NICOTIANA SYLVESTRIS, A MODEL PLANT FOR CELL BIOLOGY: ORGANELLE MOVEMENT AND RETROTRANSPOSON MUTAGENESIS

By

GREGORY N. THYSSEN

A Dissertation submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Plant Biology

written under the direction of

Dr. Pal Maliga

and approved by

_____

_____

_____

_____

New Brunswick, New Jersey

May 2012

ABSTRACT OF THE DISSERTATION

Nicotiana sylvestris, a Model Plant for Cell Biology: Organelle Movement and Retrotransposon Mutagenesis

By

GREGORY N. THYSSEN

Dissertation Director:

Dr. Pal Maliga

Nicotiana sylvestris **is a diploid tobacco plant that is amenable to laboratory manipulation including facile transformation of nuclear and plastid (chloroplast) genomes. In three separate studies, I used this model organism to observe biological processes with evolutionary and biotechnological implications.**

The first addresses the mechanisms of horizontal gene transfer by demonstrating cell-to-cell movement of plastids. We grafted *Nicotiana sylvestris* plants with selectable transgenic plastid genomes to *Nicotiana tabacum* plants with selectable transgenic nuclear markers. Grafting triggers formation of new cell-to-cell contacts, creating an opportunity for organelle movement between the plant cells. I present evidence for cell-to-cell movement of the entire 161-kb plastid genome in these plants, most likely in intact plastids. Acquisition of plastids from neighboring cells provides a mechanism by which cells may be repopulated with functioning organelles.

My second objective was to determine whether exceptional pollen transmission

of plastids is accompanied by paternal mitochondria transmission in *Nicotiana sylvestris*. Plastids and mitochondria in *Nicotiana* are normally both inherited from the maternal parent. We observed that plastids from the *N. sylvestris* father were transmitted at a low (~0.002%) frequency via pollen. The plants that inherited paternal plastids did not carry paternal mitochondrial DNA, indicating that leakage of plastids via pollen can produce plant lines with unrelated plastids and mitochondria.

My third objective was to observe the behavior of an individual high-copy retrotransposon in *N. sylvestris*, its native host. Long terminal repeat (LTR) retrotransposons are major components of the nuclear genomes of plants, animals and fungi. The "copy-and-paste" life cycle of retrotransposons accounts for their accumulation in host genomes and permits the assumption that LTRs are identical at the time of insertion. Our objective was to experimentally determine if an introduced synthetic element would interact with native high-copy elements during retrotransposition. I present evidence that S-TNT1 co-packaged with native TNT1 elements to produce hybrid insertions with swapped LTRs and multiple recombinations within the *gag-pol* gene. We can best explain our observations by dimerization and co-packaging of TNT1 gRNAs in the cytoplasm, followed by template-switching during minus-strand DNA synthesis, which we term the "mix-and-paste" pseudodiploid mating system for LTR-retroelements.

ACKNOWLEDGEMENTS

I was supported by a Waksman Institute of Microbiology Predoctoral Fellowship and a teaching assistantship from the Department of Genetics, Rutgers University. I appreciate helpful discussions with Drs. Hugo Dooner, Yubin Li, Zora Svab, Derek Gordon, Jody Hey and Joachim Messing. Part of this work has already been published: Thyssen G, Svab Z, & Maliga P Cell-to-cell movement of plastids in plants. Proceedings of the National Academy of Sciences 109, 2439-2443.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER ONE

Cell-to-Cell Movement of Plastids in Plants

Abstract

Our objective was to test whether or not plastids and mitochondria, the two DNA-containing organelles, move between cells in plants. As our experimental approach, we grafted two different species of tobacco, *Nicotiana tabacum* and *Nicotiana sylvestris*. Grafting triggers formation of new cell-to-cell contacts, creating an opportunity to detect cell-to-cell organelle movement between the genetically distinct plants. We initiated tissue culture from sliced graft junctions and selected for clonal lines in which gentamycin resistance encoded in the *N. tabacum* nucleus was combined with spectinomycin resistance encoded in *N. sylvestris* plastids. Here, we present evidence for cell-to-cell movement of the entire 161-kb plastid genome in these plants, most likely in intact plastids. We also found that the related mitochondria were absent, suggesting independent movement of the two DNA-containing organelles. Acquisition of plastids from neighboring cells provides a mechanism by which cells may be repopulated with functioning organelles. Our finding supports the universality of intercellular organelle trafficking and may enable development of future biotechnological applications.

Introduction

Plant cells have three DNA-containing cellular compartments: the nucleus, plastids, and mitochondria. The plastid and mitochondrial genomes (ptDNA and mtDNA) in all plant

species have been massively reduced relative to their prokaryotic ancestors through the evolutionary process of intracellular gene transfer. The 155-kb plastid and 430-kb mitochondrial genomes of *Nicotiana tabacum* encode only 112 (1) and 60 (2) genes, respectively. Experimental reconstruction of this evolutionary process in the laboratory revealed that plastid-to-nucleus gene transfer occurs at a surprisingly high frequency (3). The recent demonstration of the exchange of genetic material between cells in plant tissue grafts reconstructed the evolutionary process of intercellular gene transfer (4).

Intercellular movement of mitochondria in mammalian cells was found to be a basic biological process and involved in tissue repair (5). In coculture, donor cells extend cytoplasmic projections toward target cells and mitochondria stream from cell to cell (6–8). This transfer was shown to result in replacement of diseased mitochondrial genomes with mtDNA from the donor cells (5).

However, there is no report yet on the intercellular movement of DNA-containing organelles, plastids, and mitochondria, between plant cells. In contrast to animal cells, plant cells have a rigid cell wall. However, plant cells are connected by sophisticated intercellular channels (plasmodesmata), which actively and passively regulate cell-to-cell movement of nutrients, hormones, and information macromolecules, including transcription factors, phloem proteins, mRNA, and sRNAs (9, 10).

Our objective was to determine whether chloroplasts or mitochondria could move from cell to cell in plants. To test this hypothesis, we grafted two different species of tobacco with genetic markers in their plastids, mitochondria and nuclei. Grafting triggers formation of new cell-to-cell connections (11) that creates an opportunity for cell-to-cell movement of organelles. Here, we report evidence supporting the movement of plastids

(ptDNA) between cells in graft tissue. However, the related (nonselected) mitochondria were absent in the same plants, suggesting independent transfer of plastids through the graft junction. We discuss acquisition of plastids from neighboring cells as a potential mechanism to repopulate cells with functional organelles and the possibilities of cell-to-cell movement of plastids for biotechnological applications.

Results

Experimental Design.

Because of the difficulty to directly observe rare intercellular organelle movement, we chose graft partners with distinct nuclear and organellar genomes to test for cell-to-cell transfer of plastids and mitochondria in graft junctions (Fig. 1-1). We grafted two species of tobacco, *N. tabacum* (partner P1) with a selectable transgenic nuclear gentamycin resistance gene and *Nicotiana sylvestris* (partner P2) with plastids carrying a selectable spectinomycin resistance (*aadA*) gene and the aurea young leaf color phenotype (bar^{Au} gene). The *N. sylvestris* partner carried the plastids and mitochondria of a third species, *N. undulata*, providing a large number of organellar DNA markers. The P1 partner with the *N. tabacum* nucleus was fertile and the P2 partner with the *N. sylvestris* nucleus cytoplasmic male sterile (CMS) (Fig. 1-1B), a trait controlled by mitochondria (12). The grafted plants were grown in culture for ten days (Fig. 1-2A) and sections of the graft junctions were selected for the gentamycin and spectinomycin resistance traits carried by the P1 nucleus and in P2 plastids, respectively (Fig. 1-2B). Of 30 graft junctions, a total of 3 plastid graft transmission (PGT) events (G1, G3, and G4) were recovered. The plants regenerated from the graft junction displayed the leaf morphology, growth habit, and pink

flowers associated with the selected *N. tabacum* nucleus but the aurea leaf color of the P2 partner, a plastid trait (Fig. 1-1 A and B).

No Exchange of Chromosomes in the PGT Plants.

To investigate the contribution of nuclear genetic material to the PGT plants, we examined 24 simple sequence repeat (SSR) (or microsatellite) polymorphic DNA markers previously mapped to each of the *N. tabacum* chromosomes (13). These markers distinguished *N. tabacum* from *N. sylvestris* ecotype TW137 and indicated the presence of the chromosomes of the *N. tabacum* P1 partner that carried the selectable nuclear gene without contribution from the nonselected P2 *N. sylvestris* nucleus (Fig. 1-3). The presence of chromosomal markers from one partner excluded chimera formation as the source of double resistance of the G1, G3, and G4 PGT plants. However, we cannot exclude limited transfer of chromosome fragments that remained undetected in the study.

Mitochondria Remain Associated with the Selected Nucleus.

The graft partners carried distinct mitochondrial genomes determining the flower type (Fig. 1-1B). The P1 partner with the *N. tabacum* nucleus had normal anthers and produced fertile pollen, whereas the P2 partner with the *N. sylvestris* nucleus had stigmatoid anthers, a phenotype controlled by mitochondria. The G1, G3, and G4 PGT plants were male fertile and lacked the stigmatoid anthers of the CMS P2 partner. In line with the flower morphology, the CMS92 mtDNA markers were absent in the G1, G3, and G4 plants. To determine the source of the mitochondrial genome in the PGT plants, we identified six SNP and insertion/deletion markers that are suitable to distinguish the *N.*

undulata CMS92 mtDNA (Fig. 1-4) from the fertile *N. tabacum* mtDNA (2). Sanger sequencing of PCR fragments indicated that the G1, G3, and G4 plants have the mitochondrial genome of the nuclear donor (Fig. 1-4). Thus, we did not find evidence for the transfer of mitochondrial DNA in the PGT plants. Given the tendency of mitochondria for fusion (14) and mtDNA for recombination (12, 15), should mtDNA be transferred, we would expect to find at least chimeric mtDNA. The absence of nonselected mitochondrial DNA suggests limited organelle transfer, rather than large-scale mixing of the two cytoplasms at the graft junction. Although we did not find evidence for the cotransfer of mtDNA with plastids in the lines tested, it is possible that mtDNA transfer could be detected with selection in a larger PGT plant population.

PGT Plants Contain the Entire Selected Plastid Genome.

Dual selection for the nucleus- and plastid-encoded antibiotic resistances ensured that the PGT plants would carry both transgenes. The *N. tabacum*-specific SSR markers in the G1, G3, and G4 plants indicated the presence of the P1 chromosomes alone in the PGT plants. However, the presence of the plastid markers did not distinguish between a transformation-like process that involves incorporation of ptDNA fragments and intercellular movement of plastids implied by the transfer of complete plastid genomes, either of which is compatible with the earlier report (4). To determine how much of the P2 ptDNA is present in the G1, G3, and G4 plants, we first examined markers distant from the transgenes by probing total cellular DNA on blots. Southern probing of the six previously identified RFLP markers (Fig. 1-5C) and PCR analyses (Fig. 1-5B) suggested the presence of the entire plastid genome of the P2 partner and that the PGT plants

carried a uniform population of P2 transplastomes. To exhaust the search for a contribution to the PGT plastid genomes from the nonselected P1 plastome, we performed next-generation sequencing of the plastid genomes of the P1 and P2 partners and the G1, G3, and G4 PGT plants. We report here that the sequence of the 160,743-nt transplastomes in the P2 partner and in three PGT plants are identical (GenBank accession no. JN563930). The P2 and PGT plastid genomes are larger than the 155,863-nt wild-type *N. undulata* plastid genome (GenBank accession no. JN563929) because the transplastomes also contain the spectinomycin resistance (*aadA*) and the aurea bar^{Au} transgenes. We also sequenced the plastid genome in partner P1 that carries the wild-type *N. tabacum* ptDNA of cv. Petit Havana. We have found that the sequence of cv. Petit Havana ptDNA is identical to the cv. Bright Yellow sequence deposited in GenBank (GenBank Accession number Z00044). However, the *N. undulata* ptDNA differs from the *N. tabacum* cv. Petit Havana ptDNA by 805 SNPs, 52 insertions, and 61 deletions and the transgene cassettes. Differences between the plastid genomes are depicted on the mVISTA identity plots shown in Fig. 1-5A and Fig. 1-S1: the 500-bp sliding window in Fig. 1-5A gives an overview; and the 100-bp sliding window in Fig. 1-S1 provides more precise information about the location of SNPs and insertions/deletions. Importantly, we observed all of these polymorphic loci, with an average density of 200 bp/SNP (170 bp/polymorphism) in the plastid genome of the three graft transmission plants indicating the transfer of intact ptDNA from the P2 graft partner. We also tested transmission of the plastid-encoded spectinomycin resistance in reciprocal backcrosses with the G1 PGT plant. When the G1 plant was the mother and the wild type the father, each of the 208 seedlings was resistant, whereas when the G1 plant was the father and the wild type the

mother, each of the 318 seedlings was spectinomycin sensitive. Thus, spectinomycin resistance exhibited uniform, maternal inheritance, as expected for a homoplastomic *N. tabacum*, a species with strict maternal plastid inheritance (16, 17).

Discussion

Cell-to-Cell Migration of Plastids.

Here, we report cell-to-cell movement of entire plastid genomes. We considered two possible mechanisms for the transfer of genome-size ptDNA: the intercellular transport of extraorganellar ("naked") DNA or the ptDNA traveling within an intact organelle. Selection for movement of ptDNA to the nucleus led to the discovery that incorporation of kilobase-size ptDNA fragments is frequent and that the source most probably is degraded organellar genomes (18–20). Movement of entire genomes may require more protection than the fragments. Better protection could be provided if the extraorganellar ptDNA would be encapsulated in membrane-bound vesicles that are shed from fragmented chloroplast stromules (21), although ptDNA is normally absent from stromules (22). Because of the need for capacity for translation, plastids cannot be created de novo from membranes and DNA (23). Thus, if "naked" ptDNA is transferred, an invading plastome would need to enter an existing plastid with transcription and translation machinery and displace the existing plastome by a transformation-like process to explain our observations. However, a transformation-like process would yield mosaic genomes if different genomes were present, because plastid genomes within an organelle undergo frequent recombination (24–26). The absence of chimeric genomes in the PGT

plants makes it unlikely that naked DNA transfer is the mechanism of intercellular ptDNA transfer.

More likely, vehicles of cell-to-cell movement of entire plastid genomes could be the organelles themselves. The avenue for the movement of intact organelles could be damage to cell walls that allows for some mixing of cytoplasms in the graft junctions. A more likely mechanism would be the transfer of proplastids via newly formed connections between cells that are well documented at graft junctions (11). The size of proplastids, ~1 μm, is well above the size exclusion limit of plasmodesmata normally defined by molecular mass. However, the size exclusion limit changes during development and depends on tissue type (9, 27). We speculate that the new openings, formed by thinning of opposing cell walls at the site of future plasmodesmata, permit intercellular movement of proplastids. Our preferred model of intercellular plastid transfer in graft junctions is shown in Fig. 1-6.

The Role of Cell-to-Cell Movement of Plastids.

The capacity of a plant cell to acquire organelles from a neighboring cell is a basic biological process. Acquisition of plastids from neighboring cells may be important because once the ribosomes are lost, translation cannot be restored, because some of the ribosomal proteins are encoded in the plastid genome and their translation is dependent on plastid ribosomes (23). Therefore, during certain stages of development, including dedifferentiation associated with forming new connections in grafted tissues (11), plants cells may allow intercellular transport of organelles. In this regard it is intriguing to note that the redox state of plastids regulates symplastic permeability and that ectopic

expression of the proplastid-targeted GAT1 protein increased plasmodesmal size exclusion limit (28). The functional state of mitochondria also regulates the size exclusion limit of intercellular trafficking (29) and reprogramming of diseased mammalian cells was associated with acquisition of functional mitochondria (5, 7). The discovery of intercellular movement of plastids supports the universality of intercellular organelle trafficking and calls for testing the biological significance of this process in plants.

Horizontal Gene Transfer and Cell-to-Cell Movement of Organelles.

The formation of interspecific cytoplasmic connections and exchange of genetic material has also been reported between parasitic flowering plants and their hosts. All interspecific secondary plasmodesmata have been localized in thinned-wall areas at the contact between host and parasite, which corresponds to the observations on graft unions (11). Although horizontal gene transfer (HGT) in plant mitochondrial genomes is rampant when a parasitic flowering plant is involved as a donor or recipient, it very rarely occurs in plastids (30, 31). Cell-to-cell movement of plant mitochondria and the observed massive mitochondrial fusion (14) would provide an efficient mechanism for evolutionary gene transfer. In contrast, plastid fusion has been rarely observed under experimental conditions (25, 32, 33), explaining the scarcity of HGT. The cell-to-cell movement of entire plastids is restricted to closely related species, because plastid-nucleus incompatibility prevents incorporation of entire unmodified ptDNA in a distantly related host (34, 35). However, fragments of the incoming plastid genome may find their way into the nucleus and mitochondria of the host.

Applications in Plastid Genetics and Biotechnology.

Because in most species both plastids and mitochondria are maternally inherited, they cannot be separated by crossing. Thus far protoplast fusion has been the only option to obtain new combinations of plastids and mitochondria (12). The result is intercellular transfer of parental plastids, but formation of recombinant mitochondrial genomes. The protocol we report here enables combination of parental plastids and nonrecombinant mitochondria by PGT, a significant improvement over the protoplast-based process that yields recombinant mitochondria.

An additional application of PGT could be rapid introgression of transformed plastids into commercial cultivars. Plastid transformation is a powerful tool for biotechnological applications because the transgenes that are integrated into the plastid genome are expressed at high levels, can be clustered in operons, and are not subject to silencing (36, 37). Currently, the option is to transform the plastids in permissive cultivars then introduce them into commercial lines by repeated backcrossing using the commercial cultivar as a recurrent pollen parent. Based on the findings in this report, backcrossing can be replaced in the future by graft transfer of the transformed plastids, instantly yielding a substitution line with transgenic plastids combined with the valuable commercial nuclear genome.

Materials and Methods

Partner P1 (Nt-pHC19) has an allotetraploid *N. tabacum* cv. Petit Havana (2N = 48) nucleus with the *aacC1* transgene for gentamycin resistance and wild-type *N. tabacum* plastid and mitochondrial genomes (38). Partner P2 (Ns-pCK2-6W2) has a wild-type

diploid *N. sylvestris* TW137 (2N = 24) nuclear genome, *N. undulata* plastids with *aadA* transgenes for spectinomycin selection and the aurea young leaf color phenotype (*bar*^{Au} gene), and *N. undulata* (CMS-92) mitochondria that confer cytoplasmic male sterility (39). For grafting, the plants were grown aseptically on a medium containing Murashige and Skoog salts and 3% sucrose (40). Grafting was carried out reciprocally in equal numbers. Plants were regenerated from the graft junctions on RMOP shoot regeneration media supplemented with 500 mg/L spectinomycin and 100 mg/L gentamycin (40). Southern probing for ptDNA polymorphisms was carried out using six previously identified polymorphic regions (16). Organellar DNA was amplified using total cellular DNA as a template (41) and appropriate PCR primers (Tables 1-S1 and 1-S2). Primer design for ptDNA was based on GenBank accession nos. Z00044 and JN563929; for mtDNA, primer design was based on GenBank accession no. BA000042. The P1, P2, and G1 plastid genomes were amplified in 34 PCRs using primers listed in Table 1-S3. DNA sequence was determined on an Illumina Genome Analyzer II using 80-bp paired-end (500-bp insert) library. Total leaf DNA fragments of P1, P2, G1, G3, and G4 plants were also analyzed on a SOLiD 5500xl sequencer (Applied Biosystems) using 76-nt reads. Reference guided assembly of the ptDNA was essentially carried out as described elsewhere (42). Nuclear SSR markers (13) were amplified using primers listed in Table 1-S4.

	P1	G1	P2
Nucleus	*N. tabacum* *Gent* ^{R}	*N. tabacum* *Gent* ^{R}	*N. sylvestris*
Plastids	*N. tabacum*	*N. undulata* *aadA+bar* ^{Au}	*N. undulata* *aadA+bar* ^{Au}
Mitochon-dria	*N. tabacum* Fertile	*N. tabacum* Fertile	*N. undulata* CMS

FIGURE 1-1. Phenotypes of the graft partners and the G1 graft transfer plant. (A) Partner P1 is *N. tabacum* (2N = 48) with a nuclear gentamycin resistance transgene and wild-type *N. tabacum* plastids and mitochondria. Partner P2 has a wild-type *N. sylvestris* (2N = 24) nuclear genome, *N. undulata* plastids with *aadA* transgenes for spectinomycin selection and the aurea young leaf color phenotype (*bar* ^{Au} gene), and *N. undulata* mitochondria that confer cytoplasmic male sterility (CMS-92). Shown is also the G1 plant and its markers. (Black scale bar: 10 cm.) (B) Flower morphology of the P1 and P2 partners and G1 PGT plant. (White scale bar: 1 cm.)

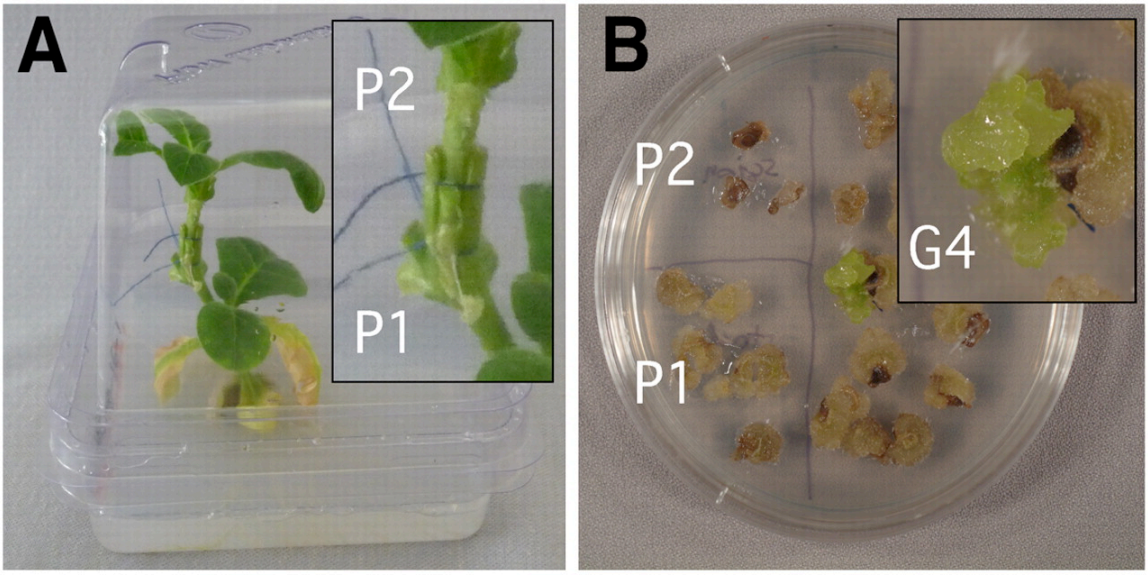


FIGURE. 1-2. Identification of plastid graft transfer events. (A) Grafted plant. Note that the P2 scion shown here is green because the expression of the *bar*^{Au} gene is restricted to fast-growing tissue and is sensitive to environmental conditions. (B) Selection in cultures of 1–2-mm graft sections for gentamycin and spectinomycin resistance. On the left are stem sections from above (P2) and below (P1) the graft and on the right from the graft region. Note a green, proliferating callus that yielded the G4 PGT plants.

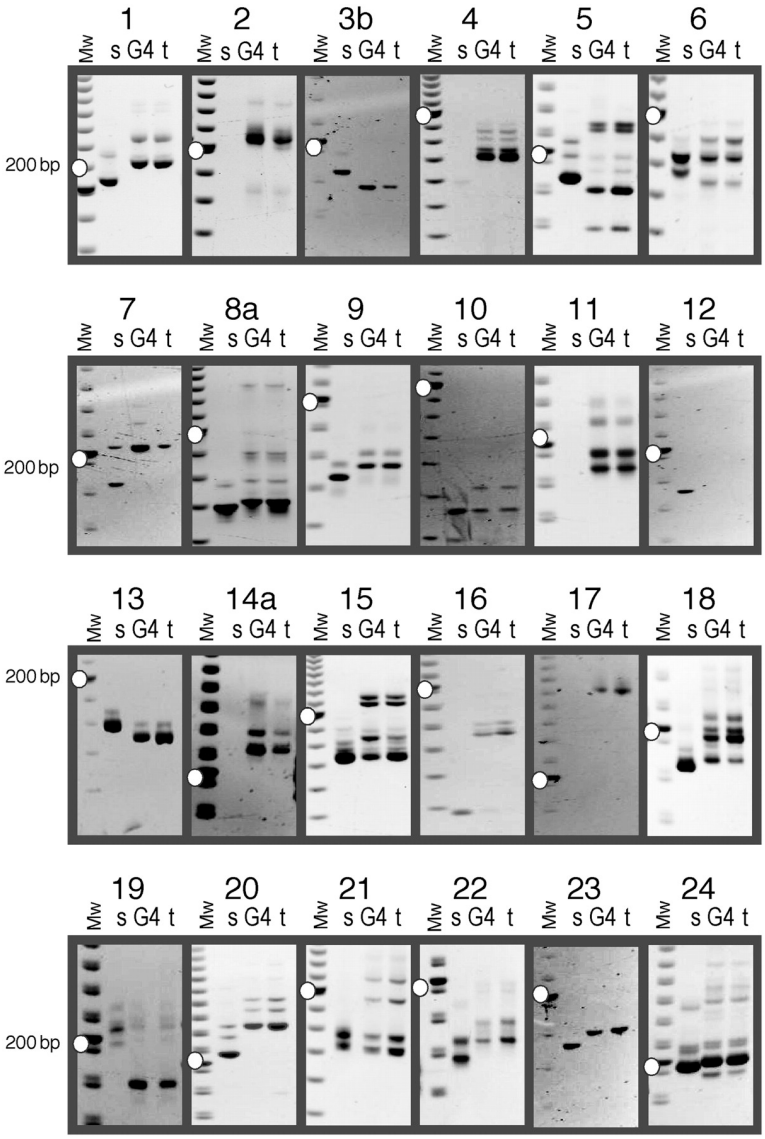
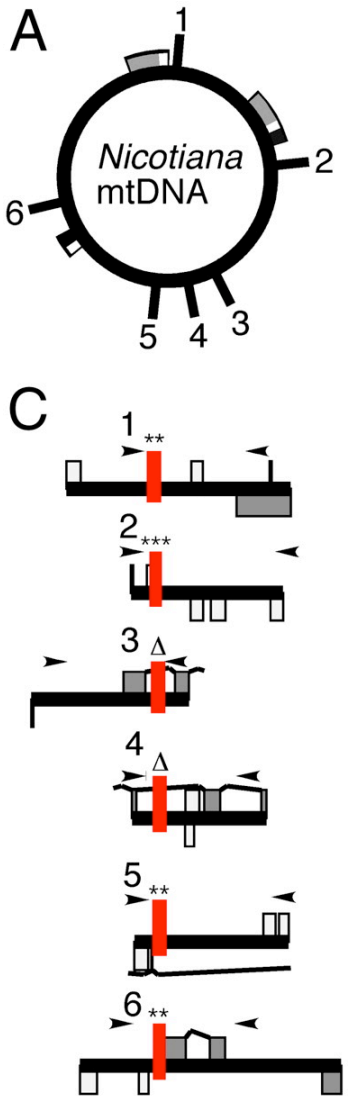


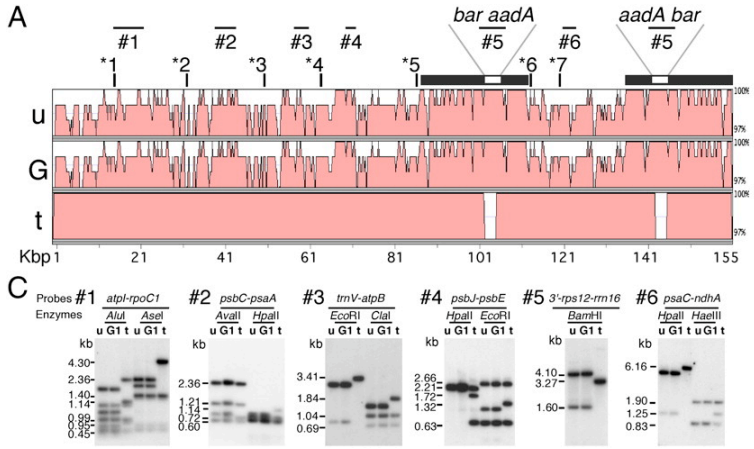
FIGURE 1-3. SSR markers confirm *N. tabacum* chromosomes in the G4 plant by testing each of the 24 chromosomes (numbered 1–24). Lanes are marked with s, G4, and t for the P2, G4, and P1 plants, respectively (see legend of Fig. 1-1). Some markers do not amplify the *N. sylvestris* template (13). White dots indicate the 200-bp fragment of the 20-bp molecular-mass ladder.



B

	Gene	Mitochondrial DNA Sequences
1	*orf125a*	*N.t.* aaaaccaat **A** tcataacttt 1027 *N.u.* aaaaccaat **C** tcataacttt G1-4 aaaaccaat **A** tcataacttt *N.t.* gaaacgaag **T** ggagcgaggg 1078 *N.u.* gaaacgaag **C** ggagcgaggg G1-4 gaaacgaag **T** ggagcgaggg
2	*orf129b*	*N.t.* aaaaaattt **A** tctcttttta 100418 *N.u.* aaaaaattt **C** tctcttttta G1-4 aaaaaattt **A** tctcttttta *N.t.* ctttgttct **G** tagtctttca 100531 *N.u.* ctttgttct **C** tagtctttca G1-4 ctttgttct **G** tagtctttca *N.t.* ctcccattt **G** aaagattatt 100821 *N.u.* ctcccattt **T** aaagattatt G1-4 ctcccattt **G** aaagattatt
3	*nad4* intron	*N.t.* tgcac **CTCCGTACAA** gtgct 183637 *N.u.* tgcac ---------- gtgct G1-4 tgcac **CTCCGTACAA** gtgct
4	*nad2* intron	*N.t.* ttccg **GCTGTTTCCG** tcatt 202083 *N.u.* ttccg ---------- tcatt G1-4 ttccg **GCTGTTTCCG** tcatt
5	*nad5* intron	*N.t.* aagaaaaaa **GA** aaagtcgag 222508 *N.u.* aagaaaaaa **TC** aaagtcgag G1-4 aagaaaaaa **GA** aaagtcgag
6	*orf115-ccmFc* spacer	*N.t.* ggatcagac **T** actcctggtg 306939 *N.u.* ggatcagac **C** actcctggtg G1-4 ggatcagac **T** actcctggtg *N.t.* agcaaaact **C** gaacggatag 307051 *N.u.* agcaaaact **A** gaacggatag G1-4 agcaaaact **C** gaacggatag

FIGURE 1-4. Identification of the source of mtDNA in the PGT plants. (A) Schematic representation of the tobacco mtDNA master circle with the position of polymorphic regions marked. Repeated regions are marked with boxes. (B) Mitochondrial DNA sequence polymorphisms. (C) Map position of polymorphic sites relative to the sequencing primers and gene features.



