CrossMark
click for updates

The genome of cultivated sweet potato contains *Agrobacterium* T-DNAs with expressed genes: An example of a naturally transgenic food crop

Tina Kyndt[a,1], Dora Quispe[a,b,1], Hong Zhai[c], Robert Jarret[d], Marc Ghislain[b], Qingchang Liu[c], Godelieve Gheysen[a], and Jan F. Kreuze[b,2]

[a]Department of Molecular Biotechnology, Ghent University, 9000 Ghent, Belgium; [b]International Potato Center, Lima 12, Peru; [c]Beijing Key Laboratory of Crop Genetic Improvement/Laboratory of Crop Heterosis and Utilization, Ministry of Education, China Agricultural University, Beijing, China, 100193; and [d]Plant Genetic Resources Unit, US Department of Agriculture, Agricultural Research Service, Griffin, GA 30223

Edited by Eugene W. Nester, University of Washington, Seattle, WA, and approved March 16, 2015 (received for review October 13, 2014)

Agrobacterium rhizogenes and _Agrobacterium tumefaciens_ are plant pathogenic bacteria capable of transferring DNA fragments [transfer DNA (T-DNA)] bearing functional genes into the host plant genome. This naturally occurring mechanism has been adapted by plant biotechnologists to develop genetically modified crops that today are grown on more than 10% of the world's arable land, although their use can result in considerable controversy. While assembling small interfering RNAs, or siRNAs, of sweet potato plants for metagenomic analysis, sequences homologous to T-DNA sequences from _Agrobacterium_ spp. were discovered. Simple and quantitative PCR, Southern blotting, genome walking, and bacterial artificial chromosome library screening and sequencing unambiguously demonstrated that two different T-DNA regions (_Ib_T-DNA1 and _Ib_T-DNA2) are present in the cultivated sweet potato (_Ipomoea batatas_ [L.] Lam.) genome and that these foreign genes are expressed at detectable levels in different tissues of the sweet potato plant. _Ib_T-DNA1 was found to contain four open reading frames (ORFs) homologous to the tryptophan-2-monooxygenase (_iaaM_), indole-3-acetamide hydrolase (_iaaH_), C-protein (_C-prot_), and agrocinopine synthase (_Acs_) genes of _Agrobacterium_ spp. _Ib_T-DNA1 was detected in all 291 cultigens examined, but not in close wild relatives. _Ib_T-DNA2 contained at least five ORFs with significant homology to the _ORF14_, _ORF17n_, rooting locus (_Rol_)_B_/_RolC_, _ORF13_, and _ORF18/ORF17n_ genes of _A. rhizogenes_. _Ib_T-DNA2 was detected in 45 of 217 genotypes that included both cultivated and wild species. Our finding, that sweet potato is naturally transgenic while being a widely and traditionally consumed food crop, could affect the current consumer distrust of the safety of transgenic food crops.

horizontal gene transfer | *Agrobacterium* spp. | food safety | sweet potato | transgenic crops

Horizontal gene transfer (HGT) has long been recognized as a natural phenomenon, especially between bacteria, but it is also being increasingly detected in eukaryotic genomes (1). Many instances of HGT include the transfer of genes from various donors in bdelloid rotifers (2) or from the intracellular bacterium *Wolbachia* into various insect and nematode genomes (3, 4). Some of the transferred genes are not functional in the recipient organism, but others are transcribed, indicating that this phenomenon represents an operational mechanism for the acquisition of new genes. Several horizontally transferred genes have been shown to be correlated with the occurrence of a specific phenotype. Examples include the transfer of carotenoid biosynthetic genes from fungi to aphids that results in the red or green coloration of the aphids (5), or the transfer of genes from hornworts to ferns that results in a more efficient photoreceptor (6). HGT from microbes to plants is well documented. Perhaps the most familiar example is the transfer of the transfer DNA (T-DNA) from *Agrobacterium* spp. This HGT often results in the occurrence of crown galls. The mechanism of this transfer has been extensively studied and is well understood (7, 8).

Crown gall is a disease that afflicts orchards and vineyards in particular. It has long been known to be caused by a bacterial agent (9). In the late 1970s, it was shown that the disease resulted from the transfer of a part of the tumor-inducing (Ti) plasmid, the T-DNA, from *Agrobacterium tumefaciens* into the host plant genome (10). The transfer of the T-DNA from the root-inducing (Ri) plasmid in a related bacterium, *Agrobacterium rhizogenes*, induces abundant root proliferation (hairy roots) at the infection site (11). Once integrated, the genes of the T-DNA are expressed and are responsible for tumor (crown gall) or hairy root formation, as well as the production of opines, in the infected plant tissue. The types of opines synthesized have been used to classify Ti and Ri plasmids into octopine, nopaline, and agropine-type plasmids (12–14).

Agrobacterium rhizogenes agropine strains contain two physically separated T-DNA regions (the TR-DNA and the TL-DNA)

Significance

We communicate the rather remarkable observation that among 291 tested accessions of cultivated sweet potato, all contain one or more transfer DNA (T-DNA) sequences. These sequences, which are shown to be expressed in a cultivated sweet potato clone ("Huachano") that was analyzed in detail, suggest that an *Agrobacterium* infection occurred in evolutionary times. One of the T-DNAs is apparently present in all cultivated sweet potato clones, but not in the crop's closely related wild relatives, suggesting the T-DNA provided a trait or traits that were selected for during domestication. This finding draws attention to the importance of plant–microbe interactions, and given that this crop has been eaten for millennia, it may change the paradigm governing the "unnatural" status of transgenic crops.

Author contributions: J.F.K. designed research; Q.L. and G.G. supervised research; T.K., D.Q., H.Z., R.J., Q.L., and J.F.K. performed research; T.K., D.Q., H.Z., R.J., M.G., G.G., and J.F.K. analyzed data; and T.K., D.Q., R.J., M.G., G.G., and J.F.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The data reported in this paper have been deposited in the Gene Expression Omnibus (GEO) database, www.ncbi.nlm.nih.gov/geo [accession nos. KM052616 (IbT-DNA1 "Huachano"), KM052617 (IbT-DNA2 "Huachano"), KM113766 (IbT-DNA1 BAC clone "Xu781"), KM658948 (ORF13 CIP_420065), KM658949 (ORF13 CIP_440031), KM658950 (ORF13 CIP_440166), KM658951 (ORF13 PI_561255), KM658952 (ORF13 CIP_403552), KM658953 (ORF13 CIP107665.9), KM658954 (iaaM CIP_440132), KM658955 (iaaM CIP_440166), KM658956 (iaaM CIP_440274), KM658957 (iaaM: CIP_440398), KM658958 (iaaM CIP_400450), KM658959 (iaaM CIP_441724), KM658960 (iaaM CIP_440116), KM658961 (iaaM CIP_440146), and KM658962 (iaaM CIP_440031)].

