *rolC* sequences were detected in *L. arenaria*, *L. minor*, *L. repens*, *L. supina* and *L. triornithophora*. Finally, within the *Veronica* genus, *V. agrestis*, *V. beccabunga* and *V. chamaedrys* carried several *rolC* haplotypes

Table 3 Uncultivated plant species of *Antirrhinum*, *Digitalis*, *Linaria* and *Veronica* genera tested for the presence of *rolC* sequences of bacterial origin. *rolC* haplotype a-e; common *rolC* homologous, u; unique *rolC* sequences, nd; not detected. New plant species found to contain ancient bacterial *rolC* sequences are marked in bold. Accessions denoted with a * were included in expression analyses. All accessions were analysed twice from independent DNA extractions

Genus	Species	Abbreviation	Origin	*rolC* haplotype
Antirrhinum	*braun-blanquetii*	ABB	DK	nd
Antirrhinum	***majus****	AMA	DK	a, u
Digitalis	***purpurea****	DPU1	DK	a, u
Digitalis	***purpurea****	DPU2	UK	a, b
Digitalis	***purpurea****	DPU3	ES (Kew)	a, d
Digitalis	***purpurea****	DPU4	DE	b, e, u
Digitalis	***purpurea***	DPU_CO	DK[a, b]	a, b, d
Linaria	*arenaria*	LAR	UK	nd
Linaria	*minor*	LMI	DK	nd
Linaria	***purpurea****	LPU	UK	a
Linaria	*repens*	LRE	UK	nd
Linaria	*supina*	LSU	UK	nd
Linaria	*triornithophora*	LTR	UK	nd
Linaria	*vulgaris**	LVU	UK	d, u
Linaria	*vulgaris*	LVU_CO	DK[b]	a, b, c, u
Veronica	***agrestis****	VAG	UK	a, b, u
Veronica	***agrestis***	VAG_CO	DK[b]	a, b, u
Veronica	*arvensis*	VAR1	UK	nd
Veronica	*arvensis*	VAR2	DK	nd
Veronica	*arvensis*	VAR_CO	DK[a, b]	nd
Veronica	***beccabunga****	VBE	UK	a, b
Veronica	***chamaedrys***	VCH	UK	a, e
Veronica	***chamaedrys***	VCH_CO	DK[c]	a, b, c
Veronica	*persica*	VPE1	UK	nd
Veronica	*persica*	VPE2	DK	nd
Veronica	*serpyllifolia*	VSE	UK	nd

[a]GPS position 56°04'14.4″ N 12°33'03.1″ E

[b]GPS position 56°02'57.2″ N 12°26'27.4″ E

[c]GPS position 55°40'06.3″ N 12°18'12.9″ E

whereas in *V. arvensis*, *V. persica* and *V. serpyllifolia rolC* sequences were not detected.

Sequence variations of ancient *rolC* in Scrophulariaceae

Comparison of the isolated *rolC* sequence from different Scrophulariaceae genera revealed five different haplotypes. The haplotypes cannot be associated to any specific plant species but are distributed across the genera (Fig. 2). Most prominent was *rolCa* which was present in *A. majus, D. purpurea*, *L. purpurea*, *L. vulgaris*, *V. agrestis*, *V. beccabunga* and *V. chamaedrys*. When aligned to the 221 bp *rolCa* fragment identified here, *rolCb*, *rolCc*, *rolCd* and *rolCe* exhibited sequence variations compared to *rolCa* of 2, 3, 13 and 14 nucleotides, respectively. *rolCd* and *rolCe* differed by a single nucleotide. Additionally, seven sequences were unique for individual species but nevertheless, closely related to *rolCa-rolCe* (Table 3, sequence data not shown). Species having species-specific *rolC* haplotypes are