	Gene	Plastid Sequence Polymorphism
*1	*atpF* intron	*N.t.* taga **G** aatcaaaaat **T** gattacc *N.u.* taga **A** aatcaaaat **C** gatttacc G taga **A** aatcaaaat **C** gatttacc
*2	*rpoB-trnC* spacer	*N.t.* t **C** tgttatggatttcgttg **Ct A** a *N.u.* t **A** tgttatggatttcgttg **Tt C** a G t **A** tgttatggatttcgttg **Tt C** a
*3	*trnL* intron	*N.t.* tca **AAATCA** aat...aatc **T** gta *N.u.* tca ------ aat...aatc **C** gta G tca ------ aat...aatc **C** gta
*4	*accD*	*N.t.* caa **C** gccta **C** attgcatttgcag *N.u.* caa **T** gccta **T** attgcatttgcag G caa **T** gccta **T** attgcatttgcag
*5	*rpl16-rps3* spacer	*N.t.* tta **G** aaatgtt **T** cccttgattga *N.u.* tta **T** aaatgtt **C** cccttgattga G tta **T** aaatgtt **C** cccttgattga
*6	*ndhF*	*N.t.* agt **A** aattcag...ccaac **T** aag *N.u.* agt **C** aattcag...ccaac **A** aag G agt **C** aattcag...ccaac **A** aag
*7	*ndhE-ndhG* spacer	*N.t.* atc **T** cgactccttttc ----- aat *N.u.* atc **G** cgactccttttc **CTTTC** aat G atc **G** cgactccttttc **CTTTC** aat

FIGURE. 1-5. Identification of the *N. undulata* plastids in the PGT plants. (A) Identity plots of the plastid genomes of the transplastomic P2 partner carrying *N. undulata* ptDNA (u) with the *aadA* and *bar*^{Au} transgenes (GenBank accession no. JN563930); the G1, G3, and G4 (G) PGT plants; and the P1 partner with *N. tabacum* ptDNA (t) (GenBank accession no. Z00044) aligned with the mVISTA program using a 500-bp sliding window. Shown above the map are the positions of the DNA probes (#1 through #6) and DNA polymorphisms (*1 through *7). (B) Plastid DNA sequence polymorphisms. For map position, see Fig. 1-5A. (C) DNA gel blot to identify RFLP markers in ptDNA. For probes see Fig. 1-5A.

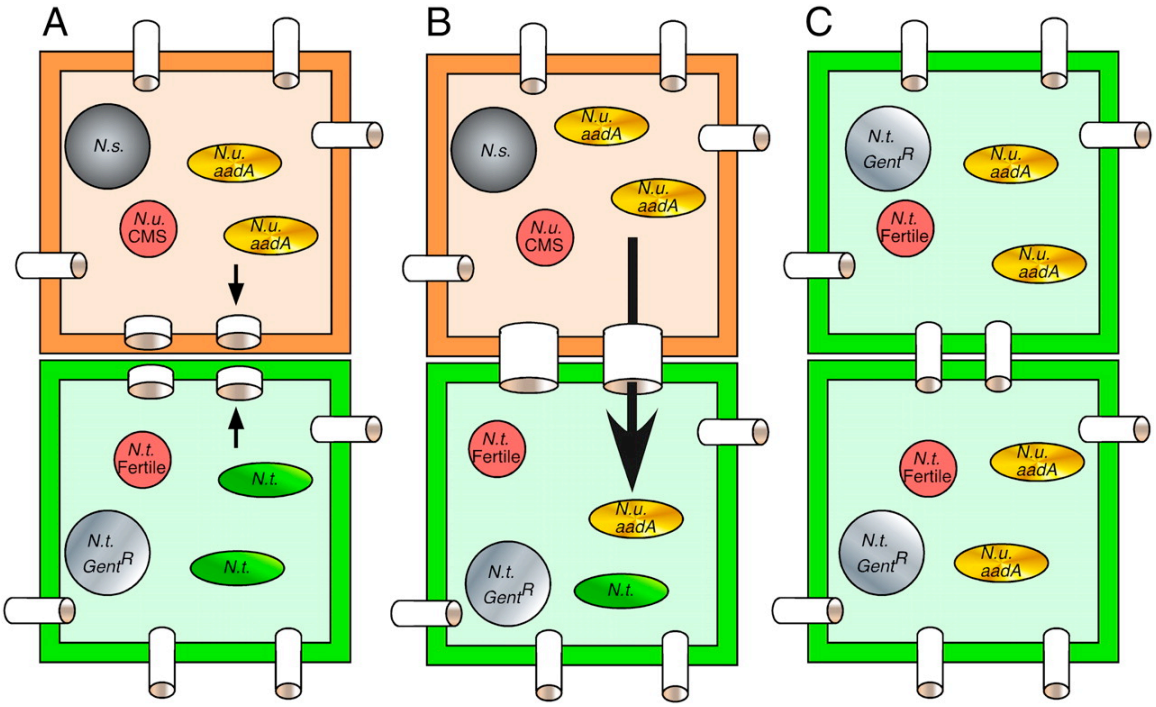


FIGURE. 1-6. Model for cell-to-cell movement of plastids via initial cytoplasmic connection in graft junctions. (A) Cells at graft junction reconnect by plasmodesmata. Arrows point to sites where opposite parts of the contact walls are synchronously thinned (11). These are future sites of plasmodesmata. Proplastids (ovals), mitochondria (small circles), and nuclei (large circles) are identified in scion and rootstock. Ns, *N. sylvestris*; Nt, *N. tabacum*; Nu, *N. undulata*. (B) Proplastid is transferred via initial cytoplasmic connection. (C) Transferred spectinomycin-resistant plastid takes over on selective medium. Note that the cells derive from the bottom cell in Fig. 1-6B.

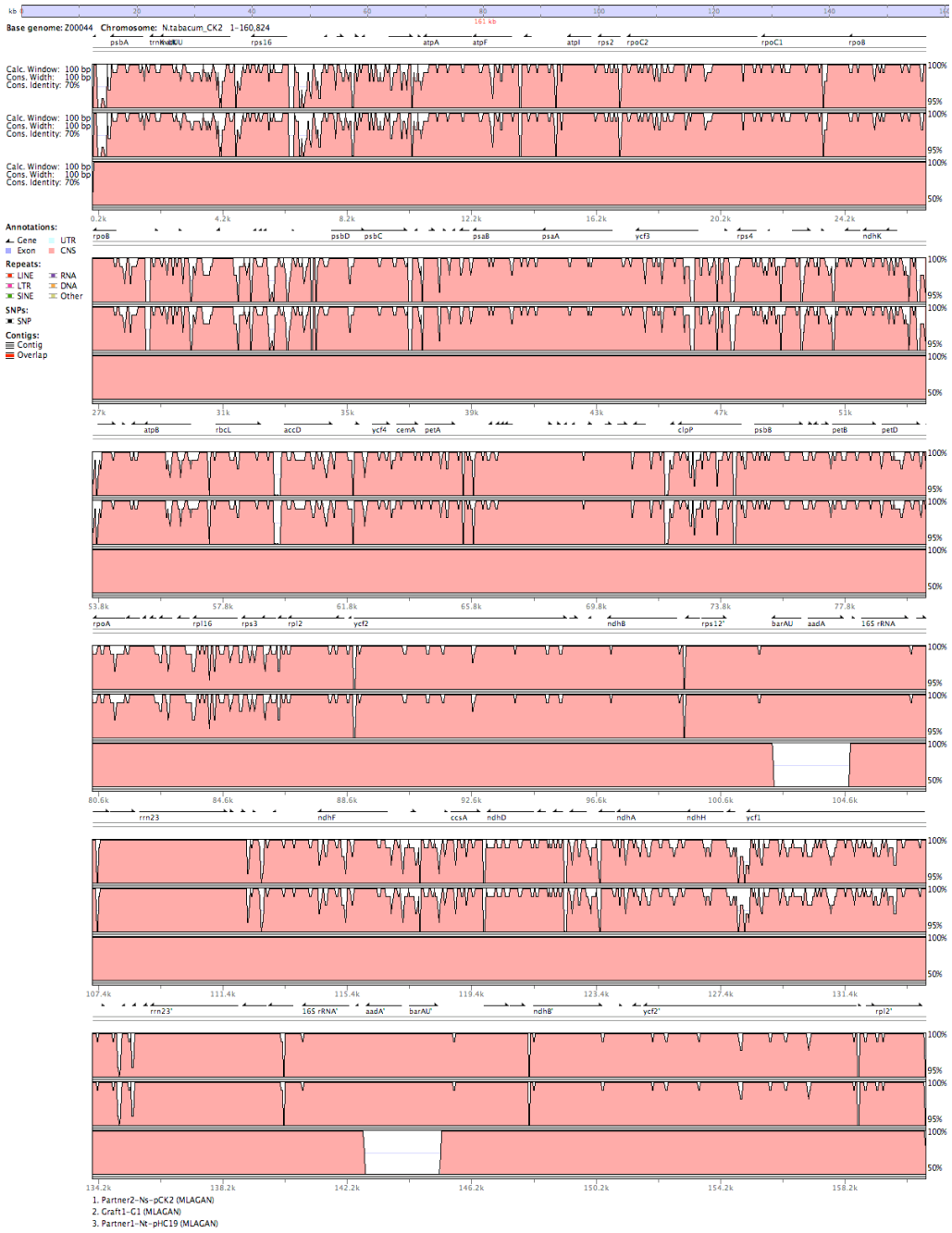


FIGURE 1-S1. Identification of the *N. undulata* plastids in the PGT plants. mVISTA-based identity plots (100-bp sliding window) showing sequence identity between sequenced chloroplast genomes of the transplastomic P2 partner carrying *N. undulata* ptDNA (u) and the *aadA* and *bar* ^{Au} transgenes (GenBank accession no. JN563930) (Top); the G1, G3, and G4 (G) PGT plants (Middle); and the P1 partner *N. tabacum* ptDNA (t) (GenBank accession no. Z00044) (Bottom) as the reference. Shown above the map are the positions of the genes.

TABLE 1-S1. Plastid primers for testing ptDNA polymorphic sites between *N. tabacum* and *N. undulata*

Pair	Primer	Position	Strand	Gene	Sequence
*1	12upF	12907	F	atpF	TCTTACTTAGAATAGGTCGTCGATTCAGCA
*1	14upR	14098	R	atpF	CCACTGATTTCTGCCGCTTCCGTT
*2	27upF	27875	F	rpoB-trnC	ACACATTCCAACCTGCTTGAATACCA
*2	29upR	29210	R	rpoB-trnC	TCTTCCGCCCCCTTCCACAACTAT
*3	48upF	48971	F	trnL	GAGACATTCCTCCGCTTTCAGGCG
*3	49upR	49945	R	trnL	TGGAACCGCTAAGGAAAGGGGGTC
*4	60upF	60806	F	accD	AACGGCATTCCCGTAGCAATTGGG
*4	62upR	62222	R	accD	GGATGAGATTGGGTCCCAGCGGAT
*5	83upF	83888	F	ndhF	TTTCCACCACGACGTGCATTTCGT
*5	85upR	85414	R	ndhF	TACAAATTGCGGGGCGTATCGACG
*6	111upF	111916	F	ndhE-ndhG	TCGGAAGAAAGGTGGGATCCGGAC
*6	113upR	113293	R	ndhE-ndhG	TGGTATGGGGTCTTATCGAAGCGC

TABLE 1-S2. Mitochondrial primers for testing mtDNA polymorphic sites between *N. tabacum* and *N. undulata*

Pair	Primer	Position	Strand	Gene	Sequence
1	mt-0-F	690	F	orf125a	CCCCGCCCAGTAGTGCCTCT
1	mt-4-R	4334	R	orf125a	CCGCGGGCATCGCGATAAGT
2	mt-100-F	100070	F	orf129b	CGGCCATCCTGGTCCTCAGGA
2	mt-104-R	104811	R	orf129b	TGGGGACTCGCACGAGGAGG
3	mt-180-F	180316	F	nad4	GGCAGGAGCGCAACGACCTT
3	mt-183-R	183813	R	nad4	AGTCGGGTTGCTCACGCAGC
4	mt-201-F	201586	F	nad2	TGGTGTGCTTCCTGCTCGCG
4	mt-204-R	204759	R	nad2	TTTCTCCGTGCCCGTTCCGC
5	mt-222-F	222140	F	nad5	AGGTGCCCGTAGTAGGCCGG
5	mt-226-R	226463	R	nad5	TTGGGCTTGGCTCTGCTCGC
6	mt-306-F	306203	F	orf115-ccmFc	CACGACTCCCCCTCTCCCCG
6	mt-309-R	309623	R	orf115-ccmFc	TGCCCGATTCCCCGACCCAT

TABLE 1-S3. Plastid primers for PCR amplification of the *N. tabacum* and *N. undulata* plastid genomes

Pair	Primer	Position	Strand	Gene	Sequence
1	0F	14	F	trnH	ACGGGAATTGAACCCGCGCA
1	4R	4410	R	trnK	CGGGTTGCTAACTCAACGG
2	3F	3704	F	trnK	TCAAATGATACATAGTGCGATACA
2	8R	8653	R	trnS	CGAATCCCTCTCTTTCCG
3	7F	7989	F	psbK	GCCTTTGTTTGGCAAGCTGCTGTAAG
3	12R	12042	R	atpA	GGCATTGCTCGTATTCACGGTCTTG
4	11F	11052	F	atpA	CCACTCTGGAAACGGAGATACCC
4	16R	16791	R	rps2	CTCGTTTTTTATCAGAAGCTTGTG
5	15F	15267	F	atpI	GATGGCCCTCCATGGATTCACC
5	20R	20888	R	rpoC2	GAGGATTAATGTCAGATCCTCAAGG
6	19F	19971	F	rpoC2	GATAGACATCGGTACTCCAGTGC
6	24R	24612	R	rpoB	GTTACACAACAACCCCTTAGAGG
7	24F	24069	F	rpoC1	GCACAAATTCCGCTTTTTATAGG
7	29R	29568	R	ycf6	GCCCAAGCAAGACTTACTATATCCAT
8	28F	28849	F	trnC	CCAGTTCAAATCCGGGTGTC
8	34R	34493	R	psbD	TACCAAGGGCTATAGTCAT
9	33F	33186	F	trnT	GCCCTTTTAACTCAGTGGTA
9	38R	38115	R	trnG	AACCCGCATCTTCTCCTTGG
10	37F	37147	F	trnS	GAGAGAGAGGGATTCGAACC
10	43R	43484	R	psaA	TTCGTTCGCCGGAACCAGAA
11	41F	41267	F	psaA	AAGAATGCCCATGTTGTGGC
11	46R	46162	R	ycf3	CCTATTACAGAGATGGTGCGATTT
12	45F	45083	F	ycf3	CGATGCATATGTAGAAAGCC
12	51R	51022	R	ndhJ	TTTTTATGAAATACAAGATGCTC
13	49F	49312	F	trnL	CGAAATCGGTAGACGCTACG
13	54R	54971	R	atpE	GAAGGAAGGAGACAAAAAATTGAGGC
14	53F	53776	F	trnV	CGAACCGTAGACCTTCTCGG
14	58R	58198	R	rbcL	GTAAAATCAAGTCCACCGCG
15	57F	57272	F	atpB	TCTAGGATTTACATATACAACAT
15	62R	62754	R	ycf4	CTAATAAGAAGCCTAATGAACC
16	61F	61145	F	accD	GCAGGTAAAAGAGTAATTGAAC
16	66R	66664	R	psbL	TACTCATTTTTGTACTTGCTGT
17	65F	65219	F	petA	GCATCTGTTATTTTGGCACA
17	71R	71704	R	clpP	ACCATAGAAACGAAGGAACCCACT
18	70F	70727	F	rps18	GCTCGTATTTTATCTTTGTTACC
18	76R	76301	R	psbB	CCCCTTGGACTGCTACGAAAAACACC
19	74F	74963	F	psbB	TGCCTTGGTATCGTGTTCATAC
19	78R	78846	R	petB	CCCAGAAATACCTTGTTTACG
20	77F	77212	F	psbH	TGGGGAACTACTCCTTTGAT
20	82R	82676	R	rps8	CGAGGTATAATGACAGACCGAG

21	81F	81880	F	rpl36	ATTCTACGTGCACCCTTACG
21	86R	86576	R	rps19	GGGCATCTACCATTATACCC
22	85F	85864	F	rps3	AGTCTGAAACCAAGTGGATTTATT
22	89R	89311	R	YCF2	GAAGATACAGGAGCGAAACAATCAAC
23	88F	88062	F	rpl2	GCTTATGACCTCCCCCTCTATGC
23	93R	93140	R	YCF2	TCTTCTAGAGAATCTCCTAATTGTTC
24	91F	91131	F	YCF2	CTTCGAATATGGAATTCAAAGGGATC
24	97R	97636	R	ndhB	CTCAAACAAGCATGAAACGTATGC
25	96F	96469	F	trnL	GAGATTTTGAGTCTCGCGTGTC
25	100R	100782	R	rps12	TCACTGCTTATATACCCGGTATTGGC
26	99F	99552	F	rps7	GTGCAAAAGCTCTATTTGCCTCTGCC
26	104R	104797	R	oriA	ATCGAAAGTTGGATCTACATTGGATC
27	103F	103454	F	rrn16	CGACACTGACACTGAGAGACGAAAGC
27	108R	108280	R	rrn23	CGCTACCTTAGGACCGTTATAGTTAC
28	107F	107056	F	rrn23	GAAACTAAGTGGAGGTCCGAACCGAC
28	111R	111882	R	ORF350	AGTGGATCCCTCTTGTTCCTGTTTAG
29	110F	110672	F	trnN	ACAGCCGACCGCTCTACCACTGAGC
29	114R	114269	R	ndhF	GGATCATACCTTTCATTCCACTTCC
30	113F	113036	F	ndhF	ATTTCATCTTTGGACCAAAAACAAGC
30	119R	119286	R	psaC	GCTAAACAAATTGCTTCTGCTCC
31	117F	117227	F	ycf5	GGTCAATCTTTTAGGAATAGGGTTAC
31	123R	123506	R	ndhA	GGACTTCTTATGTCGGGATATGGATC
32	122F	122194	F	ndhA	CTGCGCTTCCACTATATCAACTGTAC
32	128R	128835	R	ycf1	TGAAACCTTGGCATATATCT
33	127F	127391	F	ycf1	AATTTCGAGGTTCTTATTTACT
33	132R	132957	R	trnR	GACGATACTGTAGGGGAGGTC
34	154F	154629	F	rpl2	CCATAGAATACGACCCTAAT
34	1R	1533	R	psbA	CTAGCACTGAAAACCGTCTT

TABLE 1-S4. Nuclear SSR primers. F, forward; R, reverse.

Chromosome	Primer	Strand	Sequence
1	PT30307	F	AAAGAAGCACGGTCAAATAGG
1	PT30307	R	GCAACAACAAGGTGTCATGG
2	PT30242	F	TGTGTACTACCGGCCTACTGC
2	PT30242	R	TTCTGCTAAACCGATCGTGG
3b	PT30205	F	GGTCGATCCACAATTTAAACG
3b	PT30205	R	GCACTTGCTCCTTTGTACCC
4	PT30272	F	GAACCTAACCTCGCTCCACA
4	PT30272	R	AAATGGTAGCTGCGAGGAGA
5	PT30471	F	GTCTGTACCTTCGCCAAAGC
5	PT30471	R	TCCTCAGAGAACTCCAGCGT
6	PT30087	F	CTTCTTCCTAAGCCGAGGGT
6	PT30087	R	TTGATGATAGAACGCAACTCG
7	PT30138	F	AGTTGCAGGATTGTTCGCTT
7	PT30138	R	CGACTGCAAGAGTTGGCAAT
8a	PT30167	F	TGATACAGAATATGGCGAACTTT
8a	PT30167	R	CCGCTTCATCATTGAGGTTT
9	PT30140	F	AAGATGGCATATGGGATTGG
9	PT30140	R	TGAATCGGAGGAAGTGAATG
10	PT30482	F	CTTCTCTCTCCACCGCAGAC
10	PT30482	R	ACAGTTGGATATGGTGGCGT
11	PT30008	F	CGTTGCTTAGTCTCGCACTG
11	PT30008	R	GGTTGATCCGACACTATTACGA
12	PT30098	F	TTGTTGCTCTCTCGAGTTCTTT
12	PT30098	R	GCAGTCGACTCATTGGCA
13	PT30342	F	GACAACAATCAGTAAAGGAAACGA
13	PT30342	R	AATGCAAGACCCTGTCAACC
13	PT30420	F	AACAAACCGCTTTCCATTCT
13	PT30420	R	GAATTAGGCGCTTTGGGAAT
14a	PT30175	F	TTAGGCGGCGGTATTCTTAT
14a	PT30175	R	TATGCCTCAATCCCTTACGC
15	PT30463	F	AAGCTGCCCTAGCTCAATCA
15	PT30463	R	AACATCACCATTTCCACAAGTTT
16	PT30412	F	CATTTAGCCGGGAACATTCA
16	PT30412	R	CATGGGATACACACGCAAAG
17	PT30274	F	TGACAGCTAAGCTAATAACAGTAAATG
17	PT30274	R	GGACTTTGGAGTGTCAAATGC
18	PT30111	F	AGCCAGCCACCAAATTTATC
18	PT30111	R	GGAACATTGCTCAAGCCCTA

19	PT30230	F	TTTCTTTCTGTCTGATGCTTCAAT
19	PT30230	R	TTGTCCATCTCACTTGCTGC
20	PT20286	F	ACGCTAGAGCATCCAACA
20	PT20286	R	TAGTGAAAGGCAAGCAGG
21	PT30378	F	TCAAATGAGGGTTGTAGCCA
21	PT30378	R	TGCAATGGCTACACAAGAAGA
22	PT30168	F	TTGAACACCAATTGCGGTAA
22	PT30168	R	AAATTCTTGGGTCATGGTGG
23	PT30231	F	AGGAGGCGAAGAAAGAGGAG
23	PT30231	R	CCCATGAATTCGTAACAGCA
24	PT40024	F	AATGTCTGCCCAATCGAAAG
24	PT40024	R	CGAATAACGACACTCGAACG

REFERENCES FOR CHAPTER ONE

1. Shimada H, Sugiura M (1991) Fine structural features of the chloroplast genome: Comparison of the sequenced chloroplast genomes. Nucleic Acids Res 19:983–995.
2. Sugiyama Y, et al. (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: Comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics 272:603–615.
3. Bock R, Timmis JN (2008) Reconstructing evolution: Gene transfer from plastids to the nucleus. Bioessays 30:556–566.
4. Stegemann S, Bock R (2009) Exchange of genetic material between cells in plant tissue grafts. Science 324:649–651.
5. Spees JL, Olson SD, Whitney MJ, Prockop DJ (2006) Mitochondrial transfer between cells can rescue aerobic respiration. Proc Natl Acad Sci USA 103:1283–1288.
6. Onfelt B, et al. (2006) Structurally distinct membrane nanotubes between human macrophages support long-distance vesicular traffic or surfing of bacteria. J Immunol 177:8476–8483.
7. Acquistapace A, et al. (2011) Human mesenchymal stem cells reprogram adult cardiomyocytes toward a progenitor-like state through partial cell fusion and mitochondria transfer. Stem Cells 29:812–824.
8. Koyanagi M, Brandes RP, Haendeler J, Zeiher AM, Dimmeler S (2005) Cell-to-cell connection of endothelial progenitor cells with cardiac myocytes by nanotubes: A novel mechanism for cell fate changes? Circ Res 96:1039–1041.
9. Lucas WJ, Ham BK, Kim JY (2009) Plasmodesmata - Bridging the gap between neighboring plant cells. Trends Cell Biol 19:495–503.
10. Molnar A, et al. (2010) Small silencing RNAs in plants are mobile and direct epigenetic modification in recipient cells. Science 328:872–875.
11. Ehlers K, Kollmann R (2001) Primary and secondary plasmodesmata: Structure, origin, and functioning. Protoplasma 216:1–30.
12. Gillman JD, Bentolila S, Hanson MR (2009) Cytoplasmic male sterility in Petunia. Petunia, eds Gerats T, Strommer J (Springer, New York), pp 107–129.
13. Moon HS, Nicholson JS, Lewis RS (2008) Use of transferable *Nicotiana tabacum* L. microsatellite markers for investigating genetic diversity in the genus *Nicotiana*. Genome 51:547–559.
14. Sheahan MB, McCurdy DW, Rose RJ (2005) Mitochondria as a connected population: Ensuring continuity of the mitochondrial genome during plant cell dedifferentiation through massive mitochondrial fusion. Plant J 44:744–755.
15. Boeshore ML, Hanson MR, Izhar S (1985) A variant mitochondrial DNA arrangement specific to Petunia stable sterile somatic hybrids. Plant Mol Biol 4:125–132.
16. Svab Z, Maliga P (2007) Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. Proc Natl Acad Sci USA 104:7003–7008.
17. Ruf S, Karcher D, Bock R (2007) Determining the transgene containment level provided by chloroplast transformation. Proc Natl Acad Sci USA 104:6998–7002.
18. Huang CY, Ayliffe MA, Timmis JN (2003) Direct measurement of the transfer rate of

chloroplast DNA into the nucleus. Nature 422:72–76.
19. Stegemann S, Hartmann S, Ruf S, Bock R (2003) High-frequency gene transfer from the chloroplast genome to the nucleus. Proc Natl Acad Sci USA 100:8828–8833.
20. Sheppard AE, et al. (2008) Transfer of plastid DNA to the nucleus is elevated during male gametogenesis in tobacco. Plant Physiol 148:328–336.
21. Hanson MR, Sattarzadeh A (2011) Stromules: Recent insights into a long neglected feature of plastid morphology and function. Plant Physiol 155:1486–1492.
22. Newell CA, et al. (2012) Exclusion of plastid nucleoids and ribosomes from stromules in tobacco and Arabidopsis. Plant J 69:399–410.
23. Zubko MK, Day A (1998) Stable albinism induced without mutagenesis: A model for ribosome-free plastid inheritance. Plant J 15:265–271.
24. Palmer JD (1983) Chloroplast DNA exists in two orientations. Nature 301:92–93.
25. Medgyesy P, Fejes E, Maliga P (1985) Interspecific chloroplast recombination in a *Nicotiana* somatic hybrid. Proc Natl Acad Sci USA 82:6960–6964.
26. Fejes E, Engler D, Maliga P (1990) Extensive homologous chloroplast DNA recombination in the pt14 *Nicotiana* somatic hybrid. Theor Appl Genet 79:28–32.
27. Burch-Smith TM, Stonebloom S, Xu M, Zambryski PC (2011) Plasmodesmata during development: Re-examination of the importance of primary, secondary, and branched plasmodesmata structure versus function. Protoplasma 248:61–74.
28. Benitez-Alfonso Y, et al. (2009) Control of *Arabidopsis* meristem development by thioredoxin-dependent regulation of intercellular transport. Proc Natl Acad Sci USA 106:3615–3620.
29. Stonebloom S, et al. (2009) Loss of the plant DEAD-box protein ISE1 leads to defective mitochondria and increased cell-to-cell transport via plasmodesmata. Proc Natl Acad Sci USA 106:17229–17234.
30. Keeling PJ, Palmer JD (2008) Horizontal gene transfer in eukaryotic evolution. Nat Rev Genet 9:605–618.
31. Bock R (2010) The give-and-take of DNA: Horizontal gene transfer in plants. Trends Plant Sci 15:11–22.
32. Baldev A, et al. (1998) Recombination between chloroplast genomes of *Trachystoma ballii* and *Brassica juncea* following protoplast fusion. Mol Gen Genet 260:357–361.
33. Than ND, Medgyesy P (1989) Limited chloroplast gene transfer via recombination overcomes plastome-genome incompatibility between *Nicotiana tabacum* and *Solanum tuberosum*. Plant Mol Biol 12:87–93.
34. Bock R, Kössel H, Maliga P (1994) Introduction of a heterologous editing site into the tobacco plastid genome: The lack of RNA editing leads to a mutant phenotype. EMBO J 13:4623–4628.
35. Schmitz-Linneweber C, et al. (2005) Pigment deficiency in nightshade/tobacco cybrids is caused by the failure to edit the plastid ATPase alpha-subunit mRNA. Plant Cell 17: 1815–1828.
36. Maliga P, Bock R (2011) Plastid biotechnology: Food, fuel, and medicine for the 21st century. Plant Physiol 155:1501–1510.
37. Cardi T, Lenzi P, Maliga P (2010) Chloroplasts as expression platforms for plant-produced vaccines. Expert Rev Vaccines 9:893–911.
38. Carrer H, Staub JM, Maliga P (1990) Gentamycin resistance in *Nicotiana* conferred

by AAC(3)-I, a narrow substrate specificity acetyl transferase. Plant Mol Biol 17:301-303.
39. Maliga P, Svab Z (2011) Engineering the plastid genome of *Nicotiana sylvestris*, a diploid model species for plastid genetics. Plant Chromosome Engineering: Methods and Protocols, ed Birchler JJ (Springer, New York), Vol 701, pp 37–50.
40. Lutz KA, Svab Z, Maliga P (2006) Construction of marker-free transplastomic tobacco using the Cre-loxP site-specific recombination system. Nat Protoc 1:900–910.
41. Murray MG, Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321–4325.
42. Cronn R, et al. (2008) Multiplex sequencing of plant chloroplast genomes using Solexa sequencing-by-synthesis technology. Nucleic Acids Res 36:e122.