[1]T.K. and D.Q. contributed equally to this work.

[2]To whom correspondence should be addressed. Email: j.kreuze@cgiar.org.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1419685112/-/DCSupplemental.

Downloaded by guest on October 21, 2021

on their Ri plasmid. These are independently integrated into the host genome (15). The TR-DNA generally encodes the proteins required for opine synthesis (16) in addition to the auxin biosynthesis genes tryptophan-2-monooxygenase (*iaaM*) and indole-3-acetamide hydrolase (*iaaH*) (15) and the less-studied open reading frame (ORF) encoding for C-protein (17). The TL-DNA contains ORFs encoding the rooting locus (Rol) genes *RolA* (*ORF10*), *RolB* (*ORF11*), *RolC* (*ORF12*), and *RolD* (*ORF15*), which are involved in auxin/cytokinin balance, in addition to other genes such as *ORF8*, *ORF13*, *ORF14*, and *ORF18*, the functions of which are not completely understood (18, 19). The *RolB*, *RolC*, 5′ part of *ORF8*, 5′ portion of *iaaM*, and other T-DNA genes belong to a family of highly diverged genes without detectable DNA similarity but with statistically significant similarity at the protein level (20).

Soon after the discovery of *Agrobacterium*-mediated gene transfer, the fixation of *A. rhizogenes* T-DNAs in the *Nicotiana* genome was reported (21, 22). These T-DNA sequences, originating from different *A. rhizogenes* strains, are present in *Nicotiana* spp., are transmitted to the progeny, and are expressed at low levels but are not associated with the hairy root or tumor-like symptoms that are usually induced by *Agrobacterium* infections (23–25). The *Rol* genes in *Nicotiana glauca* appear to play a role in various developmental processes and in the expression of morphologic characteristics that have contributed to the evolution of the genus (23, 26). A PCR and hybridization-based search for naturally transgenic species in the family *Solanaceae* was conducted without success (23). However, a subsequent, more extensive search resulted in the identification of *A. rhizogenes*-homologous T-DNA sequences in *Linaria vulgaris* (27, 28). Neither *Nicotiana* nor *Linaria* are food crops, and such findings have so far not been associated with domesticated edible crops. In the course of a high-throughput sequence analysis of small interfering RNAs, or siRNAs, of sweet potato plants (29), siRNAs homologous to T-DNA-like sequences from *Agrobacterium* spp. were discovered in the cultivated sweet potato (*Ipomoea batatas* [L.] Lam.) landrace "Huachano." This observation prompted us to further investigate the presence of T-DNAs in the genome of this crop plant and its related wild species.

Sweet potato is one of the oldest domesticated crops in the Americas, with archeological remains from caves in the Chilca Canyon in Peru dating back to 8,000–10,000 y ago (30). The crop and its closest wild relatives belong to the genus *Ipomoea* (family *Convolvulaceae*) and are confined to section *Eriospermun* Hallier f., series *Batatas* (Choisy) D.F. Austin. This group contains 13 species and two naturally occurring hybrids (31). All of them, with the exception of *Ipomoea littoralis,* are endemic to the Americas (32). The cultivated sweet potato is hexaploid (2n = 6x = 90), although wild tetraploid (2n = 4x = 60) specimens have been collected in Ecuador, Colombia, Guatemala, and Mexico (33). The other species in series *Batatas* are either diploid (2n = 2x = 30) or tetraploid. The identity of the wild ancestors of sweet potato has not been fully elucidated, although morphologic (34, 35), molecular (36–40), and cytogenetic (41) studies all support a close relationship with *Ipomoea trifida* (2x). Therefore, our search for T-DNA in section *Batatas* genomes included a large number of both domesticated and semidomesticated sweet potatoes (hereafter referred to as cultigens which includes landraces, varieties, and feral forms of hexaploid *I. batatas*) and wild sweet potato-related species.

Results

The Sweet Potato Genome Contains *Agrobacterium* T-DNAs. In the process of performing small RNA sequencing of sweet potato landrace "Huachano" (29), a number of contigs were assembled that showed significant similarity to *Agrobacterium* genes: *agrocinopine synthase* (*Acs*), *C-protein* (*C-prot*), *iaaH*, *iaaM*, *RolB*, and *ORF18*. Sequencing of PCR products from these contigs confirmed the presence of these genes in sweet potato DNA. Genome walking techniques were subsequently used to identify two large T-DNA regions from *Agrobacterium* spp. inserted into the sweet potato "Huachano" genome (Fig. 1). Further analysis revealed that the first region, *Ipomoea batatas* T-DNA1 (*Ib*T-DNA1), had at least four successive intact ORFs homologous to the *Acs*, *C-prot*, *iaaH*, and *iaaM* sequences of *Agrobacterium* spp., as well as a truncated *iaaM* in an inverted orientation (Fig. 1*A* and Dataset S1). The second region, *Ib*T-DNA2, contained at least five intact ORFs with significant homology to the *ORF14*, *ORF17n*, *RolB*/*RolC family*, *ORF13*, and *ORF18*/*ORF17n* family of *A. rhizogenes* (Fig. 1*B* and Dataset S1). The flanking sequences of *Ib*T-DNA1 (Fig. 1*A*), obtained by genome walking, revealed at least 80% sequence identity (BlastN) to predicted *F-box* genes from *Nicotiana*, *Morus*, *Solanum*, and *Fragaria* spp. A tBlastX of this region predicts orthologous F-box proteins across numerous other plant families (Dataset S2). This presumed sweet potato sequence was a 100% nucleotide match to several transcript sequences from the sweet potato gene index (https://research.cip.cgiar.org/confluence/display/SPGI/Home; contig 02446) and a sweet potato transcriptome shotgun library at NCBI (JP111314.1), which were predicted to encode F-box proteins (e-value = e-177). These observations suggest the "Huachano" *Ib*T-DNA1 is inserted into an *F-box* gene intron.