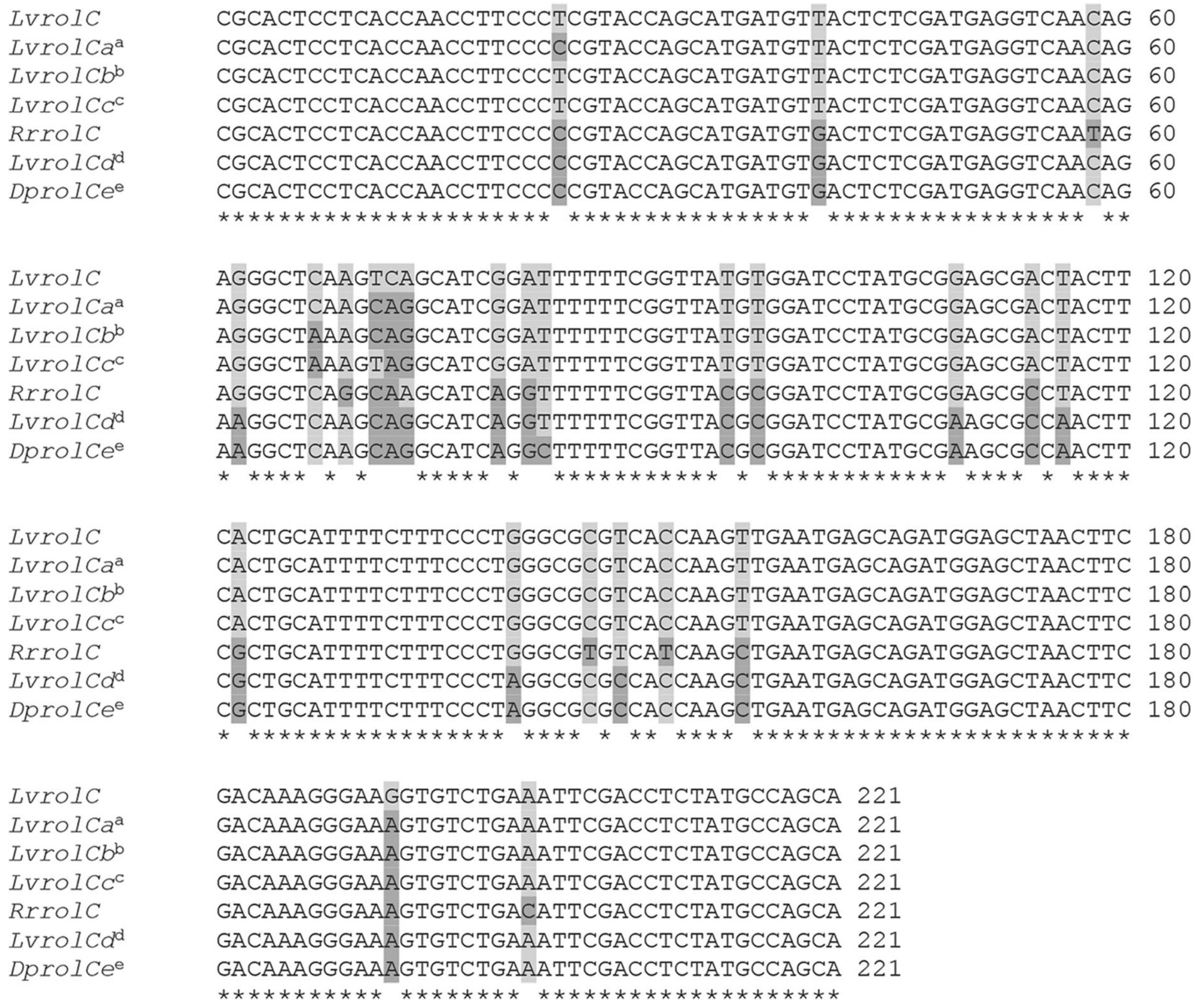

Fig. 2 Nucleotide alignment of partial genomic fragments of *rolC* sequences found in Scrophulariaceae. [a]*LvrolCa* was found in *L. vulgaris*, *A. major*, *D. purpurea*, *L. purpurea*, *V. agrestis*, *V. beccabunga* and *V. charmeadrys*. [b]*LvrolCb* was found in *L. vulgaris*, *D. purpurea*, *V. agrestis*, *V. beccabunga* and *V. charmeadrys*. [c]*LvrolCc* was found in *L. vulgaris* and *V. charmeadrys*. [d]*LvrolCd* was found in *L. vulgaris* and *D. purpurea*. [e]*DprolCe* was found in *D. purpurea*, and *V. charmeadrys*. As references corresponding fragments of *rolC* homologs from *L. vulgaris* (*LvrolC* [Genbank: EU735069]) and *R. rhizogenes* (*RrrolC* [Genbank: EF433766]) are included. The alignment was produced in Clustal Ω (Madeira et al. 2019)

documented in Table 3, common for these were that only single nucleotide changes made them differ from the *rolCa-e* nucleotides.