CHAPTER TWO

Exceptional Inheritance of Plastids via Pollen in *Nicotiana sylvestris* is Independent from Mitochondrial Transmission

Abstract

Plastids and mitochondria, the DNA-containing cytoplasmic organelles, are maternally inherited in the majority of angiosperm species. Even in plants with strict maternal inheritance, exceptional paternal transmission of plastids has been observed. Our objective was to determine if rare leakage of plastids via pollen in *Nicotiana sylvestris* results in cotransmission of paternal mitochondria. As fathers, we used *N. sylvestris* plants with transgenic, selectable plastids and wild-type mitochondria. As mother plants, we used *N. sylvestris* plants with the *N. undulata* cytoplasm, including the CMS-92 mitochondria that cause cytoplasmic male sterility (CMS) by homeotic transformation of the stamens. We report here that cytoplasmic inheritance is maternal in the diploid tobacco, *N. sylvestris*. We detected exceptional paternal plastids (ptDNA) in ~0.002% of *N. sylvestris* seedlings. However, we did not detect paternal mitochondria (mtDNA) in any of the six plastid-transmission lines, suggesting independent transmission of the cytoplasmic organelles via pollen. We also obtained fertile plastid transmission lines, which we found to be no more likely to transmit their plastids via pollen than their fathers. We discuss implications for transgene containment and plant evolutionary histories inferred from cytoplasmic phylogenies.

Introduction

Nicotiana tabacum has been widely used as a model system in plant biology because of the ease of regeneration from cells or tissue and facile transformation of nuclear and plastid genomes (1, 2). However, homeologous genes in the allotetrapoloid *N. tabacum* nucleus present an obstacle for mutant screening. Recently, the diploid *Nicotiana sylvestris* is gaining acceptance as a replacement for *N. tabacum* in studies on developmental and plastid biology (3-5). Exceptional transmission of cytoplasmic organelles has been studied in *N. tabacum* (6-9), and in this study we establish *N. sylvestris* as a model for the dissection of cytoplasmic inheritance.

In 1909 non-Mendelian maternal inheritance was observed for pigment deficiency in plants, revealing that genetic information resides in the cytoplasm as well as the nucleus (Review see (10)). Plastids and mitochondria, the DNA-containing cytoplasmic organelles, are maternally inherited in the majority (80%) of angiosperms (11) and, in *Nicotiana*, carry 112 (12) and 60 (13) genes, respectively. The exclusion of plastids from the generative cell in pollen accounts for maternal inheritance in plants with Lycoperisicon-type plastid inheritance (14). Therefore, transgenes that are incorporated in the plastid genome should not normally be transmitted by pollen, providing a natural strategy for biocontainment (8, 9). However, low-level leakage of plastids in pollen has been reported in several species (6-9, 15-17).

We sought to characterize the parental contributions of mitochondria to *N. sylvestris* plants with exceptional paternal plastids, so we used pollen donors with selectable antibiotic resistance transgenes in their plastid genomes (Fig. 2-1). The presence of alien (*N. undulata*) plastids and mitochondria in the *N. sylvestris* mother

plants ensured that there would be abundant molecular markers to distinguish the genomes of the cytoplasmic organelles. Here we report that ~0.002% of *N. sylvestris* F1 seedlings inherit plastids via pollen and such plants contain maternal mitochondria.

Results

Experimental Design

To screen for rare transmission of paternal plastids via pollen, we chose two selectable transplastomic lines to serve as fathers in the cross. The Ns-RB8 (4) plant carries the *aadA* gene for streptomycin and spectinomycin resistance in the large single copy region of the plastid genome (Fig 2-2A), while Ns-MSK56 (4) carries an *aadA-gfp* fusion marker gene in the inverted repeat region (Fig 2-2B). The mother plants contained the *N. sylvestris* nuclear genome but plastids and mitochondria from *Nicotiana undulata* (18). The CMS-92 *N. undulata* mitochondria cause cytoplasmic male sterility (CMS) in the Ns-CMS92 plants by homeotic transformation of stamens (Fig 2-1A). Seeds were collected from Ns-CMS92 plants that had been pollinated by Ns-RB8 or Ns-MSK56. Seeds were surface sterilized and germinated on media that contained spectinomycin (Fig. 2-1B,C). We investigated the effect of including callus-inducing hormones in the RMOP media on our ability to recover plants with paternal plastids (Table 2-1). Seedlings with paternal plastids had green sectors on the spectinomycin containing media (Fig 2-1B). These sectors did not always contribute to the shoot meristem (Fig 2-1B). When such plants emerged on callus inducing RMOP media (8), plant lines could be recovered, while the seeds that were germinated on RM media (8) faced the additional selection that the paternal plastids had to contribute significantly to the shoot meristem to

enable growth (Fig 2-1C, Table 2-1). We also recovered spontaneous spectinomycin-resistant but streptomycin-sensitive seedlings on RMOP in both maternal lineages, as expected. We identified four NsSpc plants in 89,636 Ns-CMS92 seedlings, and four more NsSpc plants in 123,523 seedlings with the *N. sylvestris* cytoplasm.

Exceptional transmission of plastids via pollen

The streptomycin and spectinomycin resistance of the NsPSpc seedlings indicated that they carried the paternal *aadA* plastid transgene. We designed primers to flank seven evenly spaced length polymorphisms that distinguish *N. sylvestris* (AB237912) from *N. undulata* ptDNA (JN563929)(Fig 2-1, Table 2-S1). Two additional primer pairs were designed to flank the transgene integration sites (Fig 2-2, Table 2-S1). Amplifying these fragments from the father, mother, and resistant progeny (NsPSpc plants) revealed that the entire paternal ptDNA was present in each of the NsPSpc plant lines (Fig. 2-2C). Furthermore, each of the NsPSpc plants was homoplastomic, carrying a uniform population of paternal plastids (Fig. 2-2C).

Assessment of mitochondria in NsPSpc plants

We sought to clarify the contribution of mtDNA to the NsPSpc plants by amplifying and Sanger sequencing ten polymorphic loci that distinguish the *N. sylvestris* from the *N. undulata* CMS-92 mitochondria. Six of these markers had been identified previously (18) and the additional four were found by sequencing amplicons from the CMS-92 mitochondria. In each of the six NsPSpc lines, all ten markers indicated the presence of

maternal mtDNA (Fig. 2-3). Therefore, in *N. sylvestris*, leakage of plastids via pollen can create plant lines with unrelated plastids and mitochondria.

Testing for elevated pollen transmission in NsPspc plants

We considered that some plastid transmission lines could be the product of a late somatic mutation that increases the likelihood of plastid incorporation in sperm cells. Therefore, we repeated the screen with a fertile *N. sylvestris* TW137 mother and Ns-RB8 father (Ns-7834) and in 123,523 seedlings, obtained 8 paternal plastid transmission events. We tested five of these NsPSpc plants for an elevated level of paternal plastid transmission. Such a mutation would likely operate on a gametophytic level and be observable in the F1 plants because the critical steps of pollen mitosis occur in haploid cells (14, 19). We did not observe elevated levels of paternal inheritance in the five lines of NsPSpc plants (Table 2-2).

Discussion

Plants with exceptional paternal plastids contain maternal mitochondria

Plastids in *Nicotiana* are normally excluded from the generative cells that give rise to plant sperm cells, preventing paternal transmission (14, 20). However, rare leakage of plastids via pollen seems universal in plants that generally display strict maternal inheritance of the cytoplasmic organelles (16). We found a rate of plastid leakage in *N. sylvestris* that is comparable to that found in *N. tabacum* (~0.001%) (8, 9). However, we did not find paternal mtDNA in these plants. Earlier reports in *N. tabacum* have conflicted on the inheritance of mtDNA in plants that acquired paternal plastids.

However, these were based on Southern probing of a single region (7, 8). By identifying and scoring ten polymorphisms that are well distributed around the mtDNA master circle (Fig 2-3A), we were able to conclude that the NsPSpc plants contained the entire maternal mtDNA. Since the mitochondria were not selectable, we cannot exclude limited transfer of paternal mtDNA that was rapidly lost through sorting-out. However, the propensity of mitochondria to fuse and recombine suggests that we could have observed recombinant mitochondria in our NsPSpc plants if co-transfer of plastids and mitochondria were the rule.

Underlying mechanism of exceptional plastid inheritance

The basis of exceptional inheritance of plastids could be nuclear control, as nuclear genes have been implicated in the proper development of pollen from haploid gametophytic tissue (19). Levels of exceptional inheritance vary between genotypes in *Petunia* (21, 22) suggesting genetic control. We did not observe elevated levels of pollen transmission of plastids in any of the NsPspc lines that had inherited plastids from their fathers (Table 2-2). If a late somatic mutation gave rise to gametophytic tissue that failed to completely exclude plastids, we would expect half of the pollen in each NsPSpc F1 line to carry the same mutation. Since none of the tested lines exhibited detectibly higher levels of paternal transmission, we conclude that the transmission of plastids via pollen that we observed in *N. sylvestris* was due to a chance or stochastic event, rather than a heritable spontaneous mutation. The level of containment afforded by plastid localization is therefore so high that even if a non-transgenic field were completely pollenated by a transplastomic parent, the seed would be far from reaching the 0.9% transgenic threshold

that triggers labeling in Europe (23). Furthermore, while half of the pollen of a hybrid plant with an escaped nuclear transgene can carry the transgene further, we found that strict maternal inheritance is maintained for chloroplasts in *N. sylvestris* plants that acquired exceptional plastids via pollen.

Implications for the interpretation of plant evolutionary histories

Molecular phylogenies based on plastid and nuclear markers are occasionally incongruous (24), suggesting the existence of chloroplast capture (25), a mechanism for the horizontal transfer of entire chloroplast genomes (26). Our finding that plastids can be transmitted via pollen, independently of mitochondria in *N. sylvestris*, suggests the potential for further phylogenic inconsistencies that would become apparent when comparing mtDNA-derived trees with those based on ptDNA.

Materials and Methods

The mother plants were wild-type *N. sylvestris* cv. TW137 or Ns-CMS92, an *N. sylvestris* TW137 plant with plastids and mitochondria from *N. undulata* (18). The father plants were Ns-RB8 or Ns-MSK56, which carry *aadA* transgenes for streptomycin and spectinomycin resistance in their *N. sylvestris* plastid genomes, and have *N. sylvestris* nuclei and mitochondria (4). Hybrid seed were collected, surface sterilized, and sewn on selective RM or RMOP media (Table 2-1) (8). Green, streptomycin-resistant shoots or calli were isolated and subcultured on selective RMOP to generate the NsPSpc plants. Total cellular DNA was isolated from greenhouse grown leaf tissue (8). Plastid DNA markers were amplicons spanning indels that distinguish *N. sylvestris* from *N. undulata*

ptDNA (Table 2-S1). These amplicons were visualized with 5.5% polyacrylamide gel electrophoresis. Amplicons spanning transgene integration sites were visualized in 1.0% agarose. Mitochondrial DNA markers were amplicons that were Sanger sequenced to reveal SNPs or short indels that identify the *N. undulata* CMS92 mtDNA (Table 2-S2).

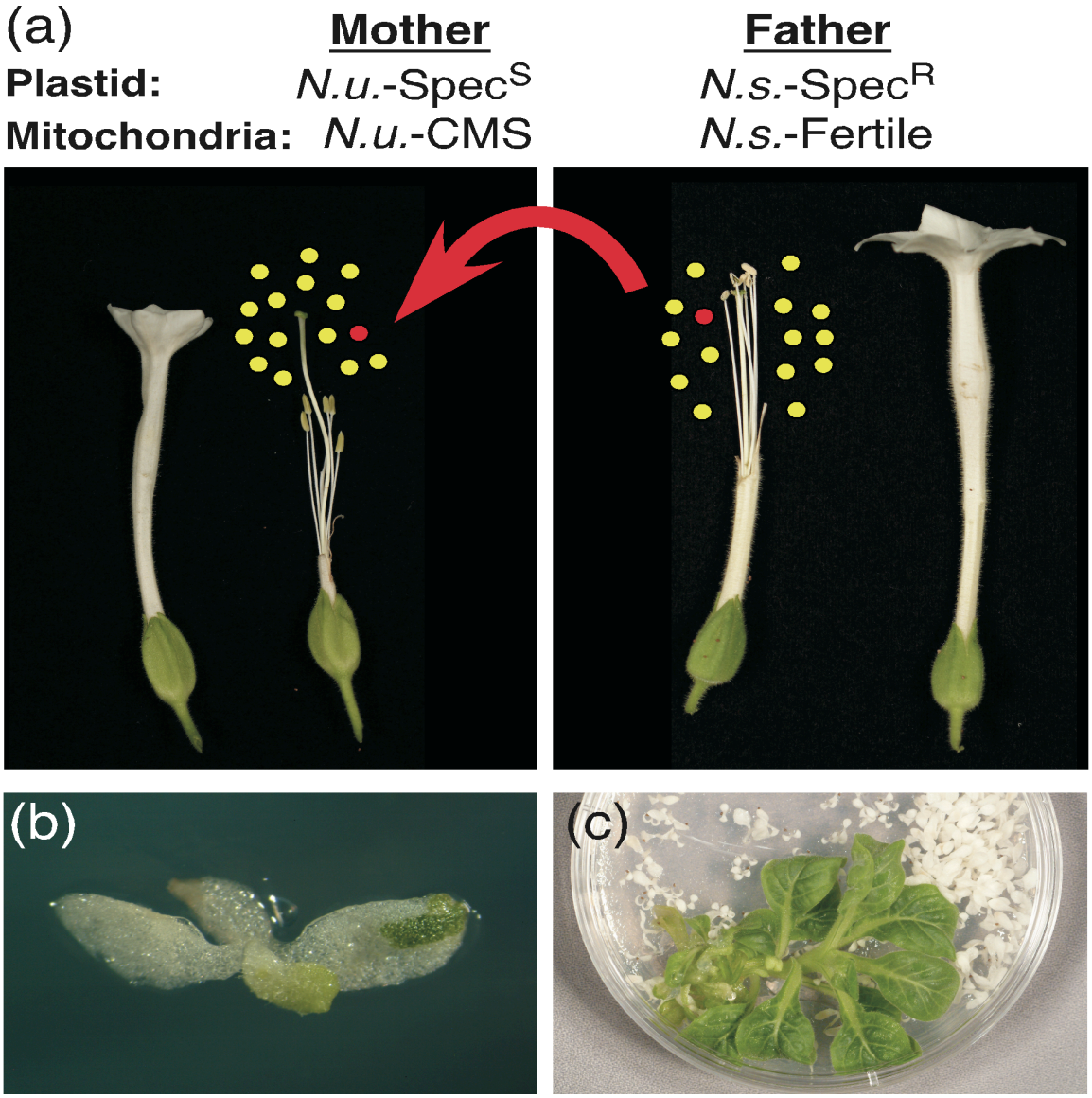


FIGURE 2-1. Detection of exceptional paternal cytoplasmic inheritance in *Nicotiana sylvestris*. (a) Crosses to screen for NsPSpc plants with paternal plastids. Genotypes of mother (Ns-CMS92) and father plants (Ns-RB8 and Ns-MSK56) are shown. (b) Identification of a spectinomycin resistant NsPSpc seedling with paternal plastids confined to a sector that will not normally contribute to the meristem. (c) Identification of an NsPSpc seedling in which paternal plastids have contributed significantly to the meristem, permitting normal growth on selective media, while siblings with maternal ptDNA are bleached.

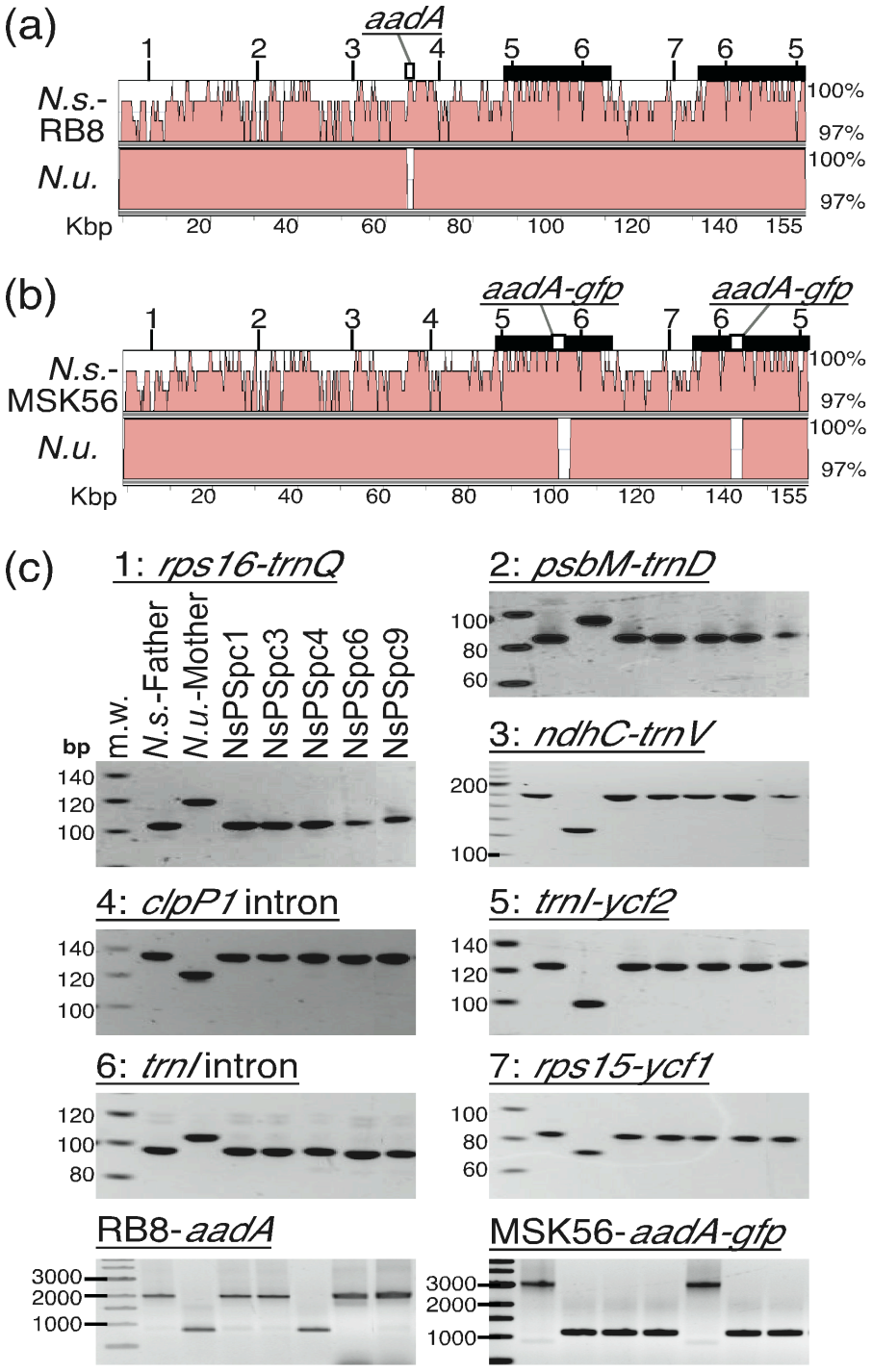
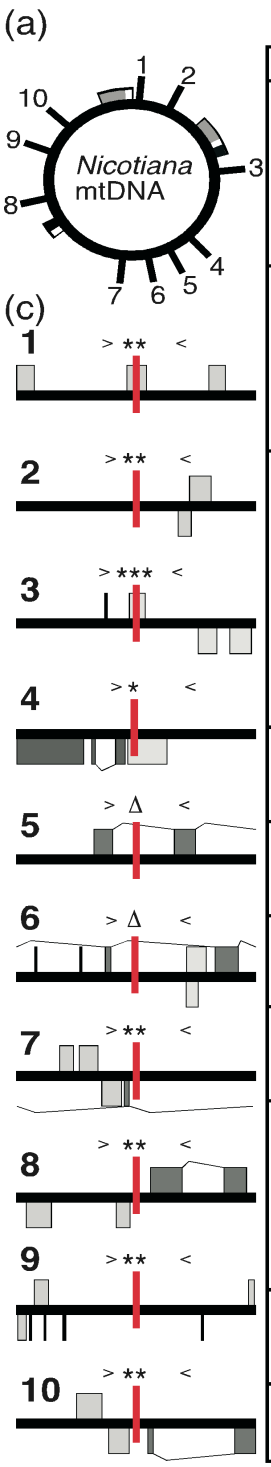


FIGURE 2-2. Plastid DNA markers in NsPSpc plant lines. Alignment between the maternal *N. undulata* plastid genome (JN563929) and (a) Ns-RB8 or (b) Ns-MSK56 plastid genomes indicates the location of the *aadA* transgene and seven length polymorphisms (1-7). (c) Length polymorphisms of amplified ptDNA fragments from father, mother, and NsPSpc plants. For primer sequences see Table S1.



(b)

Locus	Mitochondrial Sequence
orf125a 1017	*N.s.* aaaaccaat **A** tcataacttt *N.u.* aaaaccaat **C** tcataacttt PSpc aaaaccaat **C** tcataacttt
1068	*N.s.* gaaacgaag **T** ggagcgaggg *N.u.* gaaacgaag **C** ggagcgaggg PSpc gaaacgaag **C** ggagcgaggg
atp6-orf110a 30671	*N.t.* ttagacagg **T** tgactccttt *N.u.* ttagacagg **G** tgactccttt PSpc ttagacagg **G** tgactccttt
30744	*N.s.* gtatcgaac **G** agcccctatc *N.u.* gtatcgaac **A** agcccctatc PSpc gtatcgaac **A** agcccctatc
orf129b 100408	*N.s.* aaaaaattt **A** tctcttttta *N.u.* aaaaaattt **C** tctcttttta PSpc aaaaaattt **C** tctcttttta
100521	*N.s.* ctttgttct **G** tagtctttca *N.u.* ctttgttct **C** tagtctttca PSpc ctttgttct **C** tagtctttca
100811	*N.t.* ctcccattt **G** aaagattatt *N.u.* ctcccattt **T** aaagattatt PSpc ctcccattt **T** aaagattatt
orf315 131766	*N.s.* gggtctcct **C** tttctcgtcc *N.u.* gggtctcct **A** tttctcgtcc PSpc gggtctcct **A** tttctcgtcc
nad4 intron 183622	*N.s.* tgcac **CTCCGTACAA** gtgct *N.u.* tgcac ---------- gtgct PSpc tgcac ---------- gtgct
nad2 intron 202068	*N.s.* ttccg **GCTGTTTCCG** tcatt *N.u.* ttccg ---------- tcatt PSpc ttccg ---------- tcatt
nad5 intron 222498	*N.s.* aagaaaaaa **GA** aaagtcgag *N.u.* aagaaaaaa **TC** aaagtcgag PSpc aagaaaaaa **TC** aaagtcgag
orf115-ccmFc 306929	*N.s.* ggatcagac **T** actcctggtg *N.u.* ggatcagac **C** actcctggtg PSpc ggatcagac **C** actcctggtg
307041	*N.s.* agcaaaact **C** gaacggatag *N.u.* agcaaaact **A** gaacggatag PSpc agcaaaact **A** gaacggatag
trnS-trnE 328122	*N.s.* aaaactttt **TC** aaaagtatt *N.u.* aaaactttt **GA** aaaagtatt PSpc aaaactttt **GA** aaaagtatt
orf159b-nad7 361016	*N.s.* agccagctg **CT** tgcccccttt *N.u.* agccagctg **AC** tgcccccttt PSpc agccagctg **AC** tgcccccttt

FIGURE 2-3. Mitochondrial DNA markers in NsPSpc plants lines. (a) The *Nicotiana* mtDNA master circle showing the location of the ten selected polymorphisms. The black, white and grey boxes mark the repeat regions. (b) Alignment of mitochondrial sequence for paternal (N.s.), maternal (N.u.) and six NsPSpc plant lines (PSpc). (c) Diagram of mtDNA marker loci showing gene features and primer locations. For primer sequences see Table 2-S2.

TABLE 2-1. Paternal ptDNA transmission in *N. sylvestris* detected by selection for spectinomycin resistance.

Pollen donor	**Protocol**	**No. of seedlings**	**No. and % of ptDNA transfer events**	
Ns-MSK56	RM	63,268	0	0.00000%
Ns-MSK56	RMOP	31,090	1	0.00322%
Ns-RB8	RM	94,619	3	0.00317%
Ns-RB8	RMOP	58,546	2	0.00342%
Total		247,523	6	0.00242%

TABLE 2-2. Paternal ptDNA transmission in NsPSpc lines detected by selection for spectinomycin resistance.

Pollen donor	No. of seedlings	No. green seedlings on RMOP-Spec500
NsPSpc11	~5,600	0
NsPSpc12	~1,460	0
NsPSpc18	350	0
NsPSpc19	1,125	0
NsPSpc21	~4,600	0

TABLE 2-S1. Primers flanking length polymorphisms that distinguish *Nicotiana undulata* ptDNA (JN563929) from *Nicotiana sylvestris* (AB237912) and confirm transgene integration

Primer	Locus	Gene	Sequence
6pLF	6273	rps16-trnQ	TGGTTGGGCTGATGTATAAACACCA
6pLR	6375		AGCGATGGGGTCTTACTAAAGAAA
31pLF	31438	psbM-	TCATTCCCCTTTCTAAGAGGAGTAGGATCT
31pLR	31522	- trnD	TTTTGGTATAGGTGTCCCGGGGCT
53pLF	53279	ndhC-trnV	TGTGCTTCGCTAGGTCGAGGTAAGT
53pLR	53462		CAAATCATTTCACGGGCCTGGTGA
74pLF	74199	clpP1	TCTAAACGGAGCCTGGATACTTCA
74pLR	74332	intron	TGGAGCGTGAAGTGCAATTAGATCCA
88pLF	88774	trnI-ycf2	TAGCGGGGATCCTCGTACATGGTG
88pLR	88895		TGTCCTCTCATTGATTCCTCCTAAATTGC
104pLF	104880	trnI intron	TGGACAGCTATCTCTCGAGCACAGG
104pLR	104976		GGGGCGATCTCGTAGTTCTTGGTCT
125pLF	125501	rps15-ycf1	TTTCCCCTTTCTTTATTTTACAGATATGGA
125pLR	125583		ACCACGTTCAAATTACTGGCATTA
65.8F	65836	RB8-aadA	CCCTTTACCTTACCCCCACCCCC
66R	66664		TACTCATTTTTGTACTTGCTGT
102nutF	102261	MSK56-	AACTCCAGTTCCTTCGGAATCGGT
103nutR	103394	-aadA-gfp	CACCGGAAATTCCCTCTGCCCCTA

TABLE 2-S2. Primers to distinguish *Nicotiana undulata* CMS92 mitochondrial DNA (mtDNA) from *Nicotiana tabacum/sylvestris* (BA000042) by Sanger sequencing

Primer	Locus	Sequence
0muF	670	AGAAGCTGTGATCGAGGAAGCCCC
1muR	1963	GCTCTGAAGGGAGAGTTGAGCGGA
30muF	30429	TTGCCCGAGTAGGGGAAGGGATTG
31muR	31810	TCGTTTCGGGCCGATGAAGTACCT
100muF	100029	GCTTCGATGATCAACCCCTGGCAC
101muR	101468	CCAAATACAAGGGAGCGGGCACTG
131muF	131632	ATCAGAAGCATCCAGCAGCACCAC
132muR	132932	GCTCTGCTGCATGACGGAGTGATC
183muF	183262	ACCCGACCAGGGATGGACGTAAAC
184muR	184522	AGGTGCCTCTACATGAGCTTCGGG
201muF	201780	CGCCTGGAAGTCCGAGGACCTTTA
203muR	203065	CTCCGAAAGCGTTTTCCTTCCCCC
222muF	222229	CGTACGTGGAGCTTCCGCCTCATA
223muR	223505	GGGCCTGCCCTTTTGCTAGCTTTT
306muF	306541	TGTATCACCGAGACACCCGAAGGG
307muR	307907	CGGATCGAATCAGAGTTCACGCCG
327muF	327868	AGTTGCTCTTTGCCCAAAGCCCTC
329muR	329117	TGTTAGGCATTGAACCCCACCCCA
360muF	360645	GCCATTGGTTACTGGTTGAGCCAC
361muR	361983	GATGTCGTGACCGCTTAGGCTTGG

REFERENCES FOR CHAPTER TWO

1. Hajdukiewicz P, Svab Z, & Maliga P (1994) The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. Plant Mol Biol 25, 989-994.
2. Svab Z & Maliga P (1993) High-frequency plastid transformation in tobacco by selection for a chimeric *aadA* gene. Proc Natl Acad Sci U S A 90, 913-917.
3. McHale NA & Koning RE (2004) MicroRNA-directed cleavage of *Nicotiana sylvestris* PHAVOLUTA mRNA regulates the vascular cambium and structure of apical meristems. Plant Cell 16, 1730-1740.
4. Maliga P & Svab Z (2011) Engineering the plastid genome of *Nicotiana sylvestris*, a diploid model species for plastid genetics. Methods Mol Biol 701, 37-50.
5. Pineau B, Mathieu C, Gerard-Hirne C, De Paepe R, & Chetrit P (2005) Targeting the NAD7 subunit to mitochondria restores a functional complex I and a wild type phenotype in the *Nicotiana sylvestris* CMS II mutant lacking *nad7*. J Biol Chem 280, 25994-26001.
6. Medgyesy P, Páy A, & Márton L (1986) Transmission of paternal chloroplasts in *Nicotiana*. Molecular and General Genetics 204, 195-198.
7. Avni A & Edelman M (1991) Direct selection for paternal inheritance of chloroplasts in sexual progeny of *Nicotiana*. Mol Gen Genet 225, 273-277.
8. Svab Z & Maliga P (2007) Exceptional transmission of plastids and mitochondria from the transplastomic pollen parent and its impact on transgene containment. Proc Natl Acad Sci U S A 104, 7003-7008.
9. Ruf S, Karcher D, & Bock R (2007) Determining the transgene containment level provided by chloroplast transformation. Proc Natl Acad Sci U S A 104, 6998-7002.
10. Hagemann R (2009) The foundation of extranuclear inheritance: plastid and mitochondrial genetics. Mol Genet Genomics 283, 199-209.
11. Zhang Q & Sodmergen (2007) Why does biparental plastid inheritance revive in angiosperms? J Plant Res 123, 201-206.
12. Shimada H & Sugiura M (1991) Fine structural features of the chloroplast genome: comparison of the sequenced chloroplast genomes. Nucleic Acids Res 19, 983-995.
13. Sugiyama Y, Watase Y, Nagase M, Makita N, Yagura S, Hirai A, & Sugiura M (2005) The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Mol Genet Genomics 272, 603-615.
14. Birky CW, Jr. (1995) Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc Natl Acad Sci U S A 92, 11331-11338.
15. Wang T, Li Y, Shi Y, Reboud X, Darmency H, & Gressel J (2004) Low frequency transmission of a plastid-encoded trait in *Setaria italica*. Theor Appl Genet 108, 315-320.
16. Azhagiri AK & Maliga P (2007) Exceptional paternal inheritance of plastids in *Arabidopsis* suggests that low-frequency leakage of plastids via pollen may be universal in plants. Plant J 52, 817-823.