Southern blot analyses were performed to confirm the insertion of both T-DNAs into the sweet potato ("Huachano") genome (Fig. 2). An estimated four copies of each T-DNA were present (Fig. 2 *A* and *C*). When comparing these copies with the two T-DNA sequences identified through genome walking (Fig. 1), it is not possible to determine whether the three homologous sequences of each T-DNA are the result of polymorphic *Spe*I sites, and thus allelic, or arose independently as the result of separate integration events or from rearrangements subsequent to a single integration event. Hybridization of total genomic DNA, digested with *Spe*I, with a probe for *Ib*T-DNA1 (*C-prot*, probe 1; Fig. 1*A*) produced several bands (Fig. 2*A*) that also appeared (Fig. 2*B*) when hybridized to the probe corresponding to the flanking plant DNA (*F-box*, probe 2; Fig. 1*A*), confirming their physical linkage. Additional bands hybridizing to probe 2 (Fig. 2*B*) suggest that at least two additional DNA regions homologous to the *F-box* gene, unlinked to *Ib*T-DNA1, are present

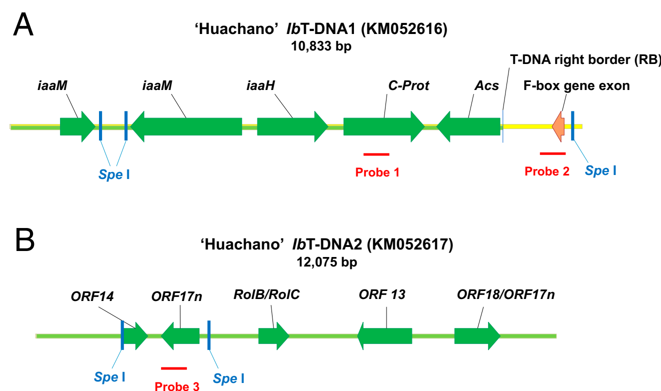

Fig. 1. Organization of *Ib*T-DNA1 and *Ib*T-DNA2 in the genome of sweet potato. (*A*) *Ib*T-DNA1 of landrace "Huachano," including ORFs showing significant homology to *iaaM*, *iaaH*, *C-prot*, and *Acs* and a truncated *iaaM* in inverted orientation. This T-DNA is located in an intron of a gene showing strong homology to plant *F-box* genes. The regions with significant similarity to plant sequences is shown as a yellow line. (*B*) *Ib*T-DNA2 of "Huachano," including ORFs with significant homology to *ORF14*, *ORF17n*, *RolB/RolC*, *ORF13*, and *ORF18/ORF17n*. The locations of the probes used for Southern blots are indicated in red below the corresponding gene regions, and the positions of the *Spe*I restriction sites are shown in blue. National Center for Biotechnology Information (NCBI) GenBank sequence accession numbers are indicated between parentheses for each T-DNA.

PLANT BIOLOGY

Downloaded by guest on October 21, 2021

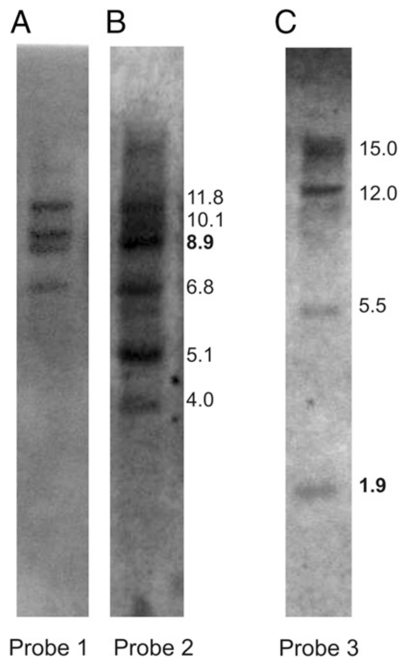

Fig. 2. Southern blot analyses showing the integration of *Ib*T-DNA1 and *Ib*T-DNA2 into the sweet potato genome. Total genomic DNA of landrace "Huachano" was digested with *Spe*I and hybridized with various probes. (*A*) Probe 1 complementary to the ORF coding for C-protein of *Ib*T-DNA1 revealing the presence of multiple (four estimated) insertions into the sweet potato genome. (*B*) Probe 2 complementary to the *F-box* gene in the region flanking *Ib*T-DNA1, revealing the presence of probably six copies, four of which appear to correspond to similar bands in the hybridization with probe 1 for *Ib*T-DNA1. (*C*) Probe 3 complementary to the ORF coding for *ORF17n* of *Ib*T-DNA2. Band sizes as expected from the nucleotide sequence data are indicated in bold.

in the genome and are likely the origin of the identified sweet potato *F-box* transcripts. Hybridization with a probe for *Ib*T-DNA2 (*ORF17n*, probe 3; Fig. 1*B*) again revealed the presence of this locus in the sweet potato genome (Fig. 2*C*). Phylogenetic analyses at the nucleotide level were performed to elucidate the relationship between the ORFs from the *Ib*T-DNAs and the Ri/Ti plasmids of well-characterized *Agrobacterium* strains (Fig. 3). The nucleotide organization and the resulting trees from *iaaM* (*Ib*T-DNA1) and *ORF13* (*Ib*T-DNA2) confirmed all these sequences are homologous.