Phylogenetic analyses including GenBank reference sequences of *rolC* from *R. rhizogenes*, *Nicotiana* and *Linaria*, and the *rolCa-rolCe* sequences identified here showed a Scrophulariaceae specific sequence cluster. Also, the Scrophulariaceae *rolC* sequence cluster showed a closer phylogenetic relationship to the *R. rhizogenes rolC* gene than to the *Nicotiana rolC* reference sequences (Fig. 3).

Expression analyses of ancient *rolC* sequences in Scrophulariaceae

The *rolC* positive species *L. vulgaris*, *L. purpurea*, *A. majus*, *V. beccabunga*, *V. agrestis* and *D. purpurea*, were analysed in expression analyses to identify *rolC* transcripts (Table 3). RNA was extracted from leaves in vegetative and reproductive growth stages, and RNA quality and integrity were verified experimentally (Supplementary table 1 and Supplementary Fig. 1). Following cDNA synthesis, RT-PCR was conducted using the appropriate reference genes as control. However, no transcripts of ancient *rolC* could be detected in any of the species investigated (Supplementary Fig. 2).

Discussion

Horizontal gene transfer mediated by phytopathogenic species of *Rhizobium* has resulted in naturally transformed plant species belonging to the *Nicotiana*, *Ipomea* and *Linaria* genera (White et al. 1982; Fürner et al. 1986; Matveeva et al. 2012; Kyndt et al. 2015). How T-DNA mechanistically moves from a stable genomic integration in root cells into the genome of sexually transmitted cells is not clear. Additionally, the occurrence of natural transformation in nature on the plant species level remains elusive, as the genomes of most plant species are unknown. In this study, European species of Scrophulariaceae were screened for the presence of *rolC* of bacterial origin to expand

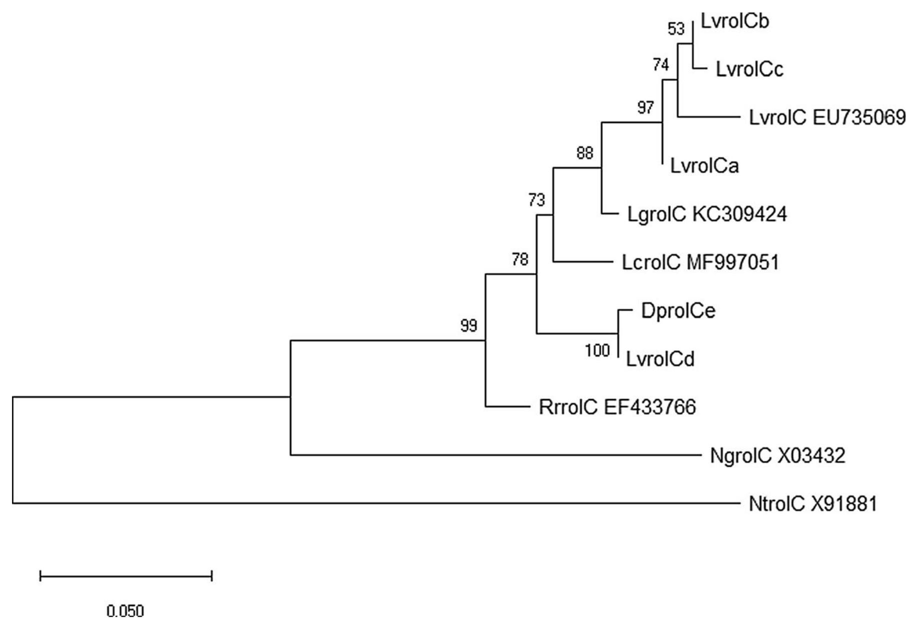

Fig. 3 Phylogenetic analyses of partial genomic fragments of *rolC* sequences in Scrophulariaceae. *LvrolCa* was found in *L. vulgaris*, *A. major*, *D. purpurea*, *L. purpurea*, *V. agrestis*, *V. beccabunga* and *V. charmeadrys*. *LvrolCb* was found in *L. vulgaris*, *D. purpurea*, *V. agrestis*, *V. beccabunga* and *V. charmeadrys*. *LvrolCc* was found in *L. vulgaris* and *V. charmeadrys*. *LvrolCd* was found in *L. vulgaris* and *D. purpurea*. *DprolCe* was found in *D. purpurea*, and *V. charmeadrys*. As references corresponding fragments of *rolC* homologs from *L. vulgaris* (*LvrolC* [Genbank: EU735069]), *L. creticola* (*LcrolC* [Genbank: MF997051]), *L. genistifolia* (*LgrolC* [Genbank: KC309424]), *N. tabacum* (*NtrolC* [Genbank: X91881]), *N. glauca* (*LvrolC* [Genbank: X03432]) and *R. rhizogenes* (*RrrolC* [Genbank: EF433766]) are included. The phylogenetic analysis was produced in Mega-X (Kumar et al. 2018)