17. Ellis JR, Bentley KE, & McCauley DE (2008) Detection of rare paternal chloroplast inheritance in controlled crosses of the endangered sunflower *Helianthus verticillatus*. Heredity (Edinb) 100, 574-580.
18. Thyssen G, Svab Z, & Maliga P (2012) Cell-to-cell movement of plastids in plants. Proc Natl Acad Sci U S A 109, 2439-2443.
19. Berger F & Twell D (2011) Germline specification and function in plants. Annu Rev Plant Biol 62, 461-484.
20. Russell SD (1994) Populations of plastids and mitochondria during male reproductive cell maturation in *Nicotiana tabacum* L.: A cytological basis for occasional biparental inheritance. Planta 03, 115-122.
21. Derepas A & Dulieu H (1992) Inheritance of the Capacity to Transfer Plastids by the Pollen Parent in *Petunia hybrida* Hort. Journal of Heredity 83, 6-10.
22. Cornu A & Dulieu H (1988) Pollen Transmission of Plastid DNA Under Genotypic Control in *Petunia hybrida* Hort. Journal of Heredity 79, 40-44.
23. Gruere G & Rao S (2007) A review of international labeling policies of genetically modified food to evaluate India's proposed rule. AgBioForum 10, 51-64.
24. Rieseberg L & Soltis D (1991) Phylogenetic consequences of cytoplasmic gene flow in plants. Evolutionary Trends in Plants 5, 65-84.
25. Tsitrone A, Kirkpatrick M, & Levin DA (2003) A model for chloroplast capture. Evolution 57, 1776-1782.
26. Stegemann S, Keuthe M, Greiner S, & Bock R (2012) Horizontal transfer of chloroplast genomes between plant species. Proc Natl Acad Sci U S A 109, 2434-2438.

CHAPTER THREE

Mix-and-Paste Replication of High-Copy TNT1 LTR-Retrotransposons

Abstract

Our objective was to experimentally determine if related high-copy elements interact during retrotransposition. To enable detection of template-switching events, we introduced silent mutations to the *gag-pol* gene of our synthetic 5334-bp S-TNT1, including an 831-bp codon-shuffled 'patch' that could be used to distinguish S-TNT1 from the hundreds of endogenous TNT1-like elements in the *Nicotiana sylvestris* genome. S-TNT1 was introduced on a T-DNA linked to a kanamycin resistance gene. We generated 101 kanamycin resistant plants that contained the T-DNA copy of S-TNT1. Of these, 47 plants contained one or more S-TNT1 transpositions. In three of four insertion events, S-TNT1 interacted with native N-TNT1 elements to produce recombinant insertions. The distinct LTRs and evidence of multiple recombination points within the *gag-pol* gene revealed a mixed history. One non-collinear, but micro-homology-directed recombination produced a 389-bp deletion and a "dead-on-arrival" nonautonomous element. Surprisingly, two of the new insertions had non-identical LTRs, with the 5'LTR shorter than the 3'LTR. We can best explain our observations by dimerization and co- packaging of N-TNT1 and S-TNT1 gRNA in the cytoplasm, followed by template- switching during synthesis of a single hybrid minus-strand DNA, with occasional homology-directed slippage during plus-strand strong-stop DNA synthesis, which we term the "mix-and-paste" pseudodiploid mating system for LTR-retroelements. Identification of one essentially unchanged S-TNT1 insertion indicates

that homodimerization is also occurring and suggests that our S-TNT1 transcript made up a significant fraction of the transposition competent TNT1 gRNA mating population. Our study indicates that retrotransposon mating is a major source of diversification of high- copy retrotransposons.

Introduction

Long terminal repeat (LTR) retrotransposons are among the repetitive and self-replicating sequences termed "selfish DNA" that make up the majority of eukaryotic chromosomes (1, 2). Variation in plant genome size is largely a function of repetitive DNA content: the compact 125-Mb *Arabidopsis* nuclear genome is only 14% (3), while the 2.3-Gb *Zea mays* nuclear genome is 85% transposable elements (4). In *Arabidopsis* the genome contains only 1594 LTR-retrotransposons (3) while the 406 families in *Zea mays* make up 75% of the genome (4). The remarkable variation between modern maize haplotypes was produced in part by an explosion of retrotransposition (5). Plant retrotransposons can be activated by polyploidization (6, 7), microbial attack (8) and tissue culture (9, 10). Retrotransposons accumulate in host genomes due to their canonical "copy-and-paste" life cycle, beginning as DNA in the host genome that is transcribed to mRNA, which is then reverse-transcribed via the self-encoded *gag-pol* protein apparatus to generate the progeny DNA insertion. Retrotransposons are present as large families with variation in both LTR and coding sequences (11). Historic insertions are dated by comparing the LTRs of a retroelement and assuming that all substitutions have accumulated randomly since integration (12).

Retrotransposon insertions are purged from the host genome over evolutionary time

by direct repeat-mediated deletion, producing solo LTRs and counterbalancing the genomic expansion produced by retrotransposition (13). Direct-repeat mediated deletion between adjacent integrated LTR-retrotransposons has been invoked to explain apparently recombinant retrotransposons in the *Zea mays* genome (14). Recombination between adjacent insertions in the host genome was proposed to generate progenitor hybrid retroelements that would then undergo a burst of retrotransposition (14). In contrast, apparently recombinant retrotransposons in soybean chromosomal sequence were explained by template switching during reverse-transcription rather than by host genome rearrangement (15). Template switching was also used to explain LTR-swapping for a synthetic mini-TNT1 that lacked the *gag-pol* gene (16). Template switching is consistent with what is known about replication of low-copy retrotransposons and animal retroviruses. The low-copy Ty1 LTR-retrotransposon of yeast was shown to package two copies of its transcript in a virus-like particle where template switching during reverse transcription could produce recombinant Ty1 progeny (17, 18). Frequent recombination during reverse-transcription is also known for co-packaged animal retroviruses (19), in which a pseudodiploid mating system is recognized (20). However, the plant genome sequence-based studies could not conclusively identify the lineage of an individual high-copy retrotransposon, because the specific parental retroelements may no longer exist or may have numerous indistinguishable relatives. Our objective was to experimentally determine if high-copy endogenous LTR-retrotransposons undergo a simple "copy-and-paste" life cycle or one that includes co-expression, co-packaging and recombination. We designed a synthetic, full length S-TNT1 element based on the originally described TNT1 LTR-retrotransposon from

tobacco (21). We introduced silent mutations to the *gag-pol* gene, including an 831-bp 'patch' of silent mutations in the RNaseH region that reduced homology with the wild-type sequence by 34.8%, so that we could distinguish S-TNT1 from the hundreds of native N-TNT1 elements that are endogenous to *Nicotiana sylvestris*, a diploid tobacco species (22). S-TNT1 was introduced to *N. sylvestris* by selection for a kanamycin resistance gene encoded on the same T-DNA. We report here that some of the kanamycin resistant plants contained one or more transposed copies of S-TNT1, in addition to the T-DNA encoded copy. We characterized four S-TNT1 insertions in detail. Three of these contained swapped LTRs and recombinant *gag-pol* coding sequences. The fourth insertion was nearly identical to the progenitor element, suggesting that the S-TNT1 transcript was sufficiently abundant to homodimerize or that template switching may not always occur in the pseudodiploid state. Our study indicates that mating is a major source of diversification of high-copy retrotransposons, producing hybrid retroelements with mixed histories. Furthermore, co-packaging and template switching suggests a mechanism for retroelement acquisition of genes.

Results

The S-TNT1 Retrotransposon.

A full length 5334-bp S-TNT1 retrotransposon was designed based on the TNT1 retrotransposon sequence from *Nicotiana tabacum* (X13777), including the flanking 100 bp of the *nia2* gene insertion site and 'GAAGT' target site duplication (21). Silent mutations were introduced to the TNT1 coding region to add and remove several common endonuclease-binding sites (Fig. 3-S1). Also, we re-ordered synonymous

codons in a 831-bp stretch of the *gag-pol* gene to reduce homology in this region by 34.8% without changing the length, amino acid sequence, or overall codon usage. Our S-TNT1 would therefore be classified as an autonomous element by standard nucleotide sequence analysis because it contains an uninterrupted open reading frame encoding a wild-type *gag-pol* protein. The S-TNT1 element was cloned into a pPZP212 binary vector (23), in opposite orientation to a neophosphotransferase (*nptII*) kanamycin resistance gene. T-DNA borders flank the S-TNT1 and *nptII* genes, defining the transfer cassette (Fig 3-1A). This construct, pGBT7, was introduced to *Agrobacterium* virulence strain EHA101, which was used to transform tissue of *N. sylvestris*. We generated 101 independent transgenic plants from leaf explants and root-derived calli.

Retrotransposition of S-TNT1 During Agrobacterium-mediated transformation. We expected some of the kanamycin resistant plants, which contained copies of S-TNT1 linked to the T-DNA, to contain transposed copies of S-TNT1. We used Southern probing for the codon-shuffled patch to identify the number of S-TNT1 transpositions. S-TNT1 elements that are flanked by the T-DNA right border could be recognized as a 2.4kb fragment upon DraI digestion, but transposed S-TNT1 elements were at variable distances from genomic DraI sites (Fig. 3-1*A*). As predicted, the probe did not hybridize noticeably with endogenous TNT1 elements but clearly identified the codon-shuffled patch in our transgenic plants allowing us to easily count the number of transpositions per line (Fig. 3-1*B*).

Using this protocol, we found transposition in about a third of the transformed lines derived from roots or leaf explants. We found a total of 21 retrotransposition events in

the 42 plants, with no more than three transpositions in a single plant (Fig. 3-1*B,C*). To increase the frequency of transposition, we included inhibitors of epigenetic silencing in some of the regeneration media. However, neither 5-azacytidine (AzaC), an inhibitor of DNA methylation (Fig. 3-1C)(24) nor sodium butyrate (NaB), which inhibits histone deacetylation (25), significantly increased the frequency of transposition (Fig. 3-1C). In total we obtained 101 kanamycin resistant plants, of which 47 contained new transposition events.

Since tissue culture regeneration of *Medicago truncatula* was shown to trigger a burst of TNT1 retrotransposition (10), we chose a plant, Ns-7834 (Fig. 3-1*B*, lane 2), with two T-DNA and three transposed copies of S-TNT1, for regeneration. Seeds, leaf and root tissue of Ns-7834 were placed on RMOP media (26) and new plants were regenerated. Plant regeneration alone, or inclusion of AzaC and NaB in some of the media, did not lead to reactivation of the S-TNT1 element in any of the studied 107 lines (Fig. 3-S2 and Table 3-S1). Also, we did not observe direct-repeat mediated excision or recombination of any of ~500 integrated S-TNT1 elements in the independently regenerated plants (Fig. 3-S2 and Table 3-S1). Therefore, S-TNT1 can transpose during *Agrobacterium*-mediated transformation of plant tissue but is not rapidly purged from the host genome.

S-TNT1 generates a 5-bp Target Site Duplication.

Integration of TNT1 results in the production of a characteristic 5-bp target site duplication (21). To verify de novo retroelement movement, we employed inverse PCR (27) with nested primers in the S- TNT1 ‘patch’ to capture the 3’LTR and flanking

genomic sequence of transposed S- TNT1 insertions (Fig. 3-1*A*, and Table 3-S2). The downstream sequence was used to locate reads in the Methyl-filtered Tobacco genome (Table 3-S3) [http://www.pngg.org/tgi]. These reads were used to design primers to sequence the upstream genomic sequence and the full length of four transposed S-TNT1 retroelements (Table 3-S4). Sequencing the adjacent genomic sequence confirmed the generation of the 5-bp duplication at the site of insertion, each of which was different than the *nia2* duplication (Fig. 3-2, Fig. 3-S1)(21). The upstream and downstream primers were used to sequence each locus in the wild type *N. sylvestris* genome, confirming the absence of a pre-existing TNT1 insertion at each site (Fig. 3-S1). Of the insertion events characterized in this way, one (ST-3/4) was located in the coding region of a gene (Table 3-S3). These observations are consistent with integrase-mediated retrotransposition of S-TNT1 in the *N. sylvestris* genome.

LTR-Swapping with Endogenous Elements During Minus Strand Synthesis.

The "copy-and-paste" mechanism of retrotransposon replication implies that the sequence of a new insertion will be copied from the progenitor element. Only one (ST-3/3) of the four S-TNT1 3'LTRs matched this expectation (Fig. 3-3). However, a host genome with a diverse high-copy retrotransposon population should provide an opportunity for LTR swapping during reverse-transcription if related elements dimerize and are co-packaged. The other three 3'LTRs were very similar to those of P23-like members of the TNT1-A1 subfamily of retroelements which contain four, rather than three, 27-28-bp BII repeats and lack the characteristic irregular spacer of the S-TNT1 LTR (Fig. 3-3)(28).

Homology Directed Recombination in *gag-pol*.

LTR swapping indicated heterodimerization between S-TNT1 and endogenous elements in three of the four insertions. Therefore we investigated if template switching during minus strand extension would also produce recombinant *gag-pol* coding sequences. By comparing sequences of each inserted element, a wild-type element, and the progenitor S-TNT1 element, we can infer two (ST-3/4 and ST-4/4) or three (ST-1/4) recombinations within the 4-kb coding region between S-TNT1 and distinct endogenous TNT1 retrotransposons (Fig. 3-2, Fig. 3-S1). The sites of these recombinations were within homologous regions between S-TNT1 and the native TNT1 (Fig. 3-S1).

In a sample of 37 plants, that contained transpositions, we identified one plant (Ns-3A119) with a recombinant S-TNT1 'patch.' The unusual retroelement was identified by the loss of a restriction site in fragments amplified from the patch. This retrotransposon (ST-3/4) contains 242-bp of our patch and 200-bp of wild-type *gag-pol* sequence that could only come from an endogenous TNT1 element because it is absent from the S-TNT1 element. When the sequences are aligned, it is clear also that 389-bp is missing from either template. There is, however, a 7-bp homology between the sequences at the site of recombination. The sequence "ATATGGC**"** appears at position 4394 of the wild-type TNT1 element (X13777) and at position 4005 of the S-TNT1 element (Fig. 3-S1). The resulting insertion is "dead-on-arrival" and sequence annotation would identify ST-3/4 as a non-autonomous element. We infer that ST-3/4 derives from recombination between two members of the TNT1 family of high-copy retrotransposons, one synthetic and the other native, via template switching during minus strand cDNA synthesis (Fig. 3-4).

Non-identical LTRs may form during plus strand strong stop DNA synthesis.

The 5'LTR of new retroelements is assumed to be templated from a single minus strand DNA that already contains the 3'LTR sequence, ensuring the identity of LTRs at the time of insertion. Two of the insertions that we fully sequenced had nearly identical 5' and 3'LTRs, as expected. One of these (ST-3/3) had LTRs with three BII repeats and the characteristic irregular CCTTG-spacer of S-TNT1 while another (ST-4/4) had matching LTRs with four BII repeats (Fig. 3-2, 3-3).

Contrary to this expectation, the other two retroelements had different LTRs. The 5'LTR of ST-1/4 contains only one BII repeat and the 5'LTR of ST-3/4 has three, although both have 3'LTRs with four BII-repeats (Fig. 3-2, 3-3). The smaller number of BII repeats in the 5'LTRs suggested that the reverse-transcriptase may illegitimately advance through repetitive sequence during generation of the 5'LTR resulting in deletion.

Discussion

Mix-and-Paste Mating System of the TNT1 Retrotransposon.

We report here that the plant TNT1 life cycle shares the major features of animal retroviruses and the yeast Ty1 retrotransposon including dimerization of two gRNAs to produce one hybrid minus strand DNA which templates its own 5'U3-LTR and plus strand DNA (17, 20). The most important determinant of retroviral recombination is dimerization and co-packaging within a single particle. Some retroviruses, like MMLV, appear to recombine less frequently than HIV, however this difference is not due to increased processivity of the MMLV reverse-transcriptase, but is due to preferential

dimerization between identical gRNAs (29). Dimerization between two MMLV gRNAs occurs in the nucleus near the site of transcription which increases the likelihood that the pseudodiploid genome of MMLV is completely homozygous and that recombination would go unnoticed (29). For HIV, dimerization occurs in the cytoplasm in stochastic ratios between diverse gRNAs with compatible "kissing-loop" structures. Therefore the selectivity and location of dimerization is an important determinant of the mating system of LTR-retroviruses (20).

Since three of four new S-TNT1 retrotransposons in our study were hybrids (Fig. 3-2, Fig. 3-S1), we believe that the life cycle of TNT1 is more consistent with HIV than MMLV, with dimerization and copackaging likely occurring in the cytoplasm. The exception, ST-3/3, which is identical to the progenitor S-TNT1, indicated that homodimerization is also occurring (or template switching is not obligatory) and suggested that our S-TNT1 transcript made up a significant fraction of the transposition competent TNT1 gRNA mating population.

Three of the four new S-TNT1 retroelements contained endogenous sequences that evidence LTR-swapping and recombination within the *gag-pol* gene during replication. One of these (ST-4/4) had identical LTRs of endogenous origin suggesting that the 5'LTR was templated from the 3'LTR contained in a single hybrid minus strand DNA. The other two had shorter 5'LTRs than 3'LTRs, which we interpret as the product of illegitimate advancement of the reverse-transcriptase during plus strand strong stop DNA synthesis (Fig. 3-4*D*). All three contained multiple recombinations within the *gag- pol* coding sequence (Fig. 3-2, 3-S1). Dimerization and recombination during replication have not been normally considered stages in the TNT1 life cycle (30, 31) although there

has recently been a report of LTR-swapping between a mini-TNT1 that lacked the *gag-pol* gene and an endogenous TNT1 in *N. tabacum* protoplasts (16).

Mutation During Retrotransposition and the Evolutionary Clock.

Polymorphisms between pairs of LTRs are used to estimate evolutionary dates because nucleotide substitutions are assumed to accumulate randomly after insertion (12). We compared the pairs of LTRs from the four retrotransposons in the MEGA 5.05 program (32) and observed that estimates of LTR divergence were sensitive to the quality of alignment. Using the default settings, the ST-1/4 LTRs are calculated to have 0.045 substitutions per site and appear 3.5 million years old (MYA) (12, 32, 33). Relaxing the Clustal W multiple alignment gap extension penalty from 6.66 to 4.44 produced an alignment with a single gap, that produced a substitution rate estimate of 0.0 and a 0.0 million year age, because the program only considers substitutions and ignores deletions. Regardless of alignment parameters, ST-3/3 and ST-4/4 appear to be "recent" (0.0 MYA) insertions. The fourth retrotansposon, ST-3/4 would appear to be 1.0 MYA, because the ~600-bp LTRs differ by six SNPs (Fig. 3-S3). Therefore, sequence polymorphisms that can be produced during retrotransposition are a potential source of error when inferring evolutionary dates from high-copy LTR-retroelements.

Evolution of Retroelements.

The fully sequenced *Zea mays* genome contains 180 LTR-retrotransposons that harbor additional genes that were captured from the host by an unknown mechanism (4). Some animal retroviruses are known to have captured host oncogenes through the chance entry of a host mRNA to a virus particle followed by template-switching during reverse

transcription (34). Accordingly, "mix-and-paste" replication of plant high-copy retrotransposons could provide a mechanism for acquisition of host genes. A possible route for gene capture could be read-through transcription as reported for TNT1 (35). These transcripts would contain both the compatible "kissing-loop" structure required for efficient incorporation in a TNT1 virus-like particle and additional host genetic sequence. Expressed populations of endogenous retroelements are thus sampling the host sequences downstream of their integration sites for novel contributions to their "mix-and-paste" pseudodiploid genetic system.

Materials and Methods

Transformation of *Nicotiana sylvestris* TW137 leaf explants was basically preformed as described for *Nicotiana tabacum* (36). Root tissue was precultured on RM media (36) supplemented with 1-mg/L 2,4-D, 0.1-mg/L NAA, and 0.05-mg/L BAP to induce callus prior to *Agrobacterium* co-culture and transformation. Resistant plants were regenerated on RMOP media (36) containing 200-mg/L kanamycin, alone or with 100-mM AzaC or 10-mM NaB. Various concentrations and combinations of the inhibitors of epigenetic silencing were added to the RMOP media used to regenerate plants from selfed-seed, leaf and root tissue from the primary transformant Ns-7834 as shown in Table S1.

Total cellular DNA was CTAB isolated (37) from greenhouse grown leaf tissue from individually transformed or regenerated plants. Southern blots were carried out using an 824-bp BstEII-PstI S-TNT1 patch-specific probe after overnight DraI digestion of 10ug of DNA. Inverse PCR (27) began with 2-hr digestion of 600ng DNA by DraI, SpeI, or NheI (Fig. 3-1*A*) followed by overnight ligation in a large volume (200uL).

Nested amplification with patch-specific primers was carried out beginning with 2.5uL of the ligation (Table 3-S2). The products of nested PCR were Sanger sequenced. The Ns-3A119 plant containing the ST-3/4 element was identified by PstI digestion of amplicons generated with 3U3F1 and 3U3R1 primers (Fig. 3-S1, Table 3-S4). The undigested fragment observed in Ns-3A119 was Sanger sequenced and specific inverse PCR primers were designed (Table 3-S2). Once the downstream sequence was known for each retrotransposon, reads from the Methyl-filtered Tobacco genome (Table 3-S3) [http://www.pngg.org/tgi] were used to aid design of upstream and downstream primers (Table 3-S4). The studied retrotransposons were fully Sanger sequenced from amplicons that were anchored with primers in the S-TNT1 specific patch and flanking genomic sequence (Table 3-S4). Each locus was also amplified and Sanger sequenced from wild-type *N. sylvestris*-TW137 (Fig. 3-S1).

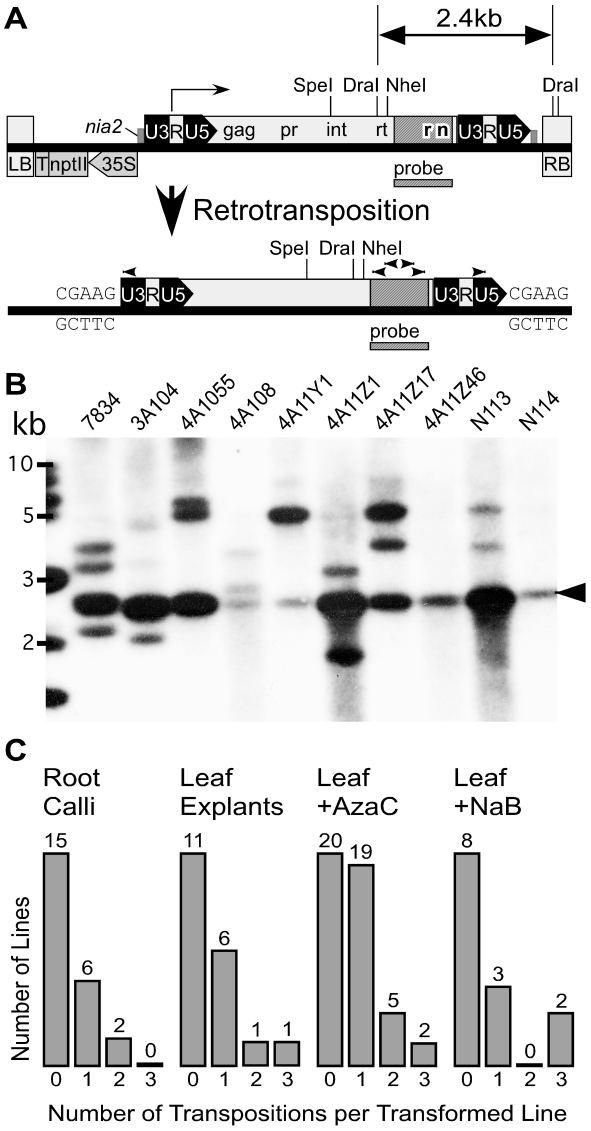


FIGURE 3-1. S-TNT1 is active during *Agrobacterium*-mediated transformation. (*A*) Diagram of S-TNT1 element in the T-DNA cassette and after retrotransposition into the *N. sylvestris* genome. The striped region represents the codon-shuffled patch of low homology between S-TNT1 and endogenous TNT1 retrotransposons. (*B*) Southern blot using the codon-shuffled region of S-TNT1 as a probe, showing the 2.4-kb fragment derived from the T-DNA (arrow) and transposed copies of S-TNT1. Each lane represents an independently transformed line. (*C*) Histograms showing numbers of transpositions in primary transformants. Abbreviations: LB, left border; T, terminator; nptII, neophosphotransferase gene; 35S, promoter sequence; *nia2*, 100-bp of *nia2* gene; U3, R, U5, regions of the long terminal repeat (LTR); gag, encapsulation protein; pr, protease; int, integrase; rt, reverse-transcriptase; rn, RNAseH; RB, right border; AzaC, 5-azacytidine; NaB, sodium butyrate.

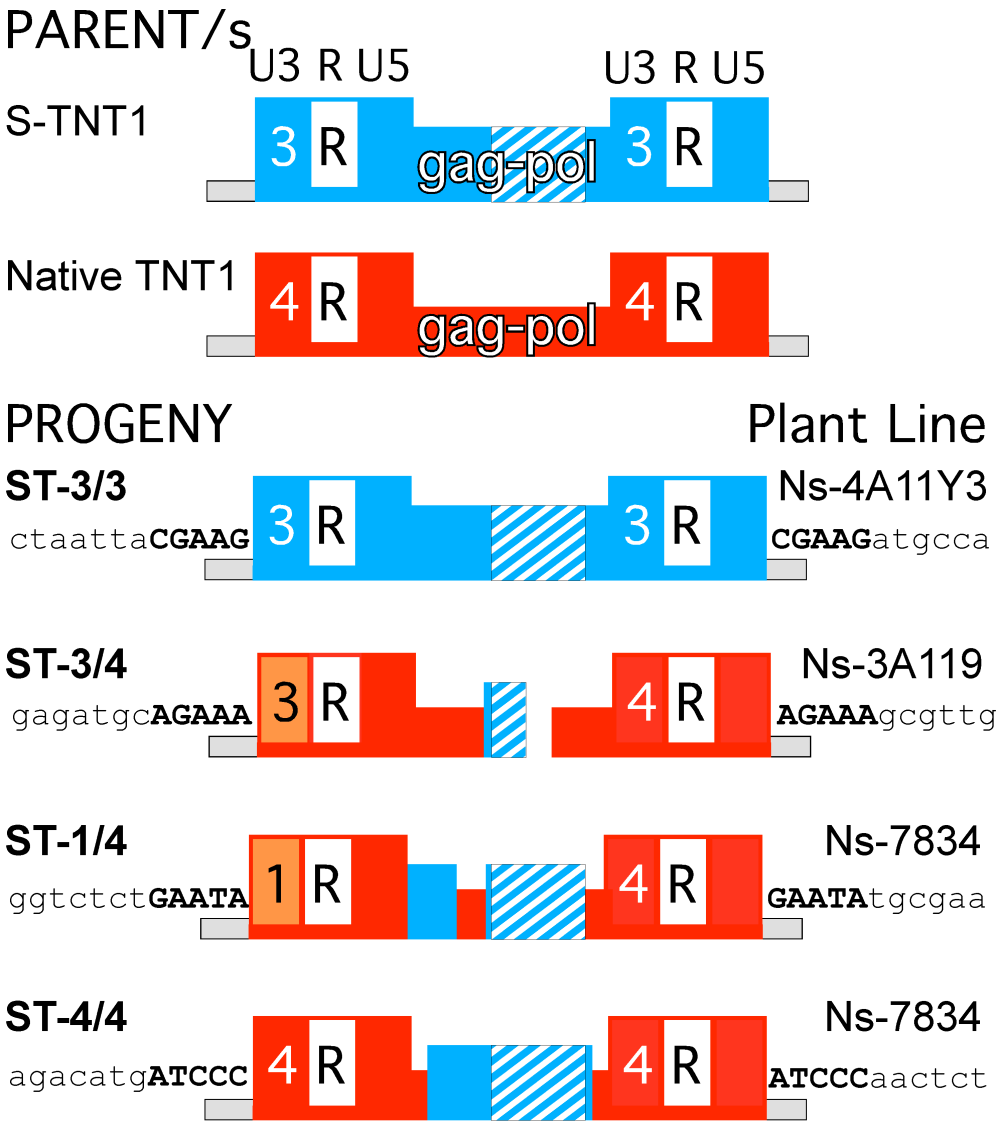


FIGURE 3-2. Newly transposed S-TNT1 retrotransposons have mixed histories. The S-TNT1 and a native TNT1 retrotransposon, symbolizing the hundreds of endogenous elements, are shown in blue and red, respectively. The four fully sequenced S-TNT1 progeny are depicted to show LTR-swapping, hybrid *gag-pol* sequence and flanking host genome sequence including the target site duplication (capitalized). Numbers in the U3 region indicate the number of BII-repeats. The striped patch represents the region of low homology between S-TNT1 and native retroelements. (See also Fig. 3-S1.)