IbT-DNA1 Interrupts an *Ipomoea* Gene and Is Expressed. The T-DNA insertion into the sweet potato *F-box* gene was corroborated by sequence analysis of a bacterial artificial chromosome (BAC) clone. This clone was identified by screening a previously generated BAC library of sweet potato variety "Xu781," using primers specific to *iaaM* and *C-prot* (Fig. 4*A*). The BAC sequence (79,099 bp) revealed that the complete *Ib*T-DNA1 in "Xu781," encompassing 22,146 bp, was located between two T-DNA border-like sequences and consisted of an inverted repeat of the *Ib*T-DNA1 described in "Huachano" with some indels (Fig. 4*B*). It comprises the *Acs*, *C-prot*, *iaaH*, and *iaaM* genes, including a region containing several short sequence repeats and similarity to Gypsy 2 type long terminal repeat transposons inserted into one of the *iaaM* copies (Fig. 4*B*), as well as a deletion of 310 bp in one of the *C-prot* copies (in contrast, the "Huachano" *Ib*T-DNA1 had a 910-bp deletion in one of the *iaaM* gene copies). It was flanked on both sides by sequences that correspond to putative exons matching transcripts of predicted F-box proteins, confirming that this *Ib*T-DNA1 was inserted into an intron of this gene. The presence of additional predicted genes in the BAC region (Fig. 4*A*) suggests *Ib*T-DNA1 is located in a transcriptionally active region of the chromosome. This conjecture is supported by near-perfect homology to sequences available in published sweet potato transcriptomes, with the exception of the predicted LINE-type retrotransposon from *I. batatas*. Hence, the *Ib*T-DNA1 insertion appears to interrupt an *Ipomoea* gene that was presumably functional before the T-DNA insertion. The presence of two uninterrupted *F-box* genes (as deduced from the Southern blot) might compensate for the loss of the interrupted paralogue.

Expression of the T-DNA ORFs was tested by quantitative RT-PCR (qRT-PCR), using the sweet potato landrace "Huachano." This analysis revealed the presence of low, but detectable, levels of the mRNAs of these genes in leaf, stem, root, shoot apex, and storage root tissues (Fig. 5). This observation confirms these ORFs are expressed in sweet potato.

IbT-DNA1 but Not IbT-DNA2 Is Ubiquitous in Sweet Potato, Whereas Closely Related Species Appear to Have None or Only One T-DNA. To examine the presence of the T-DNA sequences in the cultivated sweet potato gene pool, PCR analyses were performed for three or four of the ORFs of *Ib*T-DNA1 and *ORF13* of *Ib*T-DNA2. The analyses included a total of 291 and 204 cultigens for *Ib*T-DNA1 and *Ib*T-DNA2, respectively, which represented a random sampling of the global hexaploid sweet potato gene pool. These analyzed plant materials were collected in South and Central America, Africa, Asia, and Oceania. Ten tetraploid plants (nine *I. batatas* and one *Ipomoea tabascana*) and three diploid plants (two *I. trifida* and one *Ipomoea triloba*), representing closely related taxa, were also included in this analysis. The results are shown in Dataset S3. *Ib*T-DNA1 was detected in all 291 hexaploid sweet potato cultigens, but not in the wild relatives. In contrast, *ORF13* from *Ib*T-DNA2 was detected in 42 hexaploid cultigens, two of the nine tetraploid genotypes of *I. batatas*, and a single diploid *I. trifida*, among a total of 217 genotypes examined. The tested genotypes originated from different germplasm collections, and the analyses were conducted independently in two different laboratories.

IbT-DNA2 and Root Characteristics. In *Nicotiana*, *Rol*-containing genotypes are rooting-prone, whereas the others are shoot-producing

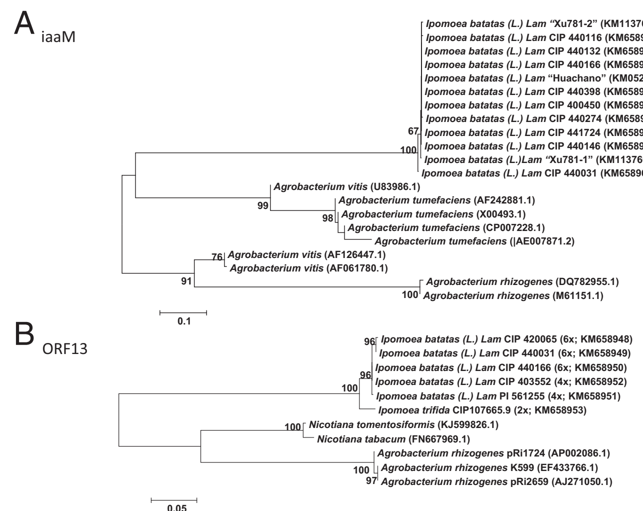

Fig. 3. Phylogenetic trees generated by neighbor joining of (*A*) *iaaM* (399-nt fragment) and (*B*) *ORF13* (722-nt fragment) alignments. Values at the nodes indicate percentage of bootstrap support (of 1,000 bootstrap replicates) and are indicated if greater than 50. International Potato Center germplasm accession numbers are indicated when available for *Ipomoea* spp. whereas strains or plasmid names are indicated for *Agrobacterium* spp. GenBank accession numbers are provided between brackets for all sequences; ploidy levels are also provided between parentheses in *B* for *Ipomoea* spp.

Downloaded by guest on October 21, 2021

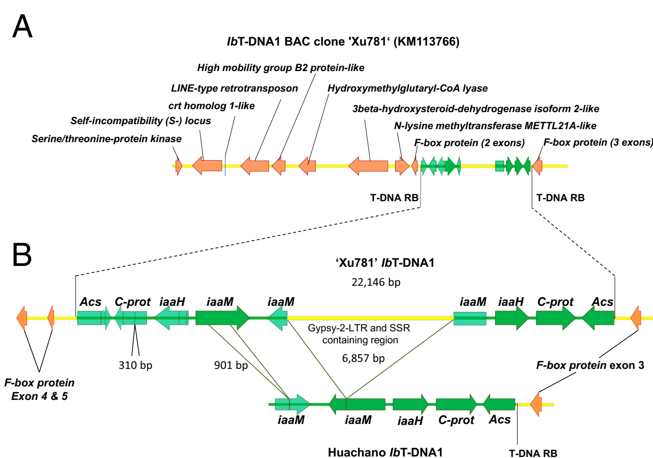

Fig. 4. Genomic structure of *Ib*T-DNA1 region in variety "Xu781" (*A*), and detailed comparison of "Xu781" *Ib*T-DNA1 to "Huachano" *Ib*T-DNA1 (*B*). NCBI GenBank sequence accession number is provided in parentheses in *A*. The genomic region includes predicted plant genes shown in orange, and T-DNA genes in green. The T-DNA region, enlarged in *B*, bears an insertion, containing three segments with homology to Gypsy-2-LTRs and several SSRs [seven in total, with the motifs $TAT_{(n)}$, $GA_{(n)}$, $GA_{(n)}$, $TA_{(n)}$, $TAT_{(n)}$, $TAT_{(n)}$ and $TA_{(n)}$], in one of the *iaaM* ORFs compared with the partial sequence determined for "Huachano" and a deletion in one copy of the *C-prot* gene. *Ib*T-DNA1 of "Huachano" contains a deletion in one of the *iaaM* copies. Semitransparent ORFs indicate they are (partially) interrupted by indels or stop-codons (indicated by lines inside the ORF) compared with the full-length and uninterrupted ORFs.