the understanding of the occurrence of horizontal gene transfer with respect to geographical regions and species.

rolC in European Scrophulariaceae

This study supports results obtained by Matveeva et al. (2012) who were the first to identify cT-DNA of bacterial origin in *L. vulgaris* of Russian origin. In the current study, not only *L. vulgaris* of European origin, but also *L. purpurea*, contains a *rolC* haplotype in its genome. Furthermore, three new genera within the same family as *L. vulgaris* can now be added to the list of naturally transformed plants namely *Antirrhinum*, *Digitalis* and *Veronica*. The distribution of Scrophulariaceae species harbouring *rolC* sequences within each genus is however not uniform.

In the study by Matveeva et al. (2012) which identified *L. vulgaris* as a naturally transformed plant, seven additional *Linaria* species were tested and did not contain cT-DNA (Matveeva et al. 2012). Here, *L. purpurea* contains *rolC* suggesting a close phylogenetic relationship between *L. purpurea* and *L. vulgaris* (Vargas et al. 2004). Experimental data from 2013 indicate that *L. purpurea and L. repens* should be in the same clade as *L. vulgaris*. However, no *rolC* sequences were detected in *L. repens* (Fernandez-Mazuecos et al. 2013).

The ability to detect cT-DNA in plants is not related to the phylogenetic relationship of the plant species. Collectively we confirmed, one species of *Digitalis*, one out of two *Antirrhinum* species and three out of seven *Veronica* species to contain fragments of *rolC* sequences. This is a more uneven occurrence of *rolC* sequences than expected if close phylogenetic ties are the determining factor. The occurrence of multiple independent transformation events within different genera of the same family would require an uncharacterized ability of Scrophulariaceae members to regenerate themselves from the initially transformed tissue e.g. regeneration of intact plants from diseased roots (hairy roots) (Chen and Otten 2017). Sequence analyses of *rolC* haplotypes showed high interspecies conservation within 5 unique *rolC* sequences (*rolCa-rolCe*) thus we speculate that the original transformation event or events could have occurred early in the evolution of Scrophulariaceae prior to the differentiation of the individual genera (Figs. 2, 3). This would however require that in several of the investigated plant species, parts of the cT-DNA have been lost again during speciation—or at least parts specific to our primers have been lost. To further investigate the hypothesis of early transformation events within Scrophulariaceae before genus differentiation, we assessed *Wulfenia carinthiaca* for presence of *rolC* sequence fragments. *W. carinthiaca* is a Miocene (approx. 23−5 MA) relic plant with disjunct distribution in Europe and could be a potential ancestor within Scrophulariaceae (Surina et al. 2014). However, our study did not identify *rolC* in *W. carinthiaca* (data not shown) and further studies are needed in that respect.

Antirrhinum and *Linaria* are phylogenetically closely related (Albach et al. 2005) but remnants of cT-DNA have not previously been detected in the *Antirrhinum* genus. As for *Antirrhinum*, *rolC* sequences have not been reported in the genus of *Digitalis*, but we have isolated all but one of the five *rolC* haplotypes, plus unique sequences, from different sources of *D. purpurea*. *V. chamaedrys* was included on the list of species screened for cT-DNA by Matveeva et al. in 2012, but *rolC* was not detected. This could be due to differences in primers used. Alternatively, the Russian population of *V. chamaedrys* does not contain *rolC* homologues, whereas the European *V. chamaedrys* does. In summary, data presented here identify *rolC* sequences in four genera within the Scrophulariaceae family *sensu stricto*. However, as we only preliminarily tested other genera within the family of Plantaginaceae *sensu lato*, further studies are needed to clarify how widespread the natural transformation event or events have become within Scrophulariaceae and Plantaginaceae.