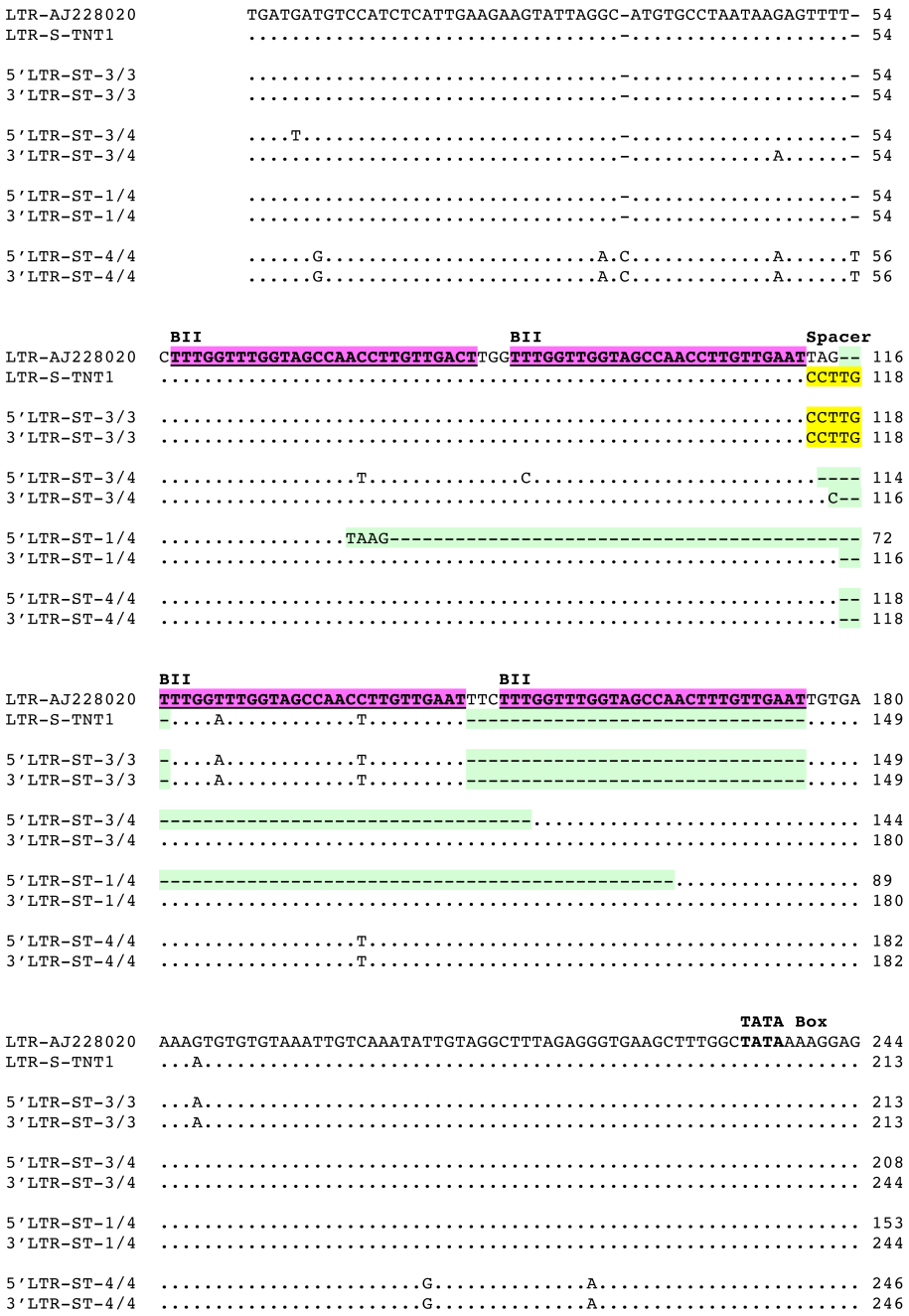


FIGURE 3-3. Alignment of U3 regions reveals LTR-swapping between S-TNT1 and endogenous syl3-AJ228020-like elements (21) and slippage during plus strand strong stop DNA synthesis. Note the variable number of BII-repeats (in magenta) and lack of the irregular CCTTG spacer (in yellow) in some LTRs relative to the progenitor S-TNT1 LTR. Gaps (in green) are shown with '-' and identical residues are shown as dots.

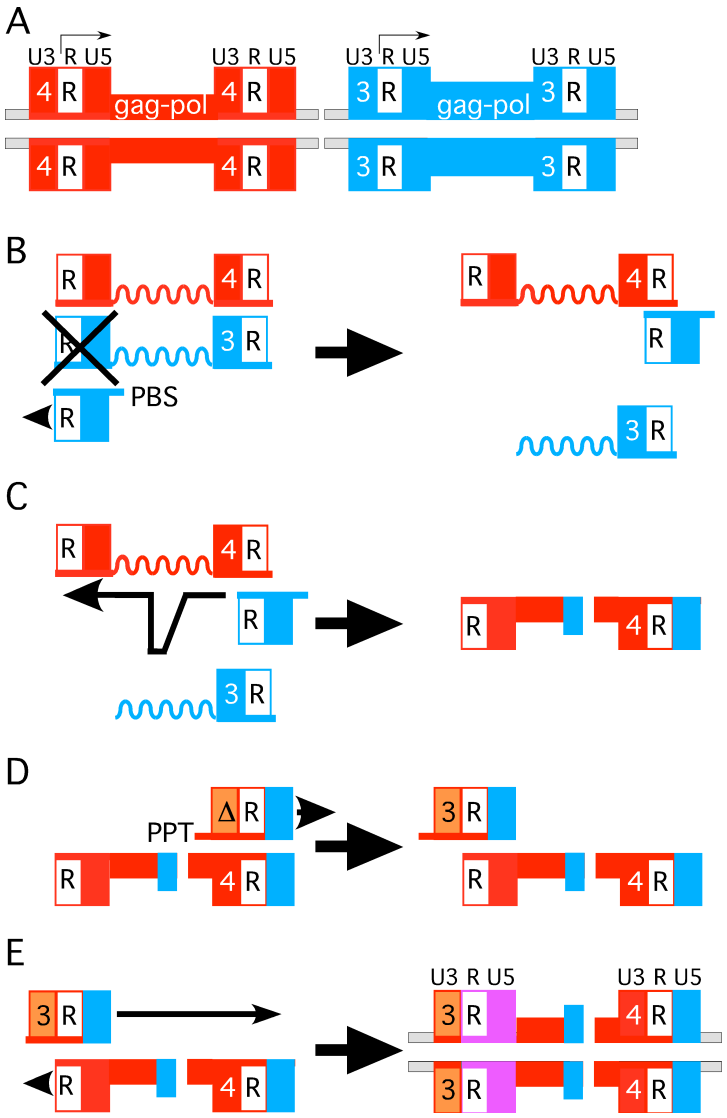
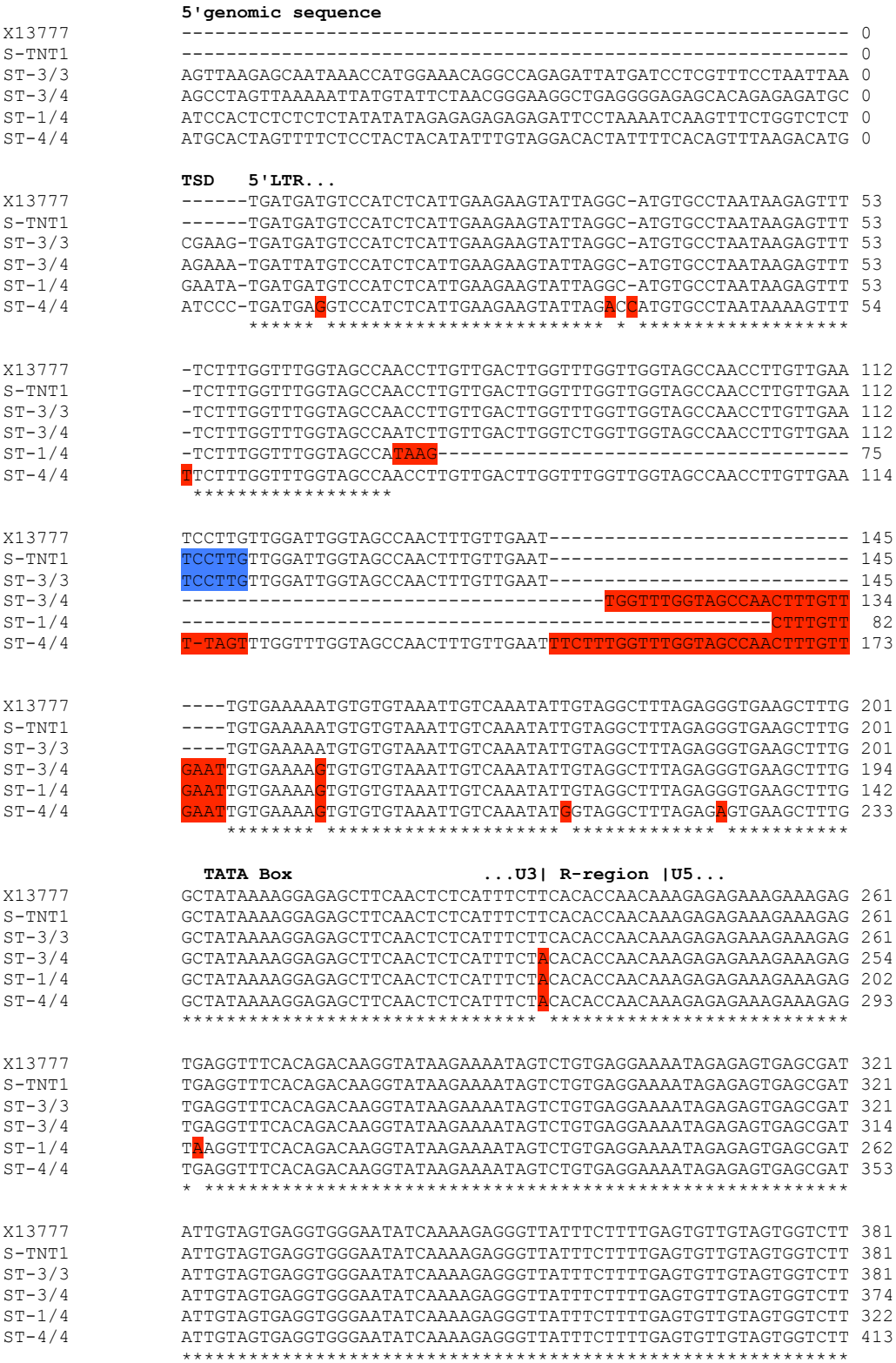


FIGURE 3-4. Mix and paste replication to produce the ST-3/4 S-TNT1 element. (*A*) Double- stranded DNA retrotransposons are transcribed from the R-region of the 5'LTR. A native element is shown in red, with 4 BII repeats in its U3 region. The S-TNT1 element with 3 BII repeats is blue. (*B*) Two distinct gRNA co-package within a virus-like particle. The strong stop (-) DNA is primed by a host tRNA at the primer binding site (PBS). The RNaseH activity of the reverse-transcriptase degrades the copied template, marked by "X". The first obligate transfer of strong stop DNA moves the strong stop (-) DNA to the 3' end of one gRNA. (*C*) Template switching occurs during extension of the minus strand DNA generating a deletion and multiple recombinations. Deletion is symbolized by a gap on the right. (*D*) Synthesis of the strong stop (+) DNA is primed from the poly- purine tract (PPT). A deletion (Δ) occurs to produce an LTR with fewer BII-repeats. The second obligate transfer moves the strong stop (+) DNA to the 3'end of the minus strand DNA. (*E*) Final extension of minus and plus strands creates a complete double stranded progeny retrotransposon. Note that the 5'U5 region is shown in purple because each DNA strand derives from a different template.

FIGURE 3-S1. Sequence of S-TNT1 and newly transposed S-TNT1 retroelements. Multiple alignment of the S-TNT1 retroelements reveals multiple recombinations within coding sequence with endogenous TNT1 elements. The top line is a native TNT1 (X13777) (20) sequence, the second line is S-TNT1. Blue highlighting shows nucleotides that evidence input from S-TNT1 while red suggests input from one of the hundreds of native TNT1s. Green shows intervals in which a recombination occurred between S-TNT1 and a native TNT1 element. Numbering indicates nucleotide position in each retrotransposon, ignoring the length of flanking sequences. Abbreviations: TSD, target site duplication; LTR, long terminal repeat; PBS, primer binding site; ATG, start codon; PPT, polypurine tract.


5'genomic sequence
X13777 ---------- 0
S-TNT1 ---------- 0
ST-3/3 AGTTAAGAGCAATAAACCATGGAAACAGGCCAGAGATTATGATCCTCGTTTCCTAATTAA 0
ST-3/4 AGCCTAGTTAAAAATTATGTATTCTAACGGGAAGGCTGAGGGGAGAGCACAGAGAGATGC 0
ST-1/4 ATCCACTCTCTCTCTATATATAGAGAGAGAGAGATTCCTAAAATCAAGTTTCTGGTCTCT 0
ST-4/4 ATGCACTAGTTTTCTCCTACTACATATTTGTAGGACACTATTTTCACAGTTTAAGACATG 0
TSD 5'LTR...
X13777 ------TGATGATGTCCATCTCATTGAAGAAGTATTAGGC-ATGTGCCTAATAAGAGTTT 53
S-TNT1 ------TGATGATGTCCATCTCATTGAAGAAGTATTAGGC-ATGTGCCTAATAAGAGTTT 53
ST-3/3 CGAAG-TGATGATGTCCATCTCATTGAAGAAGTATTAGGC-ATGTGCCTAATAAGAGTTT 53
ST-3/4 AGAAA-TGATTATGTCCATCTCATTGAAGAAGTATTAGGC-ATGTGCCTAATAAGAGTTT 53
ST-1/4 GAATA-TGATGATGTCCATCTCATTGAAGAAGTATTAGGC-ATGTGCCTAATAAGAGTTT 53
ST-4/4 ATCCC-TGATGAGGTCCATCTCATTGAAGAAGTATTAGACCATGTGCCTAATAAAAGTTT 54
X13777 -TCTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGTTGGTAGCCAACCTTGTTGAA 112
S-TNT1 -TCTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGTTGGTAGCCAACCTTGTTGAA 112
ST-3/3 -TCTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGTTGGTAGCCAACCTTGTTGAA 112
ST-3/4 -TCTTTGGTTTGGTAGCCAATCTTGTTGACTTGGTCTGGTTGGTAGCCAACCTTGTTGAA 112
ST-1/4 -TCTTTGGTTTGGTAGCCATAAG---------- 75
ST-4/4 TTCTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGTTGGTAGCCAACCTTGTTGAA 114
X13777 TCCTTGTTGGATTGGTAGCCAACTTTGTTGAAT---------- 145
S-TNT1 TCCTTGTTGGATTGGTAGCCAACTTTGTTGAAT---------- 145
ST-3/3 TCCTTGTTGGATTGGTAGCCAACTTTGTTGAAT---------- 145
ST-3/4 ----------TGGTTTGGTAGCCAACTTTGTT 134
ST-1/4 ----------CTTTGTT 82
ST-4/4 T-TAGTTTGGTTTGGTAGCCAACTTTGTTGAATTTCTTTGGTTTGGTAGCCAACTTTGTT 173
X13777 ----TGTGAAAAATGTGTGTAAATTGTCAAATATTGTAGGCTTTAGAGGGTGAAGCTTTG 201
S-TNT1 ----TGTGAAAAATGTGTGTAAATTGTCAAATATTGTAGGCTTTAGAGGGTGAAGCTTTG 201
ST-3/3 ----TGTGAAAAATGTGTGTAAATTGTCAAATATTGTAGGCTTTAGAGGGTGAAGCTTTG 201
ST-3/4 GAATTGTGAAAAGTGTGTGTAAATTGTCAAATATTGTAGGCTTTAGAGGGTGAAGCTTTG 194
ST-1/4 GAATTGTGAAAAGTGTGTGTAAATTGTCAAATATTGTAGGCTTTAGAGGGTGAAGCTTTG 142
ST-4/4 GAATTGTGAAAAGTGTGTGTAAATTGTCAAATATGGTAGGCTTTAGAGAGTGAAGCTTTG 233
TATA Box ...U3| R-region |U5...
X13777 GCTATAAAAGGAGAGCTTCAACTCTCATTTCTTCACACCAACAAAGAGAGAAAGAAAGAG 261
S-TNT1 GCTATAAAAGGAGAGCTTCAACTCTCATTTCTTCACACCAACAAAGAGAGAAAGAAAGAG 261
ST-3/3 GCTATAAAAGGAGAGCTTCAACTCTCATTTCTTCACACCAACAAAGAGAGAAAGAAAGAG 261
ST-3/4 GCTATAAAAGGAGAGCTTCAACTCTCATTTCTACACACCAACAAAGAGAGAAAGAAAGAG 254
ST-1/4 GCTATAAAAGGAGAGCTTCAACTCTCATTTCTACACACCAACAAAGAGAGAAAGAAAGAG 202
ST-4/4 GCTATAAAAGGAGAGCTTCAACTCTCATTTCTACACACCAACAAAGAGAGAAAGAAAGAG 293
X13777 TGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGTGAGGAAAATAGAGAGTGAGCGAT 321
S-TNT1 TGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGTGAGGAAAATAGAGAGTGAGCGAT 321
ST-3/3 TGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGTGAGGAAAATAGAGAGTGAGCGAT 321
ST-3/4 TGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGTGAGGAAAATAGAGAGTGAGCGAT 314
ST-1/4 TAAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGTGAGGAAAATAGAGAGTGAGCGAT 262
ST-4/4 TGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGTGAGGAAAATAGAGAGTGAGCGAT 353
X13777 ATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTTCTTTTGAGTGTTGTAGTGGTCTT 381
S-TNT1 ATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTTCTTTTGAGTGTTGTAGTGGTCTT 381
ST-3/3 ATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTTCTTTTGAGTGTTGTAGTGGTCTT 381
ST-3/4 ATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTTCTTTTGAGTGTTGTAGTGGTCTT 374
ST-1/4 ATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTTCTTTTGAGTGTTGTAGTGGTCTT 322
ST-4/4 ATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTTCTTTTGAGTGTTGTAGTGGTCTT 413

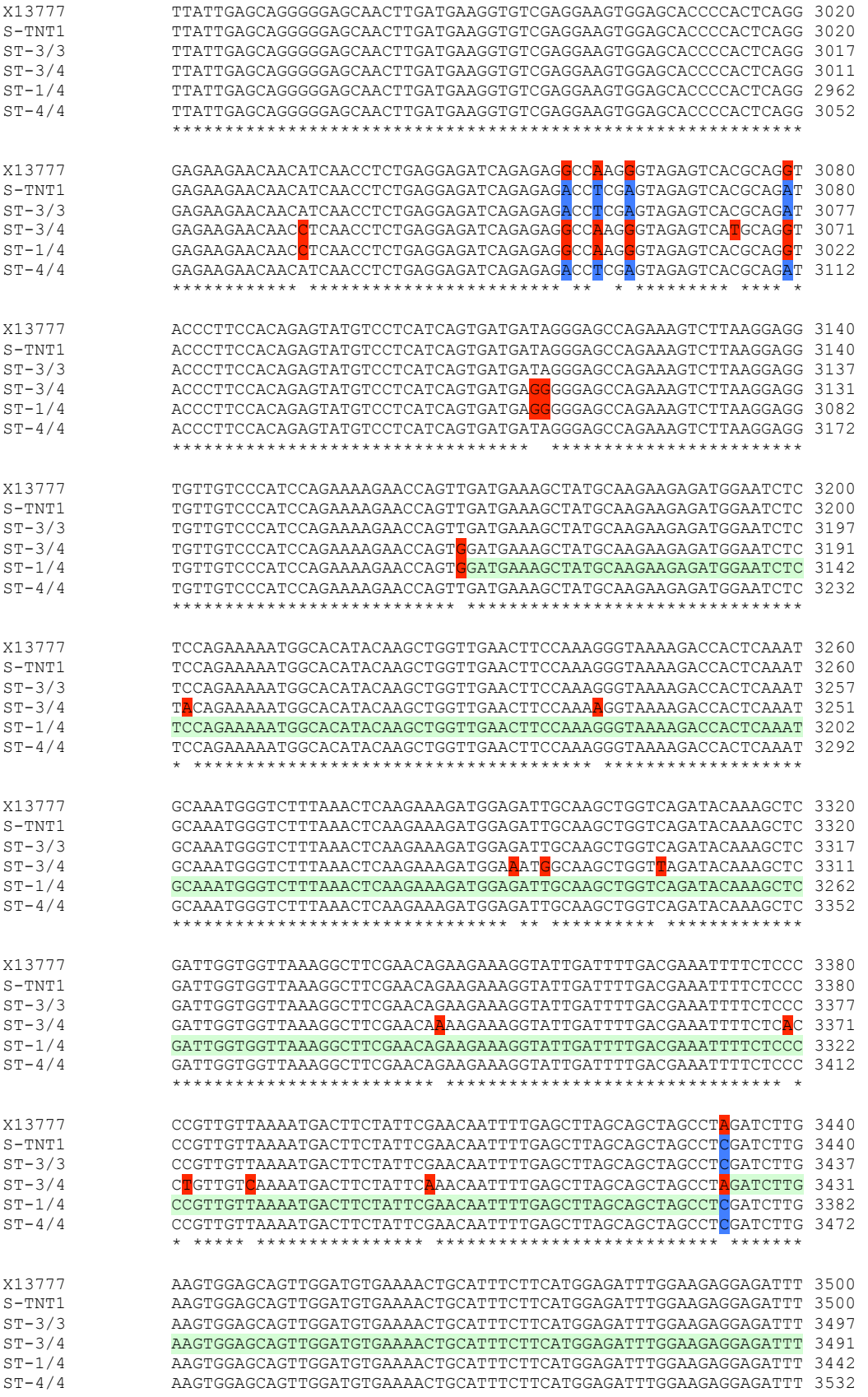
```
X13777          TGGAGTATTTACCTCCGACCTACAAAGT-GTAAAATTCCTTACTATAGTGATATCAGTTG 440
S-TNT1          TGGAGTATTTACCTCCGACCTACAAAGT-GTAAAATTCCTTACTATAGTGATATCAGTTG 440
ST-3/3          TGGAGTATT-ACCTCCGACCTACAAAGT-GTAAAATTCCTTACTATAGTGATATCAGTTG 439
ST-3/4          TGGAGTATTTATCTCCGACCTACAAAGT-GTAAAATTCCTTACTATAGTGATATCAGTTG 433
ST-1/4          TGGAGTATTTACCTCCGACCTACAAAGTTGTAAAATTCCTTACTATAGTGATATCAGTTG 382
ST-4/4          TGGAGTATTTACCTCCGACCTACAAAGT-GTAAAATTCCTTACTATAGTGATATCAGTTG 472
                ********* * ********************************************

X13777          CTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCAGAAGGGTTTTCCACGTAAAAATCT 500
S-TNT1          CTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCAGAAGGGTTTTCCACGTAAAAATCT 500
ST-3/3          CTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCAGAAGGGTTTTCCACGTAAAAATCT 499
ST-3/4          CTCCTCTCGGGGTCGTGGTTTTT--TCCCTTATTCAGAAGGGTTTTCCATGTAAAAATCT 491
ST-1/4          CTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCAGAAGGGTTTTCCACGTAAAAATTT 442
ST-4/4          CTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCAGAAGGGTTTTCCACGTAAAAATCT 532
                ********************  ********************** ******** *

X13777          TGGTGTCATTGTTACTCTTTTATTCTTGTTAATTACCGTATCTCGGTGCTACATTATTAT 560
S-TNT1          TGGTGTCATTGTTACTCTTTTATTCTTGTTAATTACCGTATCTCGGTGCTACATTATTAT 560
ST-3/3          TGGTGTCATTGTTACTCTTTTATTCTTGTTAATTACCGTATCTCGGTGCTACATTATTAT 559
ST-3/4          TGGTGTCATTGTTACTCTTTTATTCTTGTTAATTACCGTATCTCGGTGCTACATTATTAT 551
ST-1/4          TGGTGTCATTGTTACTCTTTTATTCTTGTTAATTACCGTATCTCGGTGCTACATTATTAT 502
ST-4/4          TGGTGTCATTGTTACTCTTTTATTCTTGTTAATTACCGTATCTCGGTGCTACATTATTAT 592
                ************************************************************

                                                          ...LTR       PBS
X13777          TCCGCTTTATTACCGTGAATATTATTTTGGTAAGGGGTTTATTCCCAACAACTGGTATCA 620
S-TNT1          TCCGCTTTATTACCGTGAATATTATTTTGGTAAGGGGTTTATTCCCAACAACTGGTATCA 620
ST-3/3          TCCGCTTTATTACCGTGAATATTATTTTGGTAAGGGGTTTATTCCCAACAACTGGTATCA 619
ST-3/4          TCCGCTTTATTACCGTGAATATTATTTTGGTAAGGGGTTTATTCCCAACAACTGGTATCA 611
ST-1/4          TCCGCTTTATTACCGTGAATATTATTTTGGTAAGGGGTTTATTCCCAACAACTGGTATCA 562
ST-4/4          TCCGCTTTATTACCGTGAATATTATTTTGGTAAGGGGTTTATTCCCAACAACTGGTATCA 652
                ************************************************************

X13777          GAGCACAGGTTCTGCTCGTTCACTGAAATACTATTCACTGTCGGTAGTACTATACTTGGT 680
S-TNT1          GAGCACAGGTTCTGCTCGTTCACTGAAATACTATTCACTGTCGGTAGTACTATACTTGGT 680
ST-3/3          GAGCACAGGTT-T-CTCGTTCACTGAAATACTATTCACTGTCGGTAGTACTATACTTGGT 677
ST-3/4          GAGCACAGGTTCTGCTCGTTCACTGAAATACTATTCACTGTCGGTAGTACTATACTTGGT 671
ST-1/4          GAGCACAGGTTCTGCTCGTTCACTGAAATACTATTCACTGTCGGTAGTACTATACTTGGT 622
ST-4/4          GAGCGCAGGTTTTTCTCGTTCACTGAAATACTATTCACTGTCGGTAGTACTATACTTGGT 712
                **** ****** * ***********************************************

                          ATG-Start gag-pol...
X13777          GAAAAATAAAAATGTCTGGAGTAAAGTACGAGGTAGCAAAATTCAATGGAGATAACGGTT 740
S-TNT1          GAAAAATAAAAATGTCTGGAGTAAAGTACGAGGTAGCAAAATTCAATGGAGATAACGGTT 740
ST-3/3          GAAAAATAAAAATGTCTGGAGTAAAGTACGAGGTAGCAAAATTCAATGGAGATAACGGTT 737
ST-3/4          GAAAAATAAAAATGTCCGGAGTAAAGTACGAGGTACCAAAATTCAACGGAGATAACGGTT 731
ST-1/4          GAAAAATAAAAATGTCTGGAGTAAAGTACGAGGTAGCAAAATTCAATGGAGATAACGGTT 682
ST-4/4          GAAAAATAAAAATGTCTGGAGTAAAGTACGAGGTAGCAAAATTCAATGGAGATAACGGTT 772
                *************** ***************** ********** *************

X13777          TCTCAACATGGCAAAGAAGGATGAGAGATCTGCTCATCCAACAAGGATTACACAAGGTTC 800
S-TNT1          TCTCAACATGGCAAAGAAGGATGAGAGATCTGCTCATCCAACAAGGCCTACACAAGGTTC 800
ST-3/3          TCTCAACATGGCAAAGAAGGATGAGAGATCTGCTCATCCAACAAGGCCTACACAAGGTTC 797
ST-3/4          TCTCAACATGGCAAAGAAGGATGAGAGATCTGCTCATCCAACAAGGATTACACAAGGTTC 791
ST-1/4          TCTCAACATGGCAAAGAAGGATGAGAGATCTGCTCATCCAACAAGGCCTACACAAGGTTC 742
ST-4/4          TCTCAACATGGCAAAGAAGGATGAGAGATCTGCTCATCCAACAAGGCCTACACAAGGTTC 832
                ********************************************  ************

X13777          TAGATGTTGATTCCAAAAAGCCTGATACCATGAAAGCTGAGGATTGGGCTGACTTGGATG 860
S-TNT1          TTGATGTTGATTCCAAAAAGCCTGATACCATGAAAGCTGAGGATTGGGCTGACTTGGATG 860
ST-3/3          TTGATGTTGATTCCAAAAAGCCTGATACCATGAAAGCTGAGGATTGGGCTGACTTGGATG 857
ST-3/4          TAGATGTTGATTCCAAAAAGCCTGATACCATGAAAACTGAGGATTGGGCTGACTTAGATG 851
ST-1/4          TTGATGTTGATTCCAAAAAGCCTGATACCATGAAAGCTGAGGATTGGGCTGACTTGGATG 802
ST-4/4          TTGATGTTGATTCCAAAAAGCCTGATACCATGAAAGCTGAGGATTGGGCTGACTTGGATG 892
                * ********************************** ****************** ****
```