(23). Because qRT-PCR analysis revealed expression of *RolB*/*RolC* and *ORF13* in "Huachano" tissues, this prompted us to look at root parameters in sweet potato in relation to the presence or absence of *Ib*T-DNA2 (as all clones tested positive for the presence of *Ib*T-DNA1). Using the primers listed in Dataset S4, we observed *Ib*T-DNA2 polymorphisms for two primer pairs corresponding to the genes *RolB*/*RolC* and *ORF13* in cultivars "Beauregard" (negative) and "Tanzania" (positive). Hence, we tested by segregation analysis for the presence of *Ib*T-DNA2 (using the *RolB*/*RolC* and *ORF13* primers) in an existing biparental segregating population from these parents for possible association with root parameters (total root yield, dry matter content, and harvest index). Only four of 76 progeny were negative for both markers, and nine additional genotypes were negative only for *RolB*/*RolC*, indicating the presence of variants of *Ib*T-DNA2. No association between *Ib*T-DNA2 marker alleles and root characteristics was found with the exception of root yield at one location (Wilcoxon rank sum test with continuity correction, $P = 0.04211$; Dataset S5). This preliminary genetic analysis does not prove a role for the *Ib*T-DNA in root development, and additional segregating populations will have to be generated and analyzed in dedicated experiments to detect potential associations.

Discussion

The acquisition of new genes that confer a selective advantage is an important factor in genome evolution. Significant parts of prokaryotic and eukaryotic genomes originated from the exchange of genetic material among related or unrelated species through HGT. This is considered to be a significant source of molecular variability and a driver of evolution (42–45). Our data provide evidence of an ancient HGT between an *Agrobacterium* spp. and an ancestor (or ancestral form) of the sweet potato. Our data indicate that the presence of *Ib*T-DNA1 is a general feature of the domesticated sweet potato gene pool. The presence of *Ib*T-DNA1 in all hexaploid cultigens examined, and the lack of segregation in the progeny of the analyzed cross, suggests this T-DNA fragment is fixed in the cultivated sweet potato genome, in contrast to its close wild relatives. The DNA polymorphisms between "Huachano" and "Xu781" *Ib*T-DNA1 (Fig. 4) attest to an ancient transfer of the T-DNA into an *Ipomoea* species. It is therefore conceivable that one or more of the transferred genes contributed to the expression of a trait that was subsequently selected for during domestication. Our preliminary analysis of root parameters did not reveal a consistent role for the T-DNA genes in storage root development, which was a distinctive feature of domesticated forms in the examined cross, although the significant association of the presence of ORF13 with higher root yield at one location (Dataset S5) merits further follow-up. Additional *Ib*T-DNA nucleotide sequence studies, *Ib*T-DNA segregation studies using additional populations, and further examination of alterations of gene expression are required to more fully elucidate a functional role of the T-DNA genes in storage root development, if one exists.

Remarkably, the majority of ORFs detected on *Ib*T-DNA1 and *Ib*T-DNA2 are intact, and expression of these genes was detected in different organs of "Huachano," indicating their potential functionality. Future expression studies should be conducted using several biological replicates and genotypes with contrasting root characteristics to infer possible roles of the *Ib*T-DNAs.

The gene sequences identified in *Ib*T-DNA1 and *Ib*T-DNA2 indicate that the transforming *Agrobacterium* most likely was *A. rhizogenes* (46), an ancestral form of *A. rhizogenes* or a species (perhaps extinct) closely related to *A. rhizogenes*. In this scenario, *Ib*T-DNA1 corresponds to TR-DNA (typically containing the auxin biosynthesis genes *iaaM* and *iaaH*), and *Ib*T-DNA2 corresponds to TL-DNA (harboring the *Rol* genes). The gene organization (Fig. 4) and DNA sequences (Dataset S1) of the T-DNAs are similar to, but distinctly different from, the ORFs of the Ri and Ti-plasmids in well-characterized *Agrobacterium* strains. The identified *RolB*/*RolC* region represents a new member of the *RolB* family (Fig. S1). This indicates that, unlike the T-DNA found in *Nicotiana* spp. (26), the *Agrobacterium* strain (or species) that transferred its T-DNA into the sweet potato genome is not one of the commonly studied strains.

Agrobacterium-mediated transformation has been the method of choice for the development of genetically modified crops. Despite their cultivation on more than 170 million ha, the growth and consumption of transgenic crops still faces societal opposition. This has impeded their use in efforts to contribute to a more sustainable agricultural future (47–52). Our data reveal that T-DNA integration, the interruption of an *F-box* gene, and the subsequent fixation of foreign T-DNA into the sweet potato genome occurred during the evolution and domestication of this crop, which is one of the world's most consumed foods (53). This

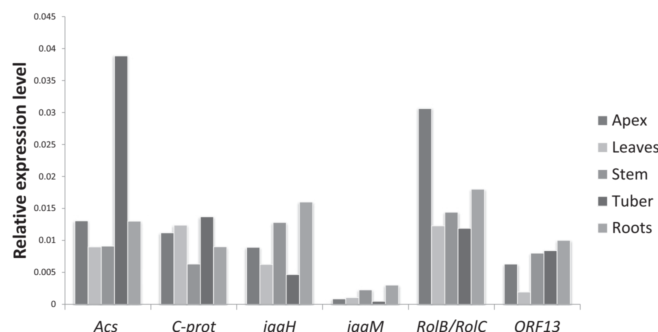

Fig. 5. Relative expression of four ORFs on *Ib*T-DNA1 and two ORFs located on *Ib*T-DNA2. The figure shows the relative presence of mRNA in different tissues of sweet potato, based on qRT-PCR. The expression levels are shown relative to the expression level of the *Cox* housekeeping gene in roots. Data were normalized using data from two reference genes (*PLD* and *RPL*).