Expression of *rolC*

Previously, *rolC* transcripts were seen at low levels in shoots and calli from in vitro grown plantlets of *L. vulgaris* (Matveeva et al. 2018). In the current study, the expression of *rolC* was investigated in leaves of *L. vulgaris* and five species revealed in this study to contain *rolC* sequences, but no transcript was detected (Table 3, Supplementary Fig. 2). This indicates that Scrophulariaceae *rolC* homologues are not expressed in leaves of plants grown *in vivo* but our results cannot exclude the presence of transcripts in other tissues such as root, meristem, or flower tissue.

Perspectives

This study demonstrates that uncultivated plants of European *Antirrhinum*, *Digitalis* and *Veronica* contain *rolC* haplotypes derived from ancient bacterial genes, however other putative sequences of the inserted cT-DNA have not been characterised. To decide if these cT-DNA in Scrophulariaceae share a common origin, the positions of the integration sites need to be determined in each plant species. Also, more sequence information of the Scrophulariaceae cT-DNA is needed to fully address why some species closely related to natural transformants in the same genus do not appear to contain *rolC* sequences in this study. Furthermore, it is also important to determine which *R. rhizogenes* strain(s) facilitated the transformation events.

Although this study raises new questions it is a critical step on the way to understanding the dynamics among plant species holding cT-DNA of ancient bacterial origin.

Acknowledgements This work was supported by the Danish Agency for Science and Higher Education as part of the innovation consortium "Innovative Plants". Botanischer Garten der Universität Hamburg in Germany, Royal Botanic Gardens (KEW) in England and the University Gardens, Frederiksberg Campus and the seed collection, Taastrup Campus, Faculty of Science, University of Copenhagen in Denmark are acknowledged for providing plant material. Martin Himmelboe and Mikkel Kok Jensen are acknowledged for design of primers. We thank Jesper Cairo Westergaard for proofreading the article.

Author contributions SSS, JPM, RM and JNH conceived and designed the study. SSS and JPM planned and performed experiments with assistance of JNH and HL. SSS and JNH wrote the manuscript. All authors critically revised the manuscript, read and approved the final manuscript.

Funding This study was funded by the Danish Agency for Science and Higher Education as part of the innovation consortium "Innovative Plants".

Data availability Sequence data that support the findings of this study have been deposited in GenBank with the accession codes *LvRolCa* (MN097818), *LvRolCb* (MN097819), *LvRolCc* (MN097820), *LvRolCd* (MN097821), *LpRolCa* (MN097822), *AmRolCa* (MN097823), *DpRolCa* (MN097824), *DpRolCb* (MN097828), *DpRolCd* (MN097833), *DpRolCe* (MN097834), *VaRolCa* (MN097825), *VaRolCb* (MN097829), *VbRolCa* (MN097826), *VbRolCb* (MN097830), *VcRolCa* (MN097827), *VcRolCb* (MN097831), *VcRolCc* (MN097832) and *VcRolCe* (MN097835).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Albach DC, Meudt HM, Oxelman B (2005) Piecing together the "new" Plantaginaceae. Am J Bot 92:297–315

Aranda PS, LaJoie DM, Jorcyk CL (2012) Bleach gel: a simple agarose gel for analyzing RNA quality. Electrophoresis 33:366–369

Benson DA, Cavanaugh M, Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW (2013) GenBank. Nucleic Acids Res 41:D36–D42

Bergthorsson U, Richardson AO, Young GJ, Goertzen LR, Palmer JD (2004) Massive horizontal transfer of mitochondrial genes from diverse land plant donors to the basal angiosperm *Amborella*. Proc Natl Acad Sci USA 101:17747–17752

Chen K, Otten L (2017) Natural Agrobacterium transformants: recent results and some theoretical considerations. Front Plant Sci 8:1600

Coordinators NR (2013) Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 41:D8–D20

Fernandez-Mazuecos M, Blanco-Pastor JL, Vargas P (2013) A phylogeny of Toadflaxes (*Linaria* Mill.) based on nuclear internal transcribed spacer sequences: systematic and evolutionary consequences. Int J Plant Sci 174:234–249

Fürner IJ, Huffman GA, Amasino RM, Garfinkel DJ, Gordon MP, Nester EW (1986) An *Agrobacterium* transformation in the evolution of the genus Nicotiana. Nature 329:422–427