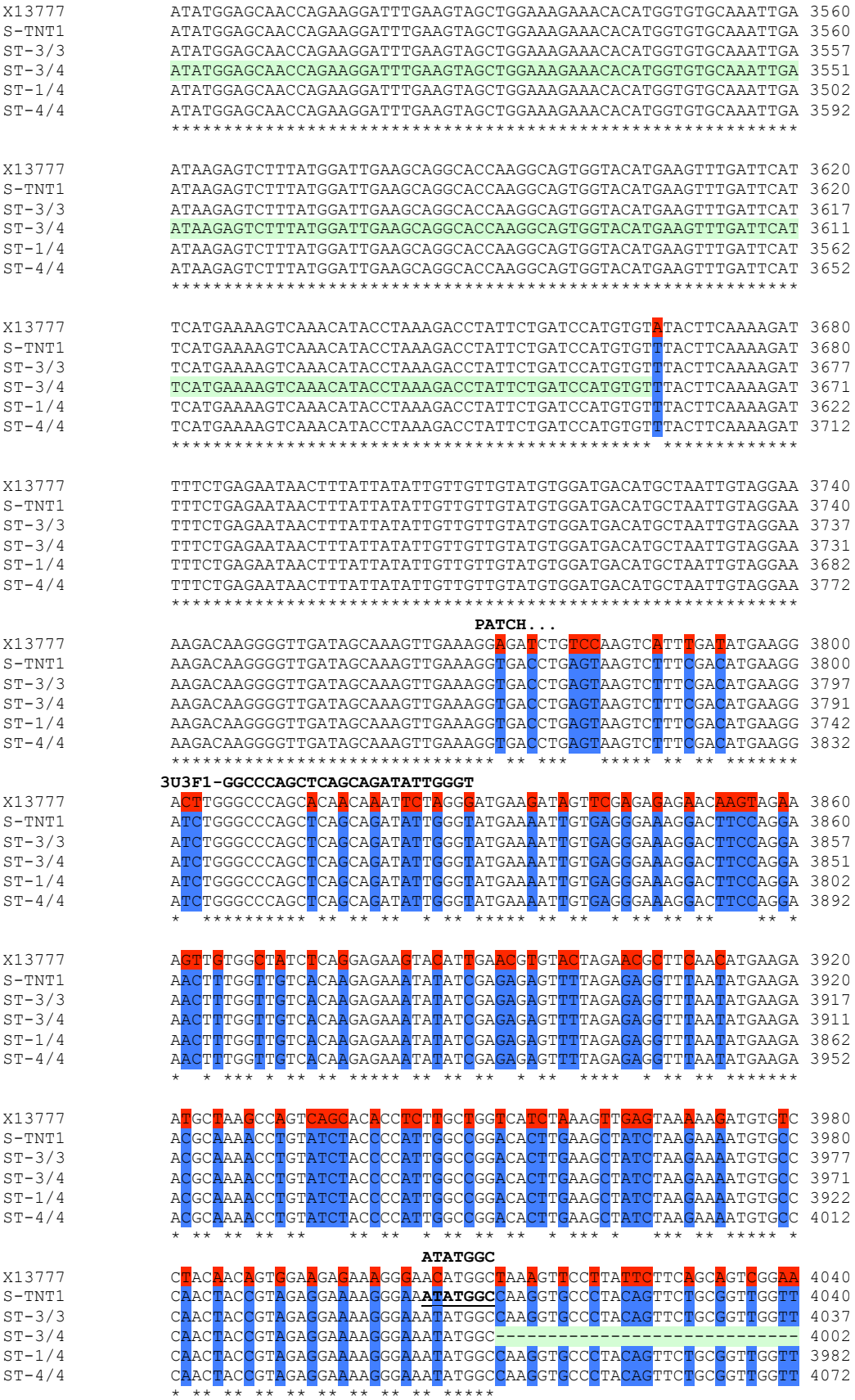
```
X13777          AAAGAGCTGCTAGTGCAATCAGGTTGCACTTATCAGATGATGTGGTAAATAACATCATTG 920
S-TNT1          AAAGAGCTGCTAGTGCAATCAGGTTGCACTTATCAGATGATGTGGTAAATAACATCATTG 920
ST-3/3          AAAGAGCTGCTAGTGCAATCAGGTTGCACTTATCAGATGATGTGGTAAATAACATCATTG 917
ST-3/4          AAAGAGATGCTAGTGCAATTAGGTTGCACTTATCAGATGATGTGGTAAATAACATCATTG 911
ST-1/4          AAAGAGCTGCTAGTGCAATCAGGTTGCACTTATCAGATGATGTGGTAAATAACATCATTG 862
ST-4/4          AAAGAGCTGCTAGTGCAATCAGGTTGCACTTATCAGATGATGTGGTAAATAACATCATTG 952
                ****** ************ ****************************************

X13777          ATGAAGACACTGCACGTGGAATTTGGACAAGGTTGGAAAGCCTATACATGTCCAAAACGC 980
S-TNT1          ATGAAGACACTGCACGTGGAATTTGGACAAGGTTGGAAAGCCTATACATGTCCAAAACGC 980
ST-3/3          ATGAAGACACTGCACGTGGAATTTGGACAAGGTTGGAAAGCCTATACATGTCCAAAACGC 977
ST-3/4          ATGAAGACACTACACGTGGAATTTGGACAAGGTTGGAAAGCCTATACATGTCCAAAACGC 971
ST-1/4          ATGAAGACACTGCACGTGGAATTTGGACAAGGTTGGAAAGCCTATACATGTCCAAAACGC 922
ST-4/4          ATGAAGACACTGCACGTGGAATTTGGACAAGGTTGGAAAGCCTATACATGTCCAAAACGC 1012
                *********** ************************************************

X13777          TGACAAATAAATTGTACCTGAAGAAGCAGTTATACGCCCTACACATGAGTGAAGGTACGA 1040
S-TNT1          TGACAAATAAATTGTACCTGAAGAAGCAGTTATACGCCCTACACATGAGTGAAGGTACGA 1040
ST-3/3          TGACAAATAAATTGTACCTGAAGAAGCAGTTATACGCCCTACACATGAGTGAAGGTACGA 1037
ST-3/4          TGACAAATAAATTGTACCTGAAGAAGCAGTTATACGCCCTACACATGAGTGAAGGTACAA 1031
ST-1/4          TGACAAATAAATTGTACCTGAAGAAGCAGTTATACGCCCTACACATGAGTGAAGGTACGA 982
ST-4/4          TGACAAATAAATTGTACCTGAAGAAGCAGTTATACGCCCTACACATGAGTGAAGGTACGA 1072
                ************************************************************ *

X13777          ATTTTTTGTCACATTTAAATGTGTTTAACGGACTAATCACACAGCTTGCCAACCTCGGAG 1100
S-TNT1          ATTTTTTGTCACATTTAAATGTGTTTAACGGACTAATCACACAGCTTGCCAACCTCGGAG 1100
ST-3/3          ATTTTTTGTCACATTTAAATGTGTTTAACGGACTAATCACACAGCTTGCCAACCTCGGAG 1097
ST-3/4          ATTTTTTGTCACATTTAAATGTGTTTAACGAACTAATCACACAGCTTGCCAACCTCGGAG 1091
ST-1/4          ATTTTTTGTCACATTTAAATGTGTTTAACGGACTAATCACACAGCTTGCCAACCTCGGAG 1042
ST-4/4          ATTTTTTGTCACATTTAAATGTGTTTAACGGACTAATCACACAGCTTGCCAACCTCGGAG 1132
                ****************************** ******************************

X13777          TGAAAATCGAGGAAGAAGATAAAGCCATCTTGCTATTGAACTCGTTGCCATCTTCGTACG 1160
S-TNT1          TGAAAATCGAGGAAGAAGATAAAGCCATCTTGCTATTGAACTCGTTGCCATCTTCGTACG 1160
ST-3/3          TGAAAATCGAGGAAGAAGATAAAGCCATCTTGCTATTGAACTCGTTGCCATCTTCGTACG 1157
ST-3/4          TGAAAATCGAGGAAGAAGATAAAGTCATCTTGCTATTGAACTCGTTGCCATCTTCGTATG 1151
ST-1/4          TGAAAATCGAGGAAGAAGATAAAGCCATCTTGCTATTGAACTCGTTGCCATCTTCGTACG 1102
ST-4/4          TGAAAATCGAGGAAGAAGATAAAGCCATCTTGCTATTGAACTCGTTGCCATCTTCGTACG 1192
                ************************ ********************************** *

X13777          ATAATTTGGCAACAACCATCCTGCACGGTAAGACTACTATTGAGTTGAAAGATGTCACAT 1220
S-TNT1          ATAATTTGGCAACAACCATCCTGCACGGTAAGACTACTATTGAGTTGAAAGATGTCACAT 1220
ST-3/3          ATAATTTGGCAACAACCATCCTGCACGGTAAGACTACTATTGAGTTGAAAGATGTCACAT 1217
ST-3/4          ATAATTTGGCAACAACCATCCTGCATGGTAAGACTACTATTGAGTTGAAAGATGTCACAT 1211
ST-1/4          ATAATTTGGCAACAACCATCCTGCACGGTAAGACTACTATTGAGTTGAAAGATGTCACAT 1162
ST-4/4          ATAATTTGGCAACAACCATCCTGCACGGTAAGACTACTATTGAGTTGAAAGATGTCACAT 1252
                ************************* ********************************

X13777          CGGCTCTTCTACTCAATGAGAAGATGAGAAAGAAGCCTGAAAATCAAGGACAGGCTCTCA 1280
S-TNT1          CGGCTCTTCTACTCAATGAGAAGATGAGAAAGAAGCCTGAAAATCAAGGACAGGCTCTCA 1280
ST-3/3          CGGCTCTTCTACTCAATGAGAAGATGAGAAAGAAGCCTGAAAATCAAGGACAGGCTCTCA 1277
ST-3/4          CGGCTCTTCTACTCAATGAGAAGATAAGAAAGAAGCCTGAAAATCAAGGACAGGCTCTCA 1271
ST-1/4          CGGCTCTTCTACTCAATGAGAAGATGAGAAAGAAGCCTGAAAATCAAGGACAGGCTCTCA 1222
ST-4/4          CGGCTCTTCTACTCAATGAGAAGATGAGAAAGAAGCCTGAAAATCAAGGACAGGCTCTCA 1312
                ************************* ********************************

X13777          TCACAGAAGGTAGAGGCAGGAGTTATCAAAGGAGTTCGAACAACTATGGTAGATCCGGAG 1340
S-TNT1          TCACAGAAGGTAGAGGCAGGAGTTATCAAAGGAGTTCGAACAACTATGGTAGATCCGGTG 1340
ST-3/3          TCACAGAAGGTAGAGGCAGGAGTTATCAAAGGAGTTCGAACAACTATGGTAGATCCGGTG 1337
ST-3/4          TCACAGAAGGTAGAGGCAGGAGTTATCAAAGGAGTTCGAACAACTATGGTAGATCCGGAG 1331
ST-1/4          TCACAGAAGGTAGAGGCAGGAGTTATCAAAGGAGTTCGAACAACTATGGTAGATCCGGTG 1282
ST-4/4          TCACAGAAGGTAGAGGCAGGAGTTATCAAAGGAGTTCGAACAACTATGGTAGATCCGGTG 1372
                ************************************************************ *
```

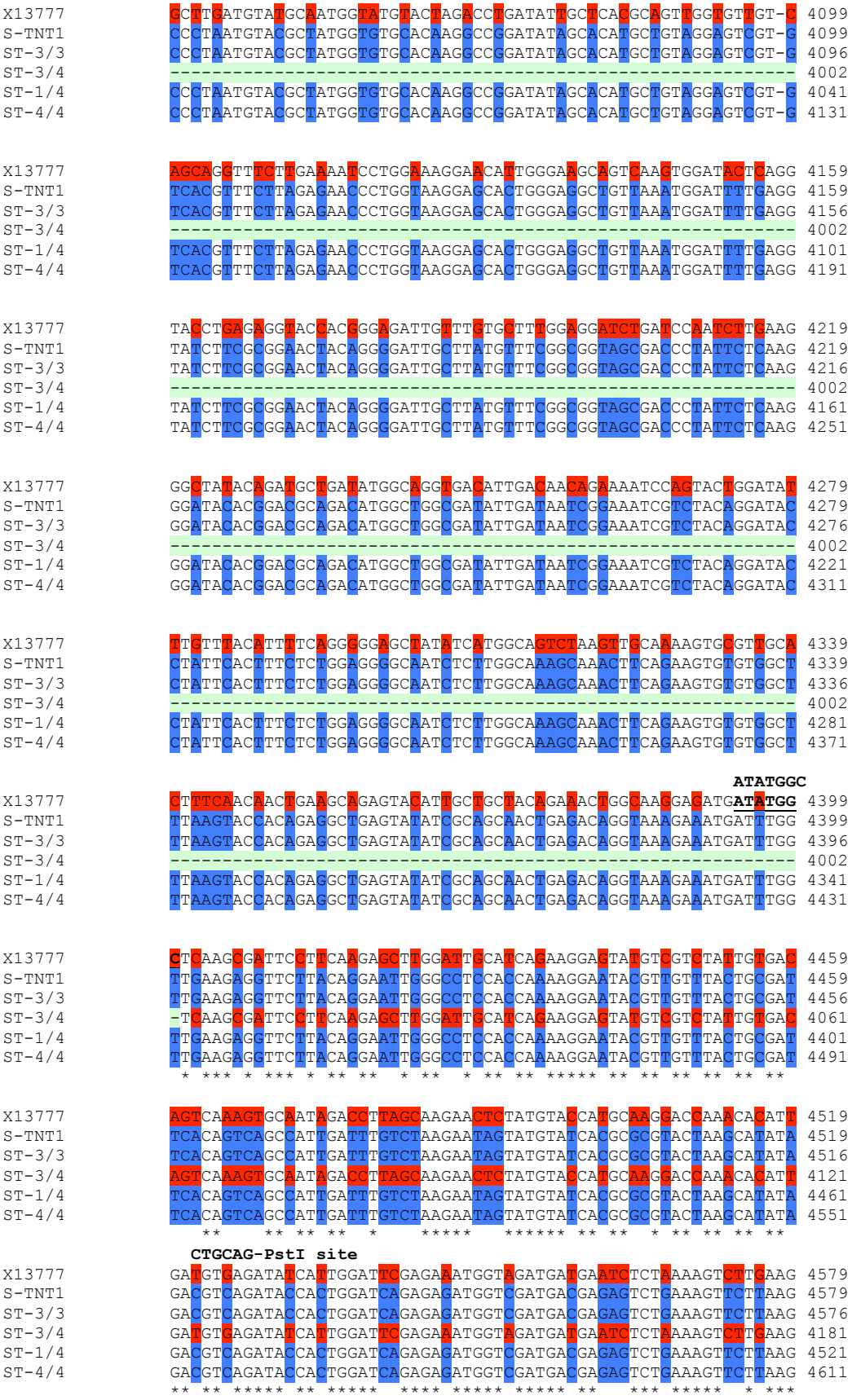
```
X13777          CTCGTGGGAAGTCAAAGAATCGATCCAAATCAAGAGTCAGAAATTGCTACAACTGTAATC 1400
S-TNT1          CTCGTGGGAAGTCAAAGAATCGATCCAAATCAAGAGTCAGAAATTGCTACAACTGTAATC 1400
ST-3/3          CTCGTGGGAAGTCAAAGAATCGATCCAAATCAAGAGTCAGAAATTGCTACAACTGTAATC 1397
ST-3/4          CTCGTGGGAAGTCAAAGAACCGATCCAAATCAAGAGTCAGAAATTGCTACAACTGTAATC 1391
ST-1/4          CTCGTGGGAAGTCAAAGAATCGATCCAAATCAAGAGTCAGAAATTGCTACAACTGTAATC 1342
ST-4/4          CTCGTGGGAAGTCAAAGAATCGATCCAAATCAAGAGTCAGAAATTGCTACAACTGTAATC 1432
                ****************** ****************************************

X13777          AACCAGGTCACTTCAAAAGAGATTGCCCAAATCCAAGGAAGGGCAAAGGTGAAACCAGTG 1460
S-TNT1          AACCAGGTCACTTCAAAAGAGATTGCCCAAATCCAAGGAAGGGCAAAGGTGAAACCAGTG 1460
ST-3/3          AACCAGGTCACTTCAAAAGAGATTGCCCAAATCCAAGGAAGGGCAAAGGTGAAACCAGTG 1457
ST-3/4          AACCAAGTCACTTCAAAAGAGATTTCCCAAATCCAAGGAAGGGCAAAGGTGAAACCAGTG 1451
ST-1/4          AACCAGGTCACTTCAAAAGAGATTGCCCAAATCCAAGGAAGGGCAAAGGTGAAACCAGTG 1402
ST-4/4          AACCAGGTCACTTCAAAAGAGATTGCCCAAATCCAAGGAAGGGCAAAGGTGAAACCAGTG 1492
                ***** ****************** ************************************

X13777          GCCAGAAGAATGACGACAACACAGCCGCCATGGTGCAAAATAATGATAATGTTGTCCTCT 1520
S-TNT1          GCCAGAAGAATGACGACAACACAGCCGCTATGGTGCAAAATAATGATAATGTTGTCCTCT 1520
ST-3/3          GCCAGAAGAATGACGACAACACAGCCGCTATGGTGCAAAATAATGATAATGTTGTCCTCT 1517
ST-3/4          GCCAGAAGAATGACGACAACACAGCCGCCATGGTACAAAATAATGATAATGTTGTCCTCT 1511
ST-1/4          GCCAGAAGAATGACGACAACACAGCCGCTATGGTGCAAAATAATGATAATGTTGTCCTCT 1462
ST-4/4          GCCAGAAGAATGACGACAACACAGCCGCTATGGTGCAAAATAATGATAATGTTGTCCTCT 1552
                **************************** ***** ************************

X13777          TTATAAATGAGGAAGAGGAATGCATGCACCTGTCAGGTCCAGAGTCGGAATGGGTGGTTG 1580
S-TNT1          TTATTAATGAGGAAGAGGAATGTATGCACCTGTCAGGTCCAGAGTCGGAATGGGTGGTTG 1580
ST-3/3          TTATTAATGAGGAAGAGGAATGTATGCACCTGTCAGGTCCAGAGTCGGAATGGGTGGTTG 1577
ST-3/4          TTATAAATGAGGAAGAGGAATGCATGCACCTGTCAGGTCCAGAGTCGGAATGGGTGGTTG 1571
ST-1/4          TTATTAATGAGGAAGAGGAATGTATGCACCTGTCAGGTCCAGAGTCGGAATGGGTGGTTG 1522
ST-4/4          TTATTAATGAGGAAGAGGAATGTATGCACCTGTCAGGTCCAGAGTCGGAATGGGTGGTTG 1612
                **** ***************** ************************************

X13777          ACACAGCGGCATCTCACCATGCCACACCGGTAAGAGATCTTTTTTGCAGATATGTAGCAG 1640
S-TNT1          ACACAGCGGCATCTCACCATGCCACACCGGTAAGAGACCTTTTTTGCAGATATGTAGCAG 1640
ST-3/3          ACACAGCGGCATCTCACCATGCCACACCGGTAAGAGACCTTTTTTGCAGATATGTAGCAG 1637
ST-3/4          ACACAGCGGCATCTCACCATGCCACACCGGTAAGAGATCTTTTTTGCAGATATGTAGCAG 1631
ST-1/4          ACACAGCGGCATCTCACCATGCCACACCGGTAAGAGACCTTTTTTGCAGATATGTAGCAG 1582
ST-4/4          ACACAGCGGCATCTCACCATGCCACACCGGTAAGAGACCTTTTTTGCAGATATGTAGCAG 1672
                ************************************ **********************

X13777          GTGATTTCGGCACAGTGAAAATGGGTAACACAAGTTACTCAAAGATTGCGGGGATTGGTG 1700
S-TNT1          GTGATTTCGGCACAGTGAAAATGGGTAACACAAGTTACTCAAAGATTGCGGGGATTGGTG 1700
ST-3/3          GTGATTTCGGCACAGTGAAAATGGGTAACACAAGTTACTCAAAGATTGCGGGGATTGGTG 1697
ST-3/4          GTGATTTCGGCACAGTGAAAATGGGTAACACAAGTTACTCAAAGATTGTGGGGATTGGTG 1691
ST-1/4          GTGATTTCGGCACAGTGAAAATGGGTAACACAAGTTACTCAAAGATTGCGGGGATTGGTG 1642
ST-4/4          GTGATTTCGGCACAGTGAAAATGGGTAACACAAGTTACTCAAAGATTGCGGGGATTGGTG 1732
                ********************************************** ************

X13777          ACATTTGTATCAAGACAAATGTCGGATGCACATTGGTTCTAAAGGATGTGCGGCATGTAC 1760
S-TNT1          ACATTTGTATCAAGACAAATGTCGGATGCACATTGGTTCTAAAGGATGTGCGGCATGTAC 1760
ST-3/3          ACATTTGTATCAAGACAAATGTCGGATGCACATTGGTTCTAAAGGATGTGCGGCATGTAC 1757
ST-3/4          ACATTTGTATCAAGACAAATGTTGGATGCACATTGGTTCTAAAGGATGTGCGGCATGTAC 1751
ST-1/4          ACATTTGTATCAAGACAAATGTCGGATGCACATTGGTTCTAAAGGATGTGCGGCATGTAC 1702
ST-4/4          ACATTTGTATCAAGACAAATGTCGGATGCACATTGGTTCTAAAGGATGTGCGGCATGTAC 1792
                ********************** ************************************

X13777          CTGATTTGCGGATGAACTTGATCTCGGGAATTGCTTTAGACCGAGATGGATACGAGAGTT 1820
S-TNT1          CTGATTTGCGGATGAACTTGATCTCGGGAATTGCTTTAGACCGAGATGGATACGAGAGTT 1820
ST-3/3          CTGATTTGCGGATGAACTTGATCTCGGGAATTGCTTTAGACCGAGATGGATACGAGAGTT 1817
ST-3/4          CTGATTTGCGGATGAACTTGATCTCGGGAATTGCTTTAGACCGAGATGGATACGAGAGTT 1811
ST-1/4          CTGATTTGCGGATGAACTTGATCTCGGGAATTGCTTTAGACCGAGATGGATACGAGAGTT 1762
ST-4/4          CTGATTTGCGGATGAACTTGATCTCGGGAATTGCTTTAGACCGAGATGGATACGAGAGTT 1852
                ************************************************************

X13777          ATTTTGCAAATCAAAAGTGGAGACTCACTAAGGGATCATTGGTGATTGCAAAGGGAGTTG 1880
S-TNT1          ATTTTGCAAATCAAAAGTGGAGACTCACTAAGGGATCATTGGTGATTGCAAAGGGAGTTG 1880
ST-3/3          ATTTTGCAAATCAAAAGTGGAGACTCACTAAGGGATCATTGGTGATTGCAAAGGGAGTTG 1877
ST-3/4          ATTTTGCAAATCAAAAGTGGAGACTCACTAAGGGATCATTGGTGATTGCAAAGGGAGTTG 1871
ST-1/4          ATTTTGCAAATCAAAAGTGGAGACTCACTAAGGGATCATTGGTGATTGCAAAGGGAGTTG 1822
ST-4/4          ATTTTGCAAATCAAAAGTGGAGACTCACTAAGGGATCATTGGTGATTGCAAAGGGAGTTG 1912
                ************************************************************
```

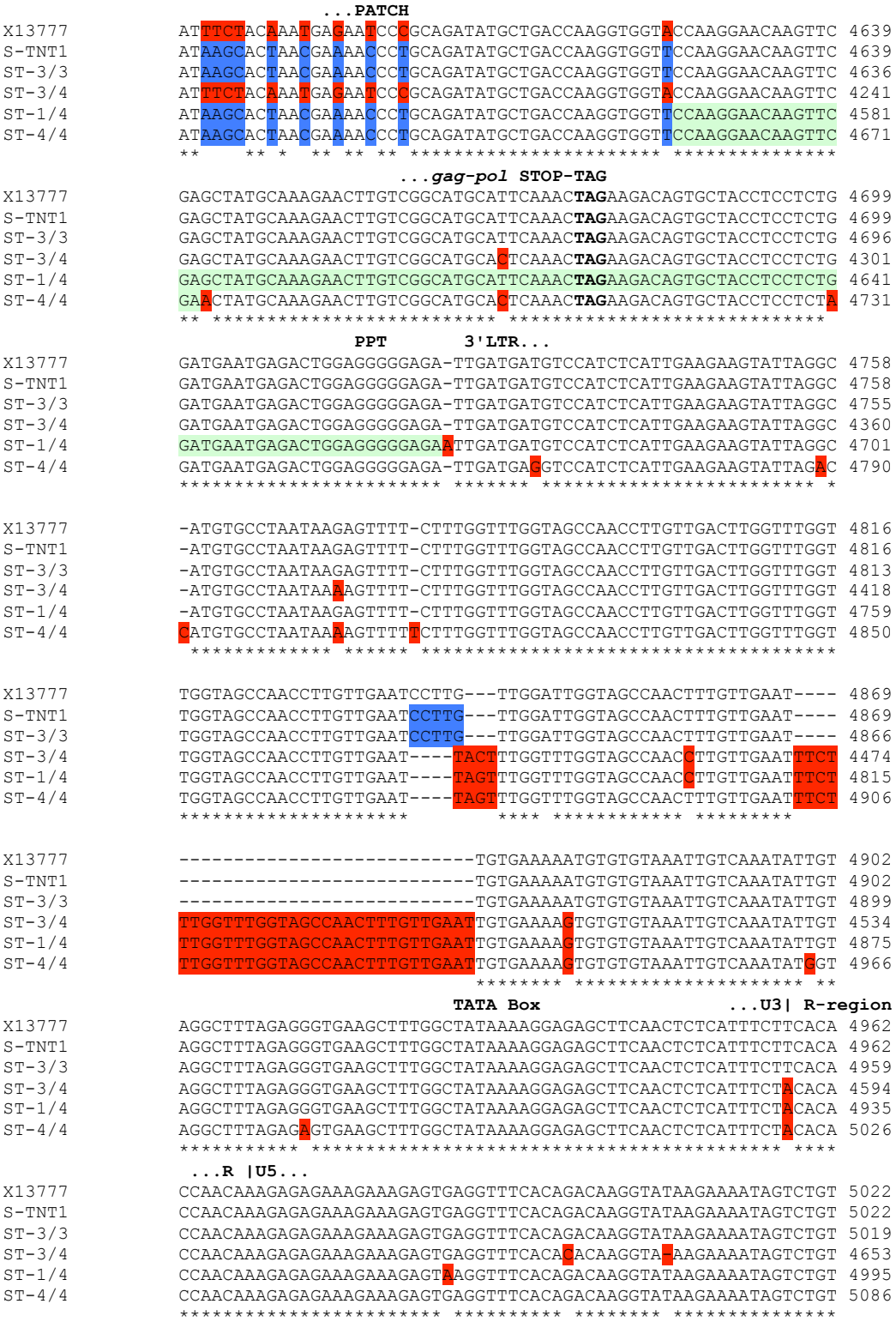
```
X13777          CTCGTGGCACGTTGTACAGGACAAATGCAGAAATATGCCAAGGTGAATTGAACGCGGCAC 1940
S-TNT1          CTCGTGGCACGTTGTACAGGACAAATGCAGAAATATGCCAAGGTGAATTGAACGCGGCAC 1940
ST-3/3          CTCGTGGCACGTTGTACAGGACAAATGCAGAAATATGCCAAGGTGAATTGAACGCGGCAC 1937
ST-3/4          CTCGTGGCACGTTGTACAGGACAAATGCAGAAATATGCCAAGGTGAATTGAACGCAGCAC 1931
ST-1/4          CTCGTGGCACGTTGTACAGGACAAATGCAGAAATATGCCAAGGTGAATTGAACGCGGCAC 1882
ST-4/4          CTCGTGGCACGTTGTACAGGACAAATGCAGAAATATGCCAAGGTGAATTGAACGCGGCAC 1972
                ********************************************************* ****

X13777          AAGATGAGATTTCTGTAGATTTATGGCACAAAAGAATGGGTCATATGAGCGAGAAGGGAT 2000
S-TNT1          AAGATGAGATTTCTGTAGATTTATGGCACAAAAGAATGGGTCACCATGAGCGAGAAGGGAT 2000
ST-3/3          AAGATGAGATTTCTGTAGATTTATGGCACAAAAGAATGGGTCACCATGAGCGAGAAGGGAT 1997
ST-3/4          AAGATGAGATTTCTGTAGATTTATGGCACAAAAGAATGGGTCATATGAGCGAGAAGGGAT 1991
ST-1/4          AAGATGAGATTTCTGTAGATTTATGGCACAAAAGAATGGGTCACCATGAGCGAGAAGGGAT 1942
ST-4/4          AAGATGAGATTTCTGTAGATTTATGGCACAAAAGAATGGGTCACCATGAGCGAGAAGGGAT 2032
                ****************************************** ****************

X13777          TGCAGATTCTTGCCAAGAAATCACTCATTTCTTATGCCAAAGGTACAACTGTAAAACCTT 2060
S-TNT1          TGCAGATTCTTGCCAAGAAATCACTCATTTCTTATGCCAAAGGTACAACTGTAAAACCTT 2060
ST-3/3          TGCAGATTCTTGCCAAGAAATCACTCATTTCTTATGCCAAAGGTACAACTGTAAAACCTT 2057
ST-3/4          TGCAGATTCTTGCCAAGAAATCACTCATTTCTTATGCCAAAGGTACAACTGTAAAACCTT 2051
ST-1/4          TGCAGATTCTTGCCAAGAAATCACTCATTTCTTATGCCAAAGGTACAACTGTAAAACCTT 2002
ST-4/4          TGCAGATTCTTGCCAAGAAATCACTCATTTCTTATGCCAAAGGTACAACTGTAAAACCTT 2092
                ************************************************************

X13777          GTGACTACTGTTTATTTGGTAAGCAGCATAGAGTCTCATTTCAGACATCGTCTGAAAGAA 2120
S-TNT1          GTGACTACTGTTTATTTGGTAAGCAGCATAGAGTCTCATTTCAGACTAGTTCTGAAAGAA 2120
ST-3/3          GTGACTACTGTTTATTTGGTAAGCAGCATAGAGTCTCATTTCAGACTAGTTCTGAAAGAA 2117
ST-3/4          GTGACTACTGTTTATTTGGTAAGCAGCATAGAGTCTCATTTAAGACATCGTCTGAAAGAA 2111
ST-1/4          GTGACTACTGTTTATTTGGTAAGCAGCATAGAGTCTCATTTCAGACTAGTTCTGAAAGAA 2062
ST-4/4          GTGACTACTGTTTATTTGGTAAGCAGCATAGAGTCTCATTTCAGACTAGTTCTGAAAGAA 2152
                ************************************** ****    **********

X13777          AATTGAATATACTTGATTTAGTATATTCTGATGTTTGCGGCCCAATGGAAATTGAATCAA 2180
S-TNT1          AATTGAATATACTTGATTTAGTATATTCTGATGTTTGCGGCCCAATGGAAATTGAATCAA 2180
ST-3/3          AATTGAATATACTTGATTTAGTATATTCTGATGTTTGCGGCCCAATGGAAATTGAATCAA 2177
ST-3/4          AATTGAATATACTTGATTTAGTATATTCTGATGTTTGCGGCCCAATGGAAATTGAATCAA 2171
ST-1/4          AATTGAATATACTTGATTTAGTATATTCTGATGTTTGCGGCCCAATGGAAATTGAATCAA 2122
ST-4/4          AATTGAATATACTTGATTTAGTATATTCTGATGTTTGCGGCCCAATGGAAATTGAATCAA 2212
                ************************************************************

X13777          TGGGCGGTAACAAATATTTTGTTACTTTTATTGATGATGCTTCACGAAAATTATGGGTTT 2240
S-TNT1          TGGGCGGTAACAAATATTTTGTTACTTTTATTGATGATGCTTCACGAAAATTATGGGTTT 2240
ST-3/3          TGGGCGGTAACAAATATTTTGTTACTTTTATTGATGATGCTTCACGAAAATTATGGGTTT 2237
ST-3/4          TGGGCGGTAACAAATATTTTGTTACTTTTATTGATGATGCTTCACGAAAATTATGGGTTT 2231
ST-1/4          TGGGCGGTAACAAATATTTTGTTACTTTTATTGATGATGCTTCACGAAAATTATGGGTTT 2182
ST-4/4          TGGGCGGTAACAAATATTTTGTTACTTTTATTGATGATGCTTCACGAAAATTATGGGTTT 2272
                ************************************************************

X13777          ATATTTTGAAAACCAAAGATCAGGTGTTTCAAGTTTTCCAGAAGTTTCATGCTCTAGTAG 2300
S-TNT1          ATATTTTGAAAACCAAAGATCAGGTGTTTCAAGTTTTCCAGAAGTTTCATGCTCTAGTAG 2300
ST-3/3          ATATTTTGAAAACCAAAGATCAGGTGTTTCAAGTTTTCCAGAAGTTTCATGCTCTAGTAG 2297
ST-3/4          ATATTTTGAAAACCAAAGATCAGGTGTTTCAAGTTTTCCAGAAGTTTCATGCTCTAGTAG 2291
ST-1/4          ATATTTTGAAAACCAAAGATCAGGTGTTTCAAGTTTTCCAGAAGTTTCATGCTCTAGTAG 2242
ST-4/4          ATATTTTGAAAACCAAAGATCAGGTGTTTCAAGTTTTCCAGAAGTTTCATGCTCTAGTAG 2332
                ************************************************************

X13777          AAAGGGAGACGGGTCGAAAGCTAAAGCGTCTCCGAAGTGACAATGGAGGTGAGTACACTT 2360
S-TNT1          AAAGGGAGACGGGTCGAAAGCTAAAGCGTCTCCGAAGTGACAATGGAGGTGAGTACACTT 2360
ST-3/3          AAAGGGAGACGGGTCGAAAGCTAAAGCGTCTCCGAAGTGACAATGGAGGTGAGTACACTT 2357
ST-3/4          AAAGGGAGACGGGTCGAAAGCTAAAGCGTCTCCGAAGTGACAATGGAGGTGAGTACACTT 2351
ST-1/4          AAAGGGAGACGGGTCGAAAGCTAAAGCGTCTCCGAAGTGACAATGGAGGTGAGTACACTT 2302
ST-4/4          AAAGGGAGACGGGTCGAAAGCTAAAGCGTCTCCGAAGTGACAATGGAGGTGAGTACACTT 2392
                ************************************************************

X13777          CAAGGGAATTTGAAGAGTATTGTTCAAGTCATGGGATCAGACATGAAAAGACAGTTCCTG 2420
S-TNT1          CAAGGGAATTTGAAGAGTATTGTTCAAGTCATGGGATCAGACATGAAAAGACAGTTCCTG 2420
ST-3/3          CAAGGGAATTTGAAGAGTATTGTTCAAGTCATGGGATCAGACATGAAAAGACAGTTCCTG 2417
ST-3/4          CAAGGGAATTTGAAGAGTATTGTTCAAGTCATGGGATCAGACATGAAAAGACAGTTCCTG 2411
ST-1/4          CAAGGGAATTTGAAGAGTATTGTTCAAGTCATGGGATCAGACATGAAAAGACAGTTCCTG 2362
ST-4/4          CAAGGGAATTTGAAGAGTATTGTTCAAGTCATGGGATCAGACATGAAAAGACAGTTCCTG 2452
                ************************************************************
```

```
X13777          GAACCCCACAGCACAATGGCGTAGCCGAGAGGATGAACCGCACCATTGTGGAGAAGGTGA 2480
S-TNT1          GAACCCCACAGCACAATGGCGTAGCCGAGAGGATGAACCGCACCATTGTGGAGAAGGTGA 2480
ST-3/3          GAACCCCACAGCACAATGGCGTAGCCGAGAGGATGAACCGCACCATTGTGGAGAAGGTGA 2477
ST-3/4          GAACCCCACAGCACAATGGCGTAGCCGAGAGGATGAACCGCACCATTGTGGAGAAGGTGA 2471
ST-1/4          GAACCCCACAGCACAATGGCGTAGCCGAGAGGATGAACCGCACCATTGTGGAGAAGGTGA 2422
ST-4/4          GAACCCCACAGCACAATGGCGTAGCCGAGAGGATGAACCGCACCATTGTGGAGAAGGTGA 2512
                ************************************************************

X13777          GAAGCATGCTCAGAATGGCTAAACTGCCTAAGTCATTCTGGGGTGAAGCAGTTCAGACAG 2540
S-TNT1          GAAGTATGCTCAGAATGGCTAAACTGCCTAAGTCATTCTGGGGTGAAGCAGTTCAGACAG 2540
ST-3/3          GAAGTATGCTCAGAATGGCTAAACTGCCTAAGTCATTCTGGGGTGAAGCAGTTCAGACAG 2537
ST-3/4          GAAGTATGCTCAGAATGGCTAAACTGCCTAAGTCATTCTGGGGTGAAGCAGTTCAGACAG 2531
ST-1/4          GAAGTATGCTCAGAATGGCTAAACTGCCTAAGTCATTCTGGGGTGAAGCAGTTCAGACAG 2482
ST-4/4          GAAGTATGCTCAGAATGGCTAAACTGCCTAAGTCATTCTGGGGTGAAGCAGTTCAGACAG 2572
                **** ********************************************************

X13777          CCTGTTACCTGATCAATAGGAGTCCATCAGTTCCGTTGGCGTTTGAAATCCCAGAGAGAG 2600
S-TNT1          CCTGTTACCTGATCAATAGGAGTCCATCAGTTCCGTTGGCGTTTGAAATCCCAGAACGCG 2600
ST-3/3          CCTGTTACCTGATCAATAGGAGTCCATCAGTTCCGTTGGCGTTTGAAATCCCAGAACGCG 2597
ST-3/4          CCTGTTACCTGATCAATAGGAGTCCATCAGTTCCGTTGGCGTTTGAAATCCCAGAGAGAG 2591
ST-1/4          CCTGTTACCTGATCAATAGGAGTCCATCAGTTCCGTTGGCGTTTGAAATCCCAGAACGCG 2542
ST-4/4          CCTGTTACCTGATCAATAGGAGTCCATCAGTTCCGTTGGCGTTTGAAATCCCAGAACGCG 2632
                ************************************************************  * *

X13777          TCTGGACCAACAAGGAGGTGTCCTACTCGCATCTGAAGGTGTTCGGTTGCAGAGCTTTTG 2660
S-TNT1          TCTGGACCAACAAGGAGGTGTCCTACTCGCATCTGAAGGTGTTCGGTTGCAGAGCTTTTG 2660
ST-3/3          TCTGGACCAACAAGGAGGTGTCCTACTCGCATCTGAAGGTGTTCGGTTGCAGAGCTTTTG 2657
ST-3/4          TTTGGACCAACAAGGAGGTGTCCTACTCGCATCTGAAGGTGTTCGGTTGCAGAGCTTTTG 2651
ST-1/4          TCTGGACCAACAAGGAGGTGTCCTACTCGCATCTGAAGGTGTTCGGTTGCAGAGCTTTTG 2602
ST-4/4          TCTGGACCAACAAGGAGGTGTCCTACTCGCATCTGAAGGTGTTCGGTTGCAGAGCTTTTG 2692
                * ************************************************************

X13777          CACATGTACCAAAAGAGCAGAGAACAAAGCTGGATGATAAATCTATTCCCTGCATATTTA 2720
S-TNT1          CACATGTACCAAAAGAGCAGAGAACAAAGCTGGATGATAAATCTATTCCCTGCATATTTA 2720
ST-3/3          CACATGTACCAAAAGAGCAGAGAACAAAGCTGGATGATAAATCTATTCCCTGCATATTTA 2717
ST-3/4          CACATGTACCAAAAGAGCAGAGAACAAAGCTGGATGATAAATCTCTTCCCTGCATATTTA 2711
ST-1/4          CACATGTACCAAAAGAGCAGAGAACAAAGCTGGATGATAAATCTATTCCCTGCATATTTA 2662
ST-4/4          CACATGTACCAAAAGAGCAGAGAACAAAGCTGGATGATAAATCTATTCCCTGCATATTTA 2752
                ******************************************** ***************

X13777          TCGGATATGGAGATGAAGAGTTCGGGTACAGACTGTGGGATCCTGTAAAGAAGAAGGTCA 2780
S-TNT1          TCGGATATGGAGATGAAGAGTTCGGGTACAGACTGTGGGACCCTGTAAAGAAGAAGGTCA 2780
ST-3/3          TCGGATATGGAGATGAAGAGTTCGGGTACAGACTGTGGGACCCTGTAAAGAAGAAGGTCA 2777
ST-3/4          TCGGATATGGAGATGAAGAGTTCGGGTACAGACTGTGGGATCCTATAAAGAAGAAGGTCA 2771
ST-1/4          TCGGATATGGAGATGAAGAGTTCGGGTACAGACTGTGGGATCCTGTAAAGAAGAAGGTCA 2722
ST-4/4          TCGGATATGGAGATGAAGAGTTCGGGTACAGACTGTGGGACCCTGTAAAGAAGAAGGTCA 2812
                ****************************************** *** ***************

X13777          TCAGAAGTAGAGATGTAGTCTTCCGAGAAAGTGAAGTTAGAACTGCTGCTGATATGTCAG 2840
S-TNT1          TCAGAAGTAGAGATGTAGTCTTCCGAGAAAGTGAAGTTAGAACTGCTGCTGATATGTCAG 2840
ST-3/3          TCAGAAGTAGAGATGTAGTCTTCCGAGAAAGTGAAGTTAGAACTGCTGCTGATATGTCAG 2837
ST-3/4          TCAGAAGCAGAGATGTAGTCTTCCGAGAAAGTGAAGTTGGAACTGCTGCTGATATGTTAG 2831
ST-1/4          TCAGAAGTAGAGATGTAGTCTTCCGAGAAAGTGAAGTTGGAACTGCTGCTGATATGTCAG 2782
ST-4/4          TCAGAAGTAGAGATGTAGTCTTCCGAGAAAGTGAAGTTAGAACTGCTGCTGATATGTCAG 2872
                ******* ****************************** ***************** **

X13777          AAAAGGTGAAGAATGGTATAATTCCTAACTTTGTTACTATTCCTTCTACTTCTAACAATC 2900
S-TNT1          AAAAGGTGAAGAATGGTATAATTCCTAACTTTGTTACTATTCCTTCTACTTCTAACAATC 2900
ST-3/3          AAAAGGTGAAGAATGGTATAATTCCTAACTTTGTTACTATTCCTTCTACTTCTAACAATC 2897
ST-3/4          AAAAGGCGAAGAATGGTATAATTCCTAACTTTGTTACTATTCCTTCTACTTCTAACAATC 2891
ST-1/4          AAAAGGCGAAGAATGGTATAATTCCTAACTTTGTTACTATTCCTTCTACTTCTAACAATC 2842
ST-4/4          AAAAGGTGAAGAATGGTATAATTCCTAACTTTGTTACTATTCCTTCTACTTCTAACAATC 2932
                ****** ******************************************************

X13777          CCACAAGTGCAGAAAGTACGACCGACGAGGTTTCCGAGCAGGGGGAGCAACCTGGTGAGG 2960
S-TNT1          CCACAAGTGCAGAAAGTACGACCGACGAGGTTTCCGAGCAGGGGGAGCAACCTGGTGAGG 2960
ST-3/3          CCACAAGTGCAGAAAGTACGACCGACGAGGTTTCCGAGCAGGGGGAGCAACCTGGTGAGG 2957
ST-3/4          CCACAAGTGCAGAAAGTACAACCGACGAGGTTGCCGAGCAGGTGGAGCAACCTGGTGAGG 2951
ST-1/4          CCACAAGTGCAGAAAGTACGACCGACGAGGTTGCCGAGCAGGGGGAGCAACCTGGTGAGG 2902
ST-4/4          CCACAAGTGCAGAAAGTACGACCGACGAGGTTTCCGAGCAGGGGGAGCAACCTGGTGAGG 2992
                ****************** ************ ********* *****************
```


X13777 TTATTGAGCAGGGGGAGCAACTTGATGAAGGTGTCGAGGAAGTGGAGCACCCCACTCAGG 3020
S-TNT1 TTATTGAGCAGGGGGAGCAACTTGATGAAGGTGTCGAGGAAGTGGAGCACCCCACTCAGG 3020
ST-3/3 TTATTGAGCAGGGGGAGCAACTTGATGAAGGTGTCGAGGAAGTGGAGCACCCCACTCAGG 3017
ST-3/4 TTATTGAGCAGGGGGAGCAACTTGATGAAGGTGTCGAGGAAGTGGAGCACCCCACTCAGG 3011
ST-1/4 TTATTGAGCAGGGGGAGCAACTTGATGAAGGTGTCGAGGAAGTGGAGCACCCCACTCAGG 2962
ST-4/4 TTATTGAGCAGGGGGAGCAACTTGATGAAGGTGTCGAGGAAGTGGAGCACCCCACTCAGG 3052
X13777 GAGAAGAACAACATCAACCTCTGAGGAGATCAGAGAGGCCAAGGGTAGAGTCACGCAGGT 3080
S-TNT1 GAGAAGAACAACATCAACCTCTGAGGAGATCAGAGAGACCTCGAGTAGAGTCACGCAGAT 3080
ST-3/3 GAGAAGAACAACATCAACCTCTGAGGAGATCAGAGAGACCTCGAGTAGAGTCACGCAGAT 3077
ST-3/4 GAGAAGAACAACCTCAACCTCTGAGGAGATCAGAGAGGCCAAGGGTAGAGTCATGCAGGT 3071
ST-1/4 GAGAAGAACAACCTCAACCTCTGAGGAGATCAGAGAGGCCAAGGGTAGAGTCACGCAGGT 3022
ST-4/4 GAGAAGAACAACATCAACCTCTGAGGAGATCAGAGAGACCTCGAGTAGAGTCACGCAGAT 3112
X13777 ACCCTTCCACAGAGTATGTCCTCATCAGTGATGATAGGGAGCCAGAAAGTCTTAAGGAGG 3140
S-TNT1 ACCCTTCCACAGAGTATGTCCTCATCAGTGATGATAGGGAGCCAGAAAGTCTTAAGGAGG 3140
ST-3/3 ACCCTTCCACAGAGTATGTCCTCATCAGTGATGATAGGGAGCCAGAAAGTCTTAAGGAGG 3137
ST-3/4 ACCCTTCCACAGAGTATGTCCTCATCAGTGATGAGGGGGAGCCAGAAAGTCTTAAGGAGG 3131
ST-1/4 ACCCTTCCACAGAGTATGTCCTCATCAGTGATGAGGGGGAGCCAGAAAGTCTTAAGGAGG 3082
ST-4/4 ACCCTTCCACAGAGTATGTCCTCATCAGTGATGATAGGGAGCCAGAAAGTCTTAAGGAGG 3172
X13777 TGTTGTCCCATCCAGAAAAGAACCAGTTGATGAAAGCTATGCAAGAAGAGATGGAATCTC 3200
S-TNT1 TGTTGTCCCATCCAGAAAAGAACCAGTTGATGAAAGCTATGCAAGAAGAGATGGAATCTC 3200
ST-3/3 TGTTGTCCCATCCAGAAAAGAACCAGTTGATGAAAGCTATGCAAGAAGAGATGGAATCTC 3197
ST-3/4 TGTTGTCCCATCCAGAAAAGAACCAGTGGATGAAAGCTATGCAAGAAGAGATGGAATCTC 3191
ST-1/4 TGTTGTCCCATCCAGAAAAGAACCAGTGGATGAAAGCTATGCAAGAAGAGATGGAATCTC 3142
ST-4/4 TGTTGTCCCATCCAGAAAAGAACCAGTTGATGAAAGCTATGCAAGAAGAGATGGAATCTC 3232
X13777 TCCAGAAAAATGGCACATACAAGCTGGTTGAACTTCCAAAGGGTAAAAGACCACTCAAAT 3260
S-TNT1 TCCAGAAAAATGGCACATACAAGCTGGTTGAACTTCCAAAGGGTAAAAGACCACTCAAAT 3260
ST-3/3 TCCAGAAAAATGGCACATACAAGCTGGTTGAACTTCCAAAGGGTAAAAGACCACTCAAAT 3257
ST-3/4 TACAGAAAAATGGCACATACAAGCTGGTTGAACTTCCAAAAGGTAAAAGACCACTCAAAT 3251
ST-1/4 TCCAGAAAAATGGCACATACAAGCTGGTTGAACTTCCAAAGGGTAAAAGACCACTCAAAT 3202
ST-4/4 TCCAGAAAAATGGCACATACAAGCTGGTTGAACTTCCAAAGGGTAAAAGACCACTCAAAT 3292
X13777 GCAAATGGGTCTTTAAACTCAAGAAAGATGGAGATTGCAAGCTGGTCAGATACAAAGCTC 3320
S-TNT1 GCAAATGGGTCTTTAAACTCAAGAAAGATGGAGATTGCAAGCTGGTCAGATACAAAGCTC 3320
ST-3/3 GCAAATGGGTCTTTAAACTCAAGAAAGATGGAGATTGCAAGCTGGTCAGATACAAAGCTC 3317
ST-3/4 GCAAATGGGTCTTTAAACTCAAGAAAGATGGAAATGGCAAGCTGGTTAGATACAAAGCTC 3311
ST-1/4 GCAAATGGGTCTTTAAACTCAAGAAAGATGGAGATTGCAAGCTGGTCAGATACAAAGCTC 3262
ST-4/4 GCAAATGGGTCTTTAAACTCAAGAAAGATGGAGATTGCAAGCTGGTCAGATACAAAGCTC 3352
X13777 GATTGGTGGTTAAAGGCTTCGAACAGAAGAAAGGTATTGATTTTGACGAAATTTTCTCCC 3380
S-TNT1 GATTGGTGGTTAAAGGCTTCGAACAGAAGAAAGGTATTGATTTTGACGAAATTTTCTCCC 3380
ST-3/3 GATTGGTGGTTAAAGGCTTCGAACAGAAGAAAGGTATTGATTTTGACGAAATTTTCTCCC 3377
ST-3/4 GATTGGTGGTTAAAGGCTTCGAACAAAAGAAAGGTATTGATTTTGACGAAATTTTCTCAC 3371
ST-1/4 GATTGGTGGTTAAAGGCTTCGAACAGAAGAAAGGTATTGATTTTGACGAAATTTTCTCCC 3322
ST-4/4 GATTGGTGGTTAAAGGCTTCGAACAGAAGAAAGGTATTGATTTTGACGAAATTTTCTCCC 3412
X13777 CCGTTGTTAAAATGACTTCTATTCGAACAATTTTGAGCTTAGCAGCTAGCCTAGATCTTG 3440
S-TNT1 CCGTTGTTAAAATGACTTCTATTCGAACAATTTTGAGCTTAGCAGCTAGCCTCGATCTTG 3440
ST-3/3 CCGTTGTTAAAATGACTTCTATTCGAACAATTTTGAGCTTAGCAGCTAGCCTCGATCTTG 3437
ST-3/4 CTGTTGTCAAAATGACTTCTATTCAAACAATTTTGAGCTTAGCAGCTAGCCTAGATCTTG 3431
ST-1/4 CCGTTGTTAAAATGACTTCTATTCGAACAATTTTGAGCTTAGCAGCTAGCCTCGATCTTG 3382
ST-4/4 CCGTTGTTAAAATGACTTCTATTCGAACAATTTTGAGCTTAGCAGCTAGCCTCGATCTTG 3472
X13777 AAGTGGAGCAGTTGGATGTGAAAACTGCATTTCTTCATGGAGATTTGGAAGAGGAGATTT 3500
S-TNT1 AAGTGGAGCAGTTGGATGTGAAAACTGCATTTCTTCATGGAGATTTGGAAGAGGAGATTT 3500
ST-3/3 AAGTGGAGCAGTTGGATGTGAAAACTGCATTTCTTCATGGAGATTTGGAAGAGGAGATTT 3497
ST-3/4 AAGTGGAGCAGTTGGATGTGAAAACTGCATTTCTTCATGGAGATTTGGAAGAGGAGATTT 3491
ST-1/4 AAGTGGAGCAGTTGGATGTGAAAACTGCATTTCTTCATGGAGATTTGGAAGAGGAGATTT 3442
ST-4/4 AAGTGGAGCAGTTGGATGTGAAAACTGCATTTCTTCATGGAGATTTGGAAGAGGAGATTT 3532


X13777 ATATGGAGCAACCAGAAGGATTTGAAGTAGCTGGAAAGAAACACATGGTGTGCAAATTGA 3560
S-TNT1 ATATGGAGCAACCAGAAGGATTTGAAGTAGCTGGAAAGAAACACATGGTGTGCAAATTGA 3560
ST-3/3 ATATGGAGCAACCAGAAGGATTTGAAGTAGCTGGAAAGAAACACATGGTGTGCAAATTGA 3557
ST-3/4 ATATGGAGCAACCAGAAGGATTTGAAGTAGCTGGAAAGAAACACATGGTGTGCAAATTGA 3551
ST-1/4 ATATGGAGCAACCAGAAGGATTTGAAGTAGCTGGAAAGAAACACATGGTGTGCAAATTGA 3502
ST-4/4 ATATGGAGCAACCAGAAGGATTTGAAGTAGCTGGAAAGAAACACATGGTGTGCAAATTGA 3592
X13777 ATAAGAGTCTTTATGGATTGAAGCAGGCACCAAGGCAGTGGTACATGAAGTTTGATTCAT 3620
S-TNT1 ATAAGAGTCTTTATGGATTGAAGCAGGCACCAAGGCAGTGGTACATGAAGTTTGATTCAT 3620
ST-3/3 ATAAGAGTCTTTATGGATTGAAGCAGGCACCAAGGCAGTGGTACATGAAGTTTGATTCAT 3617
ST-3/4 ATAAGAGTCTTTATGGATTGAAGCAGGCACCAAGGCAGTGGTACATGAAGTTTGATTCAT 3611
ST-1/4 ATAAGAGTCTTTATGGATTGAAGCAGGCACCAAGGCAGTGGTACATGAAGTTTGATTCAT 3562
ST-4/4 ATAAGAGTCTTTATGGATTGAAGCAGGCACCAAGGCAGTGGTACATGAAGTTTGATTCAT 3652
X13777 TCATGAAAAGTCAAACATACCTAAAGACCTATTCTGATCCATGTGTATACTTCAAAAGAT 3680
S-TNT1 TCATGAAAAGTCAAACATACCTAAAGACCTATTCTGATCCATGTGTTTACTTCAAAAGAT 3680
ST-3/3 TCATGAAAAGTCAAACATACCTAAAGACCTATTCTGATCCATGTGTTTACTTCAAAAGAT 3677
ST-3/4 TCATGAAAAGTCAAACATACCTAAAGACCTATTCTGATCCATGTGTTTACTTCAAAAGAT 3671
ST-1/4 TCATGAAAAGTCAAACATACCTAAAGACCTATTCTGATCCATGTGTTTACTTCAAAAGAT 3622
ST-4/4 TCATGAAAAGTCAAACATACCTAAAGACCTATTCTGATCCATGTGTTTACTTCAAAAGAT 3712
X13777 TTTCTGAGAATAACTTTATTATATTGTTGTTGTATGTGGATGACATGCTAATTGTAGGAA 3740
S-TNT1 TTTCTGAGAATAACTTTATTATATTGTTGTTGTATGTGGATGACATGCTAATTGTAGGAA 3740
ST-3/3 TTTCTGAGAATAACTTTATTATATTGTTGTTGTATGTGGATGACATGCTAATTGTAGGAA 3737
ST-3/4 TTTCTGAGAATAACTTTATTATATTGTTGTTGTATGTGGATGACATGCTAATTGTAGGAA 3731
ST-1/4 TTTCTGAGAATAACTTTATTATATTGTTGTTGTATGTGGATGACATGCTAATTGTAGGAA 3682
ST-4/4 TTTCTGAGAATAACTTTATTATATTGTTGTTGTATGTGGATGACATGCTAATTGTAGGAA 3772
PATCH...
X13777 AAGACAAGGGGTTGATAGCAAAGTTGAAAGGAGATCTGTCCAAGTCATTTGATATGAAGG 3800
S-TNT1 AAGACAAGGGGTTGATAGCAAAGTTGAAAGGTGACCTGAGTAAGTCTTTCGACATGAAGG 3800
ST-3/3 AAGACAAGGGGTTGATAGCAAAGTTGAAAGGTGACCTGAGTAAGTCTTTCGACATGAAGG 3797
ST-3/4 AAGACAAGGGGTTGATAGCAAAGTTGAAAGGTGACCTGAGTAAGTCTTTCGACATGAAGG 3791
ST-1/4 AAGACAAGGGGTTGATAGCAAAGTTGAAAGGTGACCTGAGTAAGTCTTTCGACATGAAGG 3742
ST-4/4 AAGACAAGGGGTTGATAGCAAAGTTGAAAGGTGACCTGAGTAAGTCTTTCGACATGAAGG 3832
3U3F1-GGCCCAGCTCAGCAGATATTGGGT
X13777 ACTTGGGCCCAGCACAACAAATTCTAGGGATGAAGATAGTTCGAGAGAGAACAAGTAGAA 3860
S-TNT1 ATCTGGGCCCAGCTCAGCAGATATTGGGTATGAAAATTGTGAGGGAAAGGACTTCCAGGA 3860
ST-3/3 ATCTGGGCCCAGCTCAGCAGATATTGGGTATGAAAATTGTGAGGGAAAGGACTTCCAGGA 3857
ST-3/4 ATCTGGGCCCAGCTCAGCAGATATTGGGTATGAAAATTGTGAGGGAAAGGACTTCCAGGA 3851
ST-1/4 ATCTGGGCCCAGCTCAGCAGATATTGGGTATGAAAATTGTGAGGGAAAGGACTTCCAGGA 3802
ST-4/4 ATCTGGGCCCAGCTCAGCAGATATTGGGTATGAAAATTGTGAGGGAAAGGACTTCCAGGA 3892
X13777 AGTTGTGGGTATCTCAGGAGAAGTACATTGAACGTGTACTAGAACGCTTCAACATGAAGA 3920
S-TNT1 AACTTTGGTTGTCACAAGAGAAATATATCGAGAGAGTTTTAGAGAGGTTTAATATGAAGA 3920
ST-3/3 AACTTTGGTTGTCACAAGAGAAATATATCGAGAGAGTTTTAGAGAGGTTTAATATGAAGA 3917
ST-3/4 AACTTTGGTTGTCACAAGAGAAATATATCGAGAGAGTTTTAGAGAGGTTTAATATGAAGA 3911
ST-1/4 AACTTTGGTTGTCACAAGAGAAATATATCGAGAGAGTTTTAGAGAGGTTTAATATGAAGA 3862
ST-4/4 AACTTTGGTTGTCACAAGAGAAATATATCGAGAGAGTTTTAGAGAGGTTTAATATGAAGA 3952
X13777 ATGCTAAGCCAGTCAGCACACCTCTTGCTGGTCATCTAAAGTTGAGTAAAAAGATGTGTC 3980
S-TNT1 ACGCAAAACCTGTATCTACCCCATTGGCCGGACACTTGAAGCTATCTAAGAAAATGTGCC 3980
ST-3/3 ACGCAAAACCTGTATCTACCCCATTGGCCGGACACTTGAAGCTATCTAAGAAAATGTGCC 3977
ST-3/4 ACGCAAAACCTGTATCTACCCCATTGGCCGGACACTTGAAGCTATCTAAGAAAATGTGCC 3971
ST-1/4 ACGCAAAACCTGTATCTACCCCATTGGCCGGACACTTGAAGCTATCTAAGAAAATGTGCC 3922
ST-4/4 ACGCAAAACCTGTATCTACCCCATTGGCCGGACACTTGAAGCTATCTAAGAAAATGTGCC 4012
ATATGGC
X13777 CTACAACAGTGGAAGAGAAAGGGAACATGGCTAAAGTTCCTTATTCTTCAGCAGTCGGAA 4040
S-TNT1 CAACTACCGTAGAGGAAAAGGGAAATATGGCCAAGGTGCCCTACAGTTCTGCGGTTGGTT 4040
ST-3/3 CAACTACCGTAGAGGAAAAGGGAAATATGGCCAAGGTGCCCTACAGTTCTGCGGTTGGTT 4037
ST-3/4 CAACTACCGTAGAGGAAAAGGGAAATATGGC---------- 4002
ST-1/4 CAACTACCGTAGAGGAAAAGGGAAATATGGCCAAGGTGCCCTACAGTTCTGCGGTTGGTT 3982
ST-4/4 CAACTACCGTAGAGGAAAAGGGAAATATGGCCAAGGTGCCCTACAGTTCTGCGGTTGGTT 4072


X13777 GCTTGATGTATGCAATGGTATGTACTAGACCTGATATTGCTCACGCAGTTGGTGTTGT-C 4099
S-TNT1 CCCTAATGTACGCTATGGTGTGCACAAGGCCGGATATAGCACATGCTGTAGGAGTCGT-G 4099
ST-3/3 CCCTAATGTACGCTATGGTGTGCACAAGGCCGGATATAGCACATGCTGTAGGAGTCGT-G 4096
ST-3/4 ---------- 4002
ST-1/4 CCCTAATGTACGCTATGGTGTGCACAAGGCCGGATATAGCACATGCTGTAGGAGTCGT-G 4041
ST-4/4 CCCTAATGTACGCTATGGTGTGCACAAGGCCGGATATAGCACATGCTGTAGGAGTCGT-G 4131
X13777 AGCAGGTTTCTTGAAAATCCTGGAAAGGAACATTGGGAAGCAGTCAAGTGGATACTCAGG 4159
S-TNT1 TCACGTTTCTTAGAGAACCCTGGTAAGGAGCACTGGGAGGCTGTTAAATGGATTTTGAGG 4159
ST-3/3 TCACGTTTCTTAGAGAACCCTGGTAAGGAGCACTGGGAGGCTGTTAAATGGATTTTGAGG 4156
ST-3/4 ---------- 4002
ST-1/4 TCACGTTTCTTAGAGAACCCTGGTAAGGAGCACTGGGAGGCTGTTAAATGGATTTTGAGG 4101
ST-4/4 TCACGTTTCTTAGAGAACCCTGGTAAGGAGCACTGGGAGGCTGTTAAATGGATTTTGAGG 4191
X13777 TACCTGAGAGGTACCACGGGAGATTGTTTGTGCTTTGGAGGATCTGATCCAATCTTCGAAG 4219
S-TNT1 TATCTTCGCGGAACTACAGGGGATTGCTTATGTTTCGGCGGTAGCGACCCTATTCTCAAG 4219
ST-3/3 TATCTTCGCGGAACTACAGGGGATTGCTTATGTTTCGGCGGTAGCGACCCTATTCTCAAG 4216
ST-3/4 ---------- 4002
ST-1/4 TATCTTCGCGGAACTACAGGGGATTGCTTATGTTTCGGCGGTAGCGACCCTATTCTCAAG 4161
ST-4/4 TATCTTCGCGGAACTACAGGGGATTGCTTATGTTTCGGCGGTAGCGACCCTATTCTCAAG 4251
X13777 GGCTATACAGATGCTGATATGGCAGGTGACATTGACAACAGAAAATCCAGTACTGGATAT 4279
S-TNT1 GGATACACGGACGCAGACATGGCTGGCGATATTGATAATCGGAAATCGTCTACAGGATAC 4279
ST-3/3 GGATACACGGACGCAGACATGGCTGGCGATATTGATAATCGGAAATCGTCTACAGGATAC 4276
ST-3/4 ---------- 4002
ST-1/4 GGATACACGGACGCAGACATGGCTGGCGATATTGATAATCGGAAATCGTCTACAGGATAC 4221
ST-4/4 GGATACACGGACGCAGACATGGCTGGCGATATTGATAATCGGAAATCGTCTACAGGATAC 4311
X13777 TTGTTTACATTTTCAGGGGGAGCTATATCATGGCAGTCTAAGTTGCAAAAGTGCGTTGCA 4339
S-TNT1 CTATTCACTTTCTCTGGAGGGGCAATCTCTTGGCAAAGCAAACTTCAGAAGTGTGTGGCT 4339
ST-3/3 CTATTCACTTTCTCTGGAGGGGCAATCTCTTGGCAAAGCAAACTTCAGAAGTGTGTGGCT 4336
ST-3/4 ---------- 4002
ST-1/4 CTATTCACTTTCTCTGGAGGGGCAATCTCTTGGCAAAGCAAACTTCAGAAGTGTGTGGCT 4281
ST-4/4 CTATTCACTTTCTCTGGAGGGGCAATCTCTTGGCAAAGCAAACTTCAGAAGTGTGTGGCT 4371
ATATGGC
X13777 CTTTCAACAACTGAAGCAGAGTACATTGCTGCTACAGAAACTGGGAAGGAGATGATATGG 4399
S-TNT1 TTAAGTACCACAGAGGCTGAGTATATCGCAGCAACTGAGACAGGTAAAGAAATGATTTGG 4399
ST-3/3 TTAAGTACCACAGAGGCTGAGTATATCGCAGCAACTGAGACAGGTAAAGAAATGATTTGG 4396
ST-3/4 ---------- 4002
ST-1/4 TTAAGTACCACAGAGGCTGAGTATATCGCAGCAACTGAGACAGGTAAAGAAATGATTTGG 4341
ST-4/4 TTAAGTACCACAGAGGCTGAGTATATCGCAGCAACTGAGACAGGTAAAGAAATGATTTGG 4431
X13777 CTCAAGCGATTCCTTCAAGAGCTTGGATTGCATCAGAAGGAGTATGTCGTCTATTGTGAC 4459
S-TNT1 TTGAAGAGGTTCTTACAGGAATTGGGCCTCCACCAAAAGGAATACGTTGTTTACTGCGAT 4459
ST-3/3 TTGAAGAGGTTCTTACAGGAATTGGGCCTCCACCAAAAGGAATACGTTGTTTACTGCGAT 4456
ST-3/4 -TCAAGCGATTCCTTCAAGAGCTTGGATTGCATCAGAAGGAGTATGTCGTCTATTGTGAC 4061
ST-1/4 TTGAAGAGGTTCTTACAGGAATTGGGCCTCCACCAAAAGGAATACGTTGTTTACTGCGAT 4401
ST-4/4 TTGAAGAGGTTCTTACAGGAATTGGGCCTCCACCAAAAGGAATACGTTGTTTACTGCGAT 4491
X13777 AGTCAAAGTGCAATAGACCTTAGCAAGAACTCTATGTACCATGCAAGGACCAAACACATT 4519
S-TNT1 TCACAGTCAGCCATTGATTTGTCTAAGAATAGTATGTATCACGCGCGTACTAAGCATATA 4519
ST-3/3 TCACAGTCAGCCATTGATTTGTCTAAGAATAGTATGTATCACGCGCGTACTAAGCATATA 4516
ST-3/4 AGTCAAAGTGCAATAGACCTTAGCAAGAACTCTATGTACCATGCAAGGACCAAACACATT 4121
ST-1/4 TCACAGTCAGCCATTGATTTGTCTAAGAATAGTATGTATCACGCGCGTACTAAGCATATA 4461
ST-4/4 TCACAGTCAGCCATTGATTTGTCTAAGAATAGTATGTATCACGCGCGTACTAAGCATATA 4551
CTGCAG-PstI site
X13777 GATGTGAGATATCATTGGATTCGAGAAATGGTAGATGATGAATCTCTAAAAGTCTTCAAG 4579
S-TNT1 GACGTCAGATACCACTGGATCAGAGAGATGGTCGATGACGAGAGTCTGAAAGTTCTTAAG 4579
ST-3/3 GACGTCAGATACCACTGGATCAGAGAGATGGTCGATGACGAGAGTCTGAAAGTTCTTAAG 4576
ST-3/4 GATGTGAGATATCATTGGATTCGAGAAATGGTAGATGATGAATCTCTAAAAGTCTTCAAG 4181
ST-1/4 GACGTCAGATACCACTGGATCAGAGAGATGGTCGATGACGAGAGTCTGAAAGTTCTTAAG 4521
ST-4/4 GACGTCAGATACCACTGGATCAGAGAGATGGTCGATGACGAGAGTCTGAAAGTTCTTAAG 4611


...PATCH
X13777 ATTTCTACAAATGAGAATCCGGCAGATATGCTGACCAAGGTGGTACCAAGGAACAAGTTC 4639
S-TNT1 ATAAGCACTAACGAAAACCCTGCAGATATGCTGACCAAGGTGGTTCCAAGGAACAAGTTC 4639
ST-3/3 ATAAGCACTAACGAAAACCCTGCAGATATGCTGACCAAGGTGGTTCCAAGGAACAAGTTC 4636
ST-3/4 ATTTCTACAAATGAGAATCCGGCAGATATGCTGACCAAGGTGGTACCAAGGAACAAGTTC 4241
ST-1/4 ATAAGCACTAACGAAAACCCTGCAGATATGCTGACCAAGGTGGTTCCAAGGAACAAGTTC 4581
ST-4/4 ATAAGCACTAACGAAAACCCTGCAGATATGCTGACCAAGGTGGTTCCAAGGAACAAGTTC 4671
...gag-pol STOP-TAG
X13777 GAGCTATGCAAAGAACTTGTCGGCATGCATTCAAACTAGAAGACAGTGCTACCTCCTCTG 4699
S-TNT1 GAGCTATGCAAAGAACTTGTCGGCATGCATTCAAACTAGAAGACAGTGCTACCTCCTCTG 4699
ST-3/3 GAGCTATGCAAAGAACTTGTCGGCATGCATTCAAACTAGAAGACAGTGCTACCTCCTCTG 4696
ST-3/4 GAGCTATGCAAAGAACTTGTCGGCATGCACTCAAACTAGAAGACAGTGCTACCTCCTCTG 4301
ST-1/4 GAGCTATGCAAAGAACTTGTCGGCATGCATTCAAACTAGAAGACAGTGCTACCTCCTCTG 4641
ST-4/4 GAACTATGCAAAGAACTTGTCGGCATGCACTCAAACTAGAAGACAGTGCTACCTCCTCTA 4731
PPT 3'LTR...
X13777 GATGAATGAGACTGGAGGGGGAGA-TTGATGATGTCCATCTCATTGAAGAAGTATTAGGC 4758
S-TNT1 GATGAATGAGACTGGAGGGGGAGA-TTGATGATGTCCATCTCATTGAAGAAGTATTAGGC 4758
ST-3/3 GATGAATGAGACTGGAGGGGGAGA-TTGATGATGTCCATCTCATTGAAGAAGTATTAGGC 4755
ST-3/4 GATGAATGAGACTGGAGGGGGAGA-TTGATGATGTCCATCTCATTGAAGAAGTATTAGGC 4360
ST-1/4 GATGAATGAGACTGGAGGGGGAGAATTGATGATGTCCATCTCATTGAAGAAGTATTAGGC 4701
ST-4/4 GATGAATGAGACTGGAGGGGGAGA-TTGATGAGGTCCATCTCATTGAAGAAGTATTAGAC 4790
X13777 -ATGTGCCTAATAAGAGTTTT-CTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGT 4816
S-TNT1 -ATGTGCCTAATAAGAGTTTT-CTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGT 4816
ST-3/3 -ATGTGCCTAATAAGAGTTTT-CTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGT 4813
ST-3/4 -ATGTGCCTAATAAAAGTTTT-CTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGT 4418
ST-1/4 -ATGTGCCTAATAAGAGTTTT-CTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGT 4759
ST-4/4 CATGTGCCTAATAAAAGTTTTTCTTTGGTTTGGTAGCCAACCTTGTTGACTTGGTTTGGT 4850
X13777 TGGTAGCCAACCTTGTTGAATCCTTG---TTGGATTGGTAGCCAACTTTGTTGAAT---- 4869
S-TNT1 TGGTAGCCAACCTTGTTGAATCCTTG---TTGGATTGGTAGCCAACTTTGTTGAAT---- 4869
ST-3/3 TGGTAGCCAACCTTGTTGAATCCTTG---TTGGATTGGTAGCCAACTTTGTTGAAT---- 4866
ST-3/4 TGGTAGCCAACCTTGTTGAAT----TACTTTGGTTTGGTAGCCAACGTTGTTGAATTTCT 4474
ST-1/4 TGGTAGCCAACCTTGTTGAAT----TAGTTTGGTTTGGTAGCCAACGTTGTTGAATTTCT 4815
ST-4/4 TGGTAGCCAACCTTGTTGAAT----TAGTTTGGTTTGGTAGCCAACTTTGTTGAATTTCT 4906
X13777 ----------TGTGAAAAATGTGTGTAAATTGTCAAATATTGT 4902
S-TNT1 ----------TGTGAAAAATGTGTGTAAATTGTCAAATATTGT 4902
ST-3/3 ----------TGTGAAAAATGTGTGTAAATTGTCAAATATTGT 4899
ST-3/4 TTGGTTTGGTAGCCAACTTTGTTGAATTGTGAAAAGTGTGTGTAAATTGTCAAATATTGT 4534
ST-1/4 TTGGTTTGGTAGCCAACTTTGTTGAATTGTGAAAAGTGTGTGTAAATTGTCAAATATTGT 4875
ST-4/4 TTGGTTTGGTAGCCAACTTTGTTGAATTGTGAAAAGTGTGTGTAAATTGTCAAATATGGT 4966
TATA Box ...U3| R-region
X13777 AGGCTTTAGAGGGTGAAGCTTTGGCTATAAAAGGAGAGCTTCAACTCTCATTTCTTCACA 4962
S-TNT1 AGGCTTTAGAGGGTGAAGCTTTGGCTATAAAAGGAGAGCTTCAACTCTCATTTCTTCACA 4962
ST-3/3 AGGCTTTAGAGGGTGAAGCTTTGGCTATAAAAGGAGAGCTTCAACTCTCATTTCTTCACA 4959
ST-3/4 AGGCTTTAGAGGGTGAAGCTTTGGCTATAAAAGGAGAGCTTCAACTCTCATTTCTACACA 4594
ST-1/4 AGGCTTTAGAGGGTGAAGCTTTGGCTATAAAAGGAGAGCTTCAACTCTCATTTCTACACA 4935
ST-4/4 AGGCTTTAGAGAGTGAAGCTTTGGCTATAAAAGGAGAGCTTCAACTCTCATTTCTACACA 5026
...R |U5...
X13777 CCAACAAAGAGAGAAAGAAAGAGTGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGT 5022
S-TNT1 CCAACAAAGAGAGAAAGAAAGAGTGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGT 5022
ST-3/3 CCAACAAAGAGAGAAAGAAAGAGTGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGT 5019
ST-3/4 CCAACAAAGAGAGAAAGAAAGAGTGAGGTTTCACACACAAGGTA-AAGAAAATAGTCTGT 4653
ST-1/4 CCAACAAAGAGAGAAAGAAAGAGTAAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGT 4995
ST-4/4 CCAACAAAGAGAGAAAGAAAGAGTGAGGTTTCACAGACAAGGTATAAGAAAATAGTCTGT 5086

```
X13777          GAGGAAAATAGAGAGTGAGCGATATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTT 5082
S-TNT1          GAGGAAAATAGAGAGTGAGCGATATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTT 5082
ST-3/3          GAGGAAAATAGAGAGTGAGCGATATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTT 5079
ST-3/4          GAGGAAAATAGAGAGTGAGCGATATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTT 4713
ST-1/4          GAGGAAAATAGAGAGTGAGCGATATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTT 5055
ST-4/4          GAGGAAAATAGAGAGTGAGCGATATTGTAGTGAGGTGGGAATATCAAAAGAGGGTTATTT 5146
                ************************************************************

X13777          CTTTTGAGTGTTGTAGTGGTCTTTGGAGTA-TTTACCTCCGACCTACAAAGT-GTAAAAT 5140
S-TNT1          CTTTTGAGTGTTGTAGTGGTCTTTGGAGTA-TTTACCTCCGACCTACAAAGT-GTAAAAT 5140
ST-3/3          CTTTTGAGTGTTGTAGTGGTCTTTGGAGTA-TTTACCTCCGACCTACAAAGT-GTAAAAT 5137
ST-3/4          CTTTTGAGTGTTGTAGTGGTCTTTGGAGTAATTTATCTCCGACCTACAAAGT-GTAAAAT 4772
ST-1/4          CTTTTGAGTGTTGTAGTGGTCTTTGGAGTA-TTTACCTCCGACCTACAAAGTTGTAAAAT 5114
ST-4/4          CTTTTGAGTGTTGTAGTGGTCTTTGGAGTA-TTTACCTCCGACCTACAAAGT-GTAAAAT 5204
                ****************************** **** *************** *******
                                  TATCAGTTGCTCCTCTCGGGGTCG-3U3R1
X13777          TCCTTACTATAGTGATATCAGTTGCTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCA 5200
S-TNT1          TCCTTACTATAGTGATATCAGTTGCTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCA 5200
ST-3/3          TCCTTACTATAGTGATATCAGTTGCTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCA 5197
ST-3/4          TCCTTACTATAGTGATATCAGTTGCTCCTCTCGGGGTCGTGGTTTTT--TCCCTTATTCA 4830
ST-1/4          TCCTTACTATAGTGATATCAGTTGCTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCA 5174
ST-4/4          TCCTTACTATAGTGATATCAGTTGCTCCTCTCGGGGTCGTGGTTTTTTTTCCCTTATTCA 5264
                ********************************************  ***********

X13777          GAAGGGTTTTCCACGTAAAAATCTTGGTGTCATTGTTACTCTTTTATTCTTGTTAA-TTA 5259
S-TNT1          GAAGGGTTTTCCACGTAAAAATCTTGGTGTCATTGTTACTCTTTTATTCTTGTTAA-TTA 5259
ST-3/3          GAAGGGTTTTCCACGTAAAAATCTTGGTGTCATTGTTACTCTTTTATTCTTGTTAAATTA 5257
ST-3/4          GAAGGGTTTTCCACGTAAAAATCTTGGTGTCATTGTTACTCTTTTATTCTTGTTAA-TTA 4889
ST-1/4          GAAGGGTTTTCCACGTAAAAATTTTGGTGTCATTGTTACTCTTTTATTCTTGTTAA-TTA 5233
ST-4/4          GAAGGGTTTTCCACGTAAAAATCTTGGTGTCATTGTTACTCTTTTATTCTTGTTAA-TTA 5323
                ********************** ********************************* ***

X13777          CCGTATCTCGGTGCTACATTATTATTCCGCTTTATTACCGTGAATATTATTTTGGTAAGG 5319
S-TNT1          CCGTATCTCGGTGCTACATTATTATTCCGCTTTATTACCGTGAATATTATTTTGGTAAGG 5319
ST-3/3          CCGTATCTCGGTGCTACATTATTATTCCGCTTTATTACCGTGAATATTATTTTGGTAAGG 5317
ST-3/4          CCGTATCTCGGTGCTACATTATTATTCCGCTTTATTACCGTGAATATTATTTTGGTAAGG 4949
ST-1/4          CCGTATCTCGGTGCTACATTATTATTCCGCTTTATTACCGTGAATATTATTTTGGTAAGG 5293
ST-4/4          CCGTATCTCGGTGCTACATTATTATTCCGCTTTATTACCGTGAATATTATTTTGGTAAGG 5383
                ************************************************************
                         ...LTR TSD    3'genomic sequence
X13777          GGTTTATTCCCAACA------------------------------------------------ 5334
S-TNT1          GGTTTATTCCCAACA------------------------------------------------ 5334
ST-3/3          GGTTTATTCCCAACA-CGAAGAGGCCATTATGGAATCAAATTGCAGTGATACCAGAACTA 5332
ST-3/4          GGTTTATTCCCAACA-AGAAAGCGTTGCTATATTTGTGCATCATGCATGTTGAACTGGCT 4964
ST-1/4          GGTTTATTCCCAACA-GAATATGCGAACTTTGTCAAATCGGTGCTACAATTTGTTGGTTG 5308
ST-4/4          GGTTTATTCCCAACA-ATCCCAACTCTTATTGTTATGCTAAGACTCACACATATTTTTTA 5398
                ***************

X13777          ------------------------------------------------------------
S-TNT1          ------------------------------------------------------------
ST-3/3          CTGTCATGATTAAAAACATACCCAACAAGTACAGGTAATCATCCTAAATTCAATATTTTA
ST-3/4          CCAGTATCTCTTTTCTCTGCCAAGAATTTTTTATGTATATGTTTCTTTGAAATGCAGTGA
ST-1/4          TACAAACATTTGATCTTACTGAAGATTCC----------------------------------
ST-4/4          AA------------------------------------------------------------


X13777          ------------------------------------------------------------
S-TNT1          ------------------------------------------------------------
ST-3/3          TTTCTTTATTTCCTTATATCTATTCATATTCTAATTAATTTCCCCTTTTATGTTTAAATG
ST-3/4          TGGTTTGGTCAAGTACAATAATAGTAAAAGGATAATTTGTTGTCATGAGAAAGAATCTAA
ST-1/4          ------------------------------------------------------------
ST-4/4          ------------------------------------------------------------


X13777          ------------------------------------------------------------
S-TNT1          ------------------------------------------------------------
ST-3/3          CTATAGGAAGTCCCCTTTACGCGGTGAATAACTAATGAAGGTTATATGTTGATTAAATAA
ST-3/4          GTGTATGCACCATATTTAATAAGATCCTACAGAAGCTGTTATGTGATCCTTCATTGTTTG
ST-1/4          ------------------------------------------------------------
ST-4/4          ------------------------------------------------------------
```

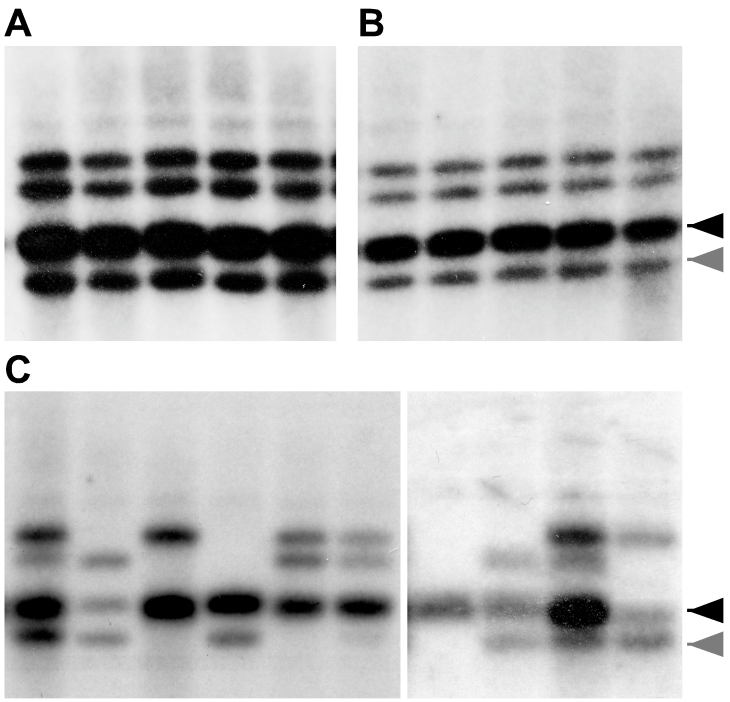


FIGURE 3-S2. S-TNT1 insertions are not activated or deleted during regeneration from (*A*) leaves, (*B*) roots or (*C*) selfed-seeds from the Ns-7834 plant. Black arrow shows the 2.4kb T-DNA fragment of S-TNT1, grey arrow shows the ST-1/4 hybrid retroelement fragments.

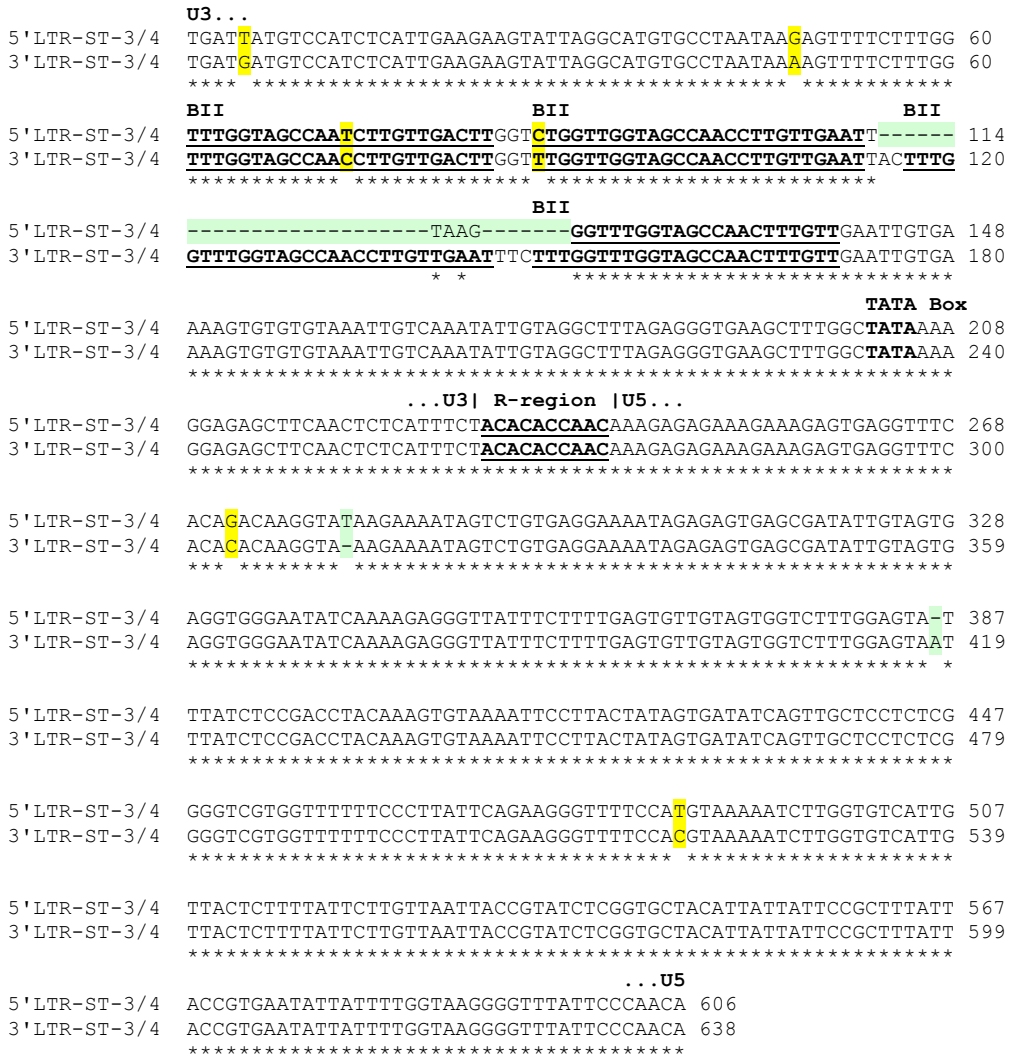


FIGURE 3-S3. Alignment of ST-3/4 long terminal repeats reveals mutation during retrotransposition. Substitutions are highlight in yellow and deletions in green. The U3, R, and U5 regions of the LTR are labeled. The BII repeats and TATA box are also labeled.

TABLE 3-S1. Transpositions and deletions absent in plants regenerated from Ns-7834 tissues on various media

Material	NaB (mM)	AzaC (mM)	Transpositions	Deletions	Plant Lines
Root	0	0	0	0	13
Leaf	50	0	0	0	8
Leaf	20	0	0	0	13
Leaf	10	100	0	0	9
Leaf	10	0	0	0	11
Leaf	0	100	0	0	7
Leaf	0	0	0	0	5
Seed	10	100	0	0	11
Seed	10	0	0	0	7
Seed	0	100	0	0	10
Seed	0	0	0	0	13
		TOTAL	**0**	**0**	**107**

TABLE 3-S2. Primers for Inverse PCR

Progeny	**Primer**	**Strand**	**Sequence**
ST-3/3 ST-1/4 ST-4/4	ip2R	R	GAACCAACCGCAGAACTGTAGGGC
	ip1R	R	AACAGCCTCCCAGTGCTCCTTACC
	ip1F	F	ATGTTTCGGCGGTAGCGACCCTAT
	ip2F	F	TCTCTGGAGGGGCAATCTCTTGGC
ST-3/4	799ipF1	F	GCGATTCCTTCAAGAGCTTGGATTGC
	799ipF2	F	ACCATGCAAGGACCAAACACATTGA
	799ipR2	R	TCCTGGAAGTCCTTTCCCTCACAA
	799ipR1	R	TTCAAGTGTCCGGCCAATGGGGTA

TABLE 3-S3. BLAST results for characterized S-TNT1 insertion sites

Progeny	**Best Similarity**	**Score, E value**	**Locus**
ST-1/4	CHO_OF3401xi23r1.ab1	1856, 5.2e-78	Methyl-filtered Assembly
ST-4/4	CHO_OF305xb23r1.ab1	931, 2.3e-34	Methyl-filtered Assembly
ST-3/3	CHO_OF3737xe19r1.ab1	3956, 7.3e-173	In terminal EAR1 homolog
ST-3/4	AGN_OF3457xm13r1.ab1	937, 2.2e-81	Methyl-filtered Assembly

TABLE 3-S4. Primers to amplify and sequence S-TNT1 insertions in transgenic plants and loci in wild type *N. sylvestris*

Position	Primer	Strand	Sequence
620	DraR0934	R	CGAGCAGAACCTGTGCTCTGATAC
640	7F740	F	AGCACAGGTTCTGCTCGTTCACTG
1990	7F2090	F	AATGGGTCACATGAGCGAGAAGGG
2280	7R2390	R	TGGAAAACTTGAAACACCTGATCTTTGGT
3530	7R3640	R	CCAGCTACTTCAAATCCTTCTGGTTGC
3810	3U3F1	F	GGCCCAGCTCAGCAGATATTGGGT
4310	5U3R1	R	GCCAAGAGATTGCCCCTCCAGAGA
4530	7F4640	F	ACCACTGGATCAGAGAGATGGTCG
4790	3U3F2	F	TTGGTTTGGTAGCCAACCTTGTTGA
5160	7F5260	F	TCAGTTGCTCCTCTCGGGGTCG
5170	3U3R1	R	CGACCCCGAGAGGAGCAACTGATA
5260	2o7F5360	F	TACCGTATCTCGGTGCTACA
5' ST-3/3	EarF0256	F	AGGCCTGGGAAAATCAAGAATGCC
3' ST-3/3	Ear-33R	R	TTCACCGCGTAAAGGGGACTTCCT
5' ST-3/4	3A119uF2	F	GACCCTCTGTACCCAGCAGCCTA
3' ST-3/4	3A119dR1	R	TGGCAGAGAAAAGAGATACTGGAGCC
5' ST-1/4	22F0235	F	CTCTCGTGATCTTCCGCCCACACT
3' ST-1/4	22-33R	R	GCAATAGAAGAGCTGAAAGTTCCTTCCT
5' ST-4/4	DraF0015	F	ATCCATTACACAATTTTACGATATTCGAGTT
3' ST-4/4	Dra38R	R	TCGTTACGCAGAAGCTCTGCCATT

REFERENCES FOR CHAPTER THREE

1. Orgel LE & Crick FH (1980) Selfish DNA: the ultimate parasite. Nature 284:604-607.
2. Kazazian HH, Jr. (2004) Mobile elements: drivers of genome evolution. Science 303:1626-1632.
3. Initiative TAG (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796-815.
4. Schnable PS, et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. Science 326:1112-1115.
5. Wang Q & Dooner HK (2006) Remarkable variation in maize genome structure inferred from haplotype diversity at the *bz* locus. Proc Natl Acad Sci U S A 103:17644-17649.
6. Kashkush K, Feldman M, & Levy AA (2002) Gene loss, silencing and activation in a newly synthesized wheat allotetraploid. Genetics 160:1651-1659.
7. Bruggmann R, Bharti AK, Gundlach H, Lai J, Young S, Pontaroli AC, Wei F, Haberer G, Fuks G, Du C, Raymond C, Estep MC, Liu R, Bennetzen JL, Chan AP, Rabinowicz PD, Quackenbush J, Barbazuk WB, Wing RA, Birren B, Nusbaum C, Rounsley S, Mayer KF, & Messing J (2006) Uneven chromosome contraction and expansion in the maize genome. Genome Res 16:1241-1251.
8. Grandbastien MA, et al. (1994) Functional analysis of the tobacco Tnt1 retrotransposon. Genetica 93:181-189.
9. Hirochika H, Sugimoto K, Otsuki Y, Tsugawa H, & Kanda M (1996) Retrotransposons of rice involved in mutations induced by tissue culture. Proc Natl Acad Sci U S A 93:7783-7788.
10. Tadege M, et al. (2008) Large-scale insertional mutagenesis using the Tnt1 retrotransposon in the model legume *Medicago truncatula*. Plant J 54:335-347.
11. Vitte C & Bennetzen JL (2006) Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution. Proc Natl Acad Sci U S A 103:17638-17643.
12. SanMiguel P, Gaut BS, Tikhonov A, Nakajima Y, & Bennetzen JL (1998) The paleontology of intergene retrotransposons of maize. Nat Genet 20:43-45.
13. Devos KM, Brown JK, & Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. Genome Res 12:1075-1079.
14. Sharma A, Schneider KL, & Presting GG (