PLANT BIOLOGY

Downloaded by guest on October 21, 2021

finding could influence the public's current perception that transgenic crops are "unnatural."

Materials and Methods

Plant Material. In total, 304 plant samples were included in this study (Dataset S3): 291 samples of cultivated *I. batatas* [L.] Lam. (hexaploid) originating from South and Central America, Africa, Asia, and Oceania; nine samples belonging to wild, uncultivated *I. batatas* [L.] Lam. (tetraploid) from South and Central America; and four samples belonging to one of the related species [*I. tabascana* (tetraploid), *I. trifida* (diploid), and *I. triloba* (diploid)]. The plant materials were obtained from the germplasm collections of the International Potato Center (Lima, Peru) and the National Genetic Resources Program (US Department of Agriculture).

DNA Extraction. DNA extraction from leaf tissues of 303 samples was performed using the Cetyl trimethylammonium bromide (CTAB) method (54). DNA quality and quantity were measured using a Nanodrop spectrophotometer (Thermo Fisher Scientific Inc.) and agarose gel electrophoresis.

Screening for *Ib*T-DNAs in *Ipomoea* spp. Detection of *Ib*T-DNA1 and *Ib*T-DNA2 genes in *Ipomoea* samples was carried out by conventional PCR or high-resolution DNA melting (55), using the primers listed in Dataset S4. Part of the *Ipomoea*-specific malate dehydrogenase gene was amplified on each DNA sample as a positive PCR control. PCR reactions were accomplished in 25-μL volumes containing 1× PCR buffer (Invitrogen); 0.4 mM each dGTP, dATP, dTTP, and dCTP; 0.3 μM forward and reverse primer; 1 U Taq DNA polymerase (Invitrogen); and 100 ng genomic DNA. The PCR conditions were 94 °C for 5 min, followed by 35 cycles of 95 °C for 30 s, 50°–60 °C for 30 s, and 72 °C for 2 min, and then a final extension at 72 °C for 10 min. PCR products were separated on 1% agarose gels for visual detection of presence or absence.

Genome Walking on "Huachano." The joining and extension of partial sequences identified through small RNA sequencing and PCR were carried out using the Genome Walker Universal kit (Clontech) according to the manufacturer's instructions and some minor modifications to the PCR conditions as follows: First-PCR: 94 °C for 2 min, followed by seven cycles of 94 °C for 25 s and 72 °C for 3 min, 32 cycles of 94 °C for 25s and 67 °C for 3 min, and a final cycle of 67 °C for 7 min. Nested-PCR: 94 °C for 1 min, followed by 5 cycles of 94 °C for 25 s and 72 °C for 3 min, 20 cycles of 94 °C for 25 s and 67 °C for 3 min, and a final cycle of 67 °C for 7 min. PCR products were recovered using the Wizard SV gel extraction kit (Promega) according to the manufacturer's recommendation. The eluted DNA was ligated into plasmid vector pCR 2.1 (Invitrogen), according to the manufacturer's instructions, and cloned in *Escherichia coli* strain DH5α. PCR products were sequenced by Macrogen, using the Sanger method.

BAC Library of Sweet Potato Variety "Xu781." A BAC library of 15,360 bacterial colonies was constructed from total genomic DNA partially restricted with *Hind*III from the sweet potato variety "Xu781," following standard procedures. PCR screening was done using the primers for amplification of the *iaaM* and *C-prot* gene (Dataset S4). After identification of a positive BAC clone, its insert was sequenced at the Beijing Genomics Institute, using the Sanger method in a shotgun approach.

Sequence Analysis, Blast Search, Annotation, and Accession Number. Sequences were assembled using the software Seqman II (DNAstar, Inc.). Coding regions were predicted using the program FGENESH2.6 (56) and annotated on the basis of the top hits when performing BlastN and tBlastX searches with the nonredundant nucleotide database of NCBI. Alignment and phylogenetic analysis were performed using the software MEGA 5 (57). All sequences were deposited in the GenBank database under the following accession numbers: KM052616 (IbT-DNA1 "Huachano"), KM052617 (IbT-DNA2 "Huachano"), KM113766 (IbT-DNA1 BAC clone "Xu781"), KM658948 (ORF13 CIP_420065), KM658949 (ORF13 CIP_440031), KM658950 (ORF13 CIP_440166), KM658951 (ORF13 PI_561255), KM658952 (ORF13 CIP_403552), KM658953 (ORF13 CIP107665.9), KM658954 (iaaM CIP_440132), KM658955 (iaaM CIP_440166), KM658956 (iaaM CIP_440274), KM658957 (iaaM: CIP_440398), KM658958 (iaaM CIP_400450), KM658959 (iaaM CIP_441724), KM658960 (iaaM CIP_440116), KM658961 (iaaM CIP_440146), and KM658962 (iaaM CIP_440031).

Southern Blot Hybridization. A total of 30 μg genomic DNA was digested with *Spe*I, separated on a 0.8% agarose gel under 25 eV for 18 h, and transferred to a Hybond-N+ nylon membrane (Amersham Pharmacia Biotech) with transfer buffer (20× SSC).

Primers (Dataset S4) were designed to amplify three different DNA probes (probe 1, *C-prot*; probe 2, the *F-box* gene sequence; and probe 3, *ORF17n*). Probe labeling was performed using the PCR DIG Probe Synthesis Kit (Roche). Prehybridization and hybridization steps were carried out using the buffer DIG Easy Hyb (Roche), according to the manufacturer's instructions. After hybridization, membranes were washed twice (5 min) at low stringency (2× SSC, 0.1% SDS) at room temperature and two additional times (15 min) at high stringency (0.1× SSC, 0.1% SDS) at 65 °C. The images were captured by chemiluminescence on photosensitive film (Fujifilm Life Science).