Goldberg RB, de Paiva G, Yadegari R (1994) Plant embryogenesis: zygote to seed. Science 266:605

Hegelund JN, Liang C, Lauridsen UB, Kemp O, Lutken H, Muller R (2018) Increasing genetic variability in oilseed rape (*Brassica napus*)—genotypes and phenotypes of oilseed rape transformed by wild type *Agrobacterium rhizogenes*. Plant Sci 271:20–26

Hildebrand EM (1940) Cane gall of brambles caused by *Phytomonas rubi* n. sp. J Agric Res 61:685–696

Hong SB, Hwang I, Dessaux Y, Guyon P, Kim KS, Farrand SK (1997) A T-DNA gene required for agropine biosynthesis by transformed plants is functionally and evolutionarily related to a Ti plasmid gene required for catabolism of agropine by Agrobacterium strains. J Bacteriol 179:4831–4840

Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol 35:1547–1549

Kyndt T, Quispe D, Zhai H, Jarret R, Ghislain M, Liu Q, Gheysen G, Kreuze JF (2015) 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 112:5844–5849

Madeira F, Park YM, Lee J, Buso N, Gur T, Madhusoodanan N, Basutkar P, Tivey ARN, Potter SC, Finn RD, Lopez R (2019) The EMBL-EBI search and sequence analysis tools APIs in 2019. Nucleic Acids Res 47:W636–W641

Matveeva TV, Bogomaz DI, Pavlova OA, Nester EW, Lutova LA (2012) Horizontal gene transfer from genus *Agrobacterium* to the plant *Linaria* in nature. Mol Plant-Microbe Interact 25:1542–1551

Matveeva TV, Bogomaz OD, Golovanova LA, Li YS, Dimitrov D (2018) Homologs of the rolC gene of naturally transgenic toadflaxes *Linaria vulgaris* and *Linaria creticola* are expressed in vitro. Vavilovskij Žurnal Genetiki i Selekcii 22:273–278

Matveeva TV, Otten L (2019) Widespread occurrence of natural genetic transformation of plants by *Agrobacterium*. Plant Mol Biol 101:415–437

Mordhorst AP, Toonen MAJ, Vries SCd (1997) Plant embryogenesis. Crit Rev Plant Sci 16:535–576

Ophel K, Kerr A (1990) Agrobacterium vitis-new species for strains of *Agrobacterium* biovar 3 from grapevine. Int J Syst Bacteriol 40:236–241

Pavlova OA, Matveeva TV, Lutova LA (2014) Genome of Linaria dalmatica contains *Agrobacterium rhizogenes rolC* gene homolog. Russ J Genet Appl Res 4:461–465

Richardson AO, Palmer JD (2006) Horizontal gene transfer in plants. J Exp Bot 58:1–9

Riker AJ (1930) Studies on infectious hairy root of nursery apple trees. J Agric Res 41:507–540

Surina B, Pfanzelt S, Einzmann HJR, Albach DC (2014) Bridging the Alps and the Middle East: evolution, phylogeny and systematics of the genus Wulfenia (Plantaginaceae). Taxon 63:843–858

Syvanen M (1994) Horizontal gene transfer: evidence and possible consequences. Annu Rev Genet 28:237–261

Tamura K, Nei M (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10:512–526

Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, Rozen SG (2012) Primer3—new capabilities and interfaces. Nucleic Acids Res 40:e115–e115

Vargas P, Rossello JA, Oyama R, Güemes J (2004) Molecular evidence for naturalness of genera in the tribe Antirrhineae (Scrophulariaceae) and three independent evolutionary lineages from the New World and the Old. Plant Syst Evol 249:151–172

Vigalondo B, Fernández-Mazuecos M, Vargas P, Sáez L (2015) Unmasking cryptic species: morphometric and phylogenetic analyses of the Ibero-North African Linaria incarnata complex. Bot J Linn Soc 177:395–417

Visscher AM, Paul A-L, Kirst M, Guy CL, Schuerger AC, Ferl RJ (2010) Growth performance and root transcriptome remodeling of Arabidopsis in response to Mars-like levels of magnesium sulfate. PLoS ONE 5:e12348

White FF, Ghidossi G, Gordon MP, Nester EW (1982) Tumor induction by *Agrobacterium rhizogenes* involves the transfer of plasmid DNA to the plant genome. Proc Natl Acad Sci USA 79:3193–3197

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