RNA Extraction, cDNA Synthesis, and qRT-PCR. RNA from leaf, stem, root, shoot apex, and storage root tissues of sweet potato plants (landrace "Huachano") was extracted using the RNeasy Plant Mini Kit (Qiagen GmbH). The RNA concentration and purity were measured using the Nanodrop spectrophotometer (Thermo Fisher Scientific Inc.). To remove contaminating DNA, the extract was treated with DNase I (Fermentas). First-strand cDNA synthesis was generated using SuperScript II Reverse Transcriptase (Invitrogen). All qRT-PCRs were performed in triplicate by using the SensiMix SYBR No Rox Kit (Bioline) in the Rotor-Gene 3,000 (Qiagen), using Rotor Discs (72, Qiagen); results were generated by the Rotor-Gene 6 software. PCRs were performed under the following conditions: 10 min at 95 °C and 45 cycles of 25 s at 95 °C, 60 s at 58 °C, and 20 s at 72 °C. Data were analyzed using the REST 2008 software (Qiagen). Normalization of the expression level of the target genes was done using two reference genes: phospholipase D1 (*PLD*) and ribosomal protein L (*RPL*) (see Dataset S4 for all used primer pairs). Expression levels are shown in comparison with the expression level of the stably expressed housekeeping gene Cytochrome *c* oxidase (*Cox)*. As negative controls, "no reverse transcriptase" and "no template" samples (*I. trifida* accession CIP 460112) were tested and confirmed to be negative.

ACKNOWLEDGMENTS. We thank Lien De Smet, Isabelle Tilmant, Milton Untiveros, Bruno Lertora, Roland Robles, and Clauvis Taning for technical assistance; Martine De Cock and Ivonne Valdizan for help with layout and formatting; Wolfgang Gruneberg for sharing phenotypic and yield data for the BxT (Beauregard by Tanzania) population; Felipe de Mendiburo and Hannele Lindqvist Kreuze for advice on statistical analysis; Genoveva Rossel and David Tay for sharing ploidy information and DNA from germplasm accessions; Zengzhi Si for helping to screen the BAC library; and Jason McKinney (Fluidigm), Cameron Gundry, and Matt Poulson (BioFire Diagnostics) for high resolution melting primer design using sequence data provided by R.J. and screening of the sweet potato collection of the US Department of Agriculture. This work was supported by grants from the Special Research Fund of Ghent University (BOF 01SB0710 and 01W02112) and the Consultative Group on International Agricultural Research (CGIAR) consortium research program on roots, tubers, and bananas. T.K. is a postdoctoral fellow of the Research Foundation-Flanders.

1. Dunning Hotopp JC (2011) Horizontal gene transfer between bacteria and animals. *Trends Genet* 27(4):157–163.
2. Gladyshev EA, Meselson M, Arkhipova IR (2008) Massive horizontal gene transfer in bdelloid rotifers. *Science* 320(5880):1210–1213.
3. Dunning Hotopp JC, et al. (2007) Widespread lateral gene transfer from intracellular bacteria to multicellular eukaryotes. *Science* 317(5845):1753–1756.
4. Nikoh N, et al. (2008) Wolbachia genome integrated in an insect chromosome: Evolution and fate of laterally transferred endosymbiont genes. *Genome Res* 18(2):272–280.
5. Moran NA, Jarvik T (2010) Lateral transfer of genes from fungi underlies carotenoid production in aphids. *Science* 328(5978):624–627.
6. Li F-W, et al. (2014) Horizontal transfer of an adaptive chimeric photoreceptor from bryophytes to ferns. *Proc Natl Acad Sci USA* 111(18):6672–6677.
7. Zupan J, Muth TR, Draper O, Zambryski P (2000) The transfer of DNA from *Agrobacterium tumefaciens* into plants: A feast of fundamental insights. *Plant J* 23(1):11–28.
8. Gelvin SB (2003) *Agrobacterium*-mediated plant transformation: The biology behind the "gene-jockeying" tool. *Microbiol Mol Biol Rev* 67(1):16–37.
9. Smith EF, Townsend CO (1907) A plant tumor of bacterial origin. *Science* 25(643): 671–673.
10. Chilton M-D, et al. (1977) Stable incorporation of plasmid DNA into higher plant cells: The molecular basis of crown gall tumorigenesis. *Cell* 11(2):263–271.
11. Chilton M-D, et al. (1982) *Agrobacterium rhizogenes* inserts T-DNA into the genomes of the host plant root cells. *Nature* 295(5848):432–434.
12. Chilton M-D, Drummond MH, Merlo DJ, Sciaky D (1978) Highly conserved DNA of Ti plasmids overlaps T-DNA maintained in plant tumours. *Nature* 275(5676):147–149.
13. Depicker A, Van Montagu M, Schell J (1978) Homologous DNA sequences in different Ti-plasmids are essential for oncogenicity. *Nature* 275(5676):150–153.
14. Engler G, et al. (1981) Physical mapping of DNA base sequence homologies between an octopine and a nopaline Ti plasmid of *Agrobacterium tumefaciens*. *J Mol Biol* 152(2):183–208.

Downloaded by guest on October 21, 2021

15. Vilaine F, Casse-Delbart F (1987) Independent induction of transformed roots by the TL and TR regions of the Ri plasmid of agropine type *Agrobacterium rhizogenes*. *Mol Gen Genet* 206(1):17–23.
16. Salomon F, et al. (1984) Genetic identification of functions of TR-DNA transcripts in octopine crown galls. *EMBO J* 3(1):141–146.
17. Otten L, et al. (1999) Sequence and functional analysis of the left-hand part of the T-region from the nopaline-type Ti plasmid, pTiC58. *Plant Mol Biol* 41(6):765–776.
18. Slightom JL, Durand-Tardif M, Jouanin L, Tepfer D (1986) Nucleotide sequence analysis of TL-DNA of *Agrobacterium rhizogenes* agropine type plasmid. Identification of open reading frames. *J Biol Chem* 261(1):108–121.
19. Capone I, et al. (1989) Induction and growth properties of carrot roots with different complements of *Agrobacterium rhizogenes* T-DNA. *Plant Mol Biol* 13(1):43–52.
20. Mohajjel-Shoja H, Clément B, Perot J, Alioua M, Otten L (2011) Biological activity of the *Agrobacterium rhizogenes*-derived *trolC* gene of *Nicotiana tabacum* and its functional relation to other *plast* genes. *Mol Plant Microbe Interact* 24(1):44–53.
21. White FD, Garfinkel J, Huffman GA, Gordon MP, Nester EW (1983) Sequences homologous to *Agrobacterium rhizogenes* T-DNA in the genomes of uninfected plants. *Nature* 301:348–350.
22. Furner IJ, et al. (1986) An *Agrobacterium* transformation in the evolution of the genus *Nicotiana*. *Nature* 319(6052):422–427.
23. Intrieri MC, Buiatti M (2001) The horizontal transfer of *Agrobacterium rhizogenes* genes and the evolution of the genus *Nicotiana*. *Mol Phylogenet Evol* 20(1):100–110.
24. Chen K, Dorlhac de Borne F, Szegedi E, Otten L (2014) Deep sequencing of the ancestral tobacco species *Nicotiana tomentosiformis* reveals multiple T-DNA inserts and a complex evolutionary history of natural transformation in the genus *Nicotiana*. *Plant J* 80(4):669–682.
25. Suzuki K, Yamashita I, Tanaka N (2002) Tobacco plants were transformed by *Agrobacterium rhizogenes* infection during their evolution. *Plant J* 32(5):775–787.
26. Aoki S (2004) Resurrection of an ancestral gene: Functional and evolutionary analyses of the Ngrol genes transferred from *Agrobacterium* to *Nicotiana*. *J Plant Res* 117(4): 329–337.
27. 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(12):1542–1551.
28. Matveeva TV, Lutova LA (2014) Horizontal gene transfer from *Agrobacterium* to plants. *Front Plant Sci* 5:326.
29. Kreuze JF, et al. (2009) Complete viral genome sequence and discovery of novel viruses by deep sequencing of small RNAs: A generic method for diagnosis, discovery and sequencing of viruses. *Virology* 388(1):1–7.
30. Woolfe JA (1992) *Sweetpotato: An untapped food resource* (Cambridge University Press, Cambridge, UK), pp 118–187.
31. Nimmakayala P, Vajja G, Reddy UK (2011) *Ipomoea. Wild Crop Relatives: Genomic and Breeding Resources, Industrial Crops*, ed Kole C (Springer, Berlin), pp 123–132.
32. Huamán Z (1992) *Systematic botany and morphology of the Sweetpotato plant. Technical Information Bulletin 25* (International Potato Center, Lima, Perú), pp 2–6.
33. Bohac J, Austin D, Jones A (1993) Discovery of wild tetraploid sweetpotatoes. *Econ Bot* 47(2):193–201.
34. Shiotani I (1987) Genomic structure and the gene flow in sweet potato and related species. *Exploration, Maintenance and Utilization of Sweet Potato Genetic Resources. Report of the First Sweet Potato Planning Conference* (International Potato Center, Lima, Peru), pp 61–73.
35. Austin DF (1988) The taxonomy, evolution and genetic diversity of sweet potatoes and related wild species. *Exploration, Maintenance and Utilization of Sweetpotato Genetic Resources. Report of the First Sweet Potato Planning Conference* (International Potato Center, Lima, Peru), pp 27–59.
36. Jarret RL, Gawel N, Whittemore A (1992) Phylogenetic relationships of the sweetpotato [*Ipomoea batatas* (L.) Lam.]. *J Am Soc Hortic Sci* 117(4):633–637.
37. Jarret RL, Austin DF (1994) Genetic diversity and systematic relationships in sweetpotatoes [(*Ipomoea batatas* (L.) Lam.] and related species as revealed by RAPD analyses. *Genet Resour Crop Evol* 41(3):165–173.
38. Buteler MI, Jarret RL, La Bonte DR (1999) Sequence characterization of microsatellites in diploid and polyploid *Ipomoea*. *Theor Appl Genet* 99(1-2):123–132.
39. Rajapakse S, et al. (2004) Phylogenetic relationships of the sweetpotato in *Ipomoea* series Batatas (*Convolvulaceae*) based on nuclear beta-amylase gene sequences. *Mol Phylogenet Evol* 30(3):623–632.
40. Roullier C, et al. (2013) Disentangling the origins of cultivated sweet potato (*Ipomoea batatas* (L.) Lam.). *PLoS ONE* 8(5):e62707.
41. Srisuwan S, Sihachakr D, Siljak-Yakovlev S (2006) The origin and evolution of sweet potato (*Ipomoea batatas* Lam.) and its wild relatives through the cytogenetic approaches. *Plant Sci* 171(3):424–433.
42. Maiden MCJ (1998) Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. *Clin Infect Dis* 27(Suppl 1):S12–S20.
43. Ochman H, Lawrence JG, Groisman EA (2000) Lateral gene transfer and the nature of bacterial innovation. *Nature* 405(6784):299–304.
44. Gogarten JP, Townsend JP (2005) Horizontal gene transfer, genome innovation and evolution. *Nat Rev Microbiol* 3(9):679–687.
45. Bock R (2010) The give-and-take of DNA: Horizontal gene transfer in plants. *Trends Plant Sci* 15(1):11–22.
46. White FF, Taylor BH, Huffman GA, Gordon MP, Nester EW (1985) Molecular and genetic analysis of the transferred DNA regions of the root-inducing plasmid of *Agrobacterium rhizogenes*. *J Bacteriol* 164(1):33–44.
47. Anonymous (2013) Fields of Gold. *Nature* 497(7447):5–6.
48. Anonymous (2013) Plant biotechnology: Tarnished promise. *Nature* 497(7447):21.
49. Anonymous (2013) GM crops: A story in numbers. *Nature* 497(7447):22–23.
50. Gilbert N (2013) Case studies: A hard look at GM crops. *Nature* 497(7447):24–26.
51. Cressey D (2013) Transgenics: A new breed. *Nature* 497(7447):27–29.
52. Whitty CJM, Jones M, Tollervey A, Wheeler T (2013) Biotechnology: Africa and Asia need a rational debate on GM crops. *Nature* 497(7447):31–33.
53. FAO (2013) FAOSTAT vol. 2014 faostat3.fao.org/browse/rankings/commodities_by_regions/E. Accessed August 2014.
54. Doyle JJ (1990) Isolation of plant DNA from fresh tissue. *Focus* 12:13–15.
55. McKinney JT, et al. (2010) Mutation screening and genotyping in plants by high resolution DNA melting. *Handbook of Plant Mutation Screening*, eds Meksem K, Kahl G (CRC Press, London), pp 149–163.
56. Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. *Genome Biol* 7(Suppl 1):1–12.
57. Tamura K, et al. (2011) MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* 28(10):2731–2739.

Downloaded by guest on October 21, 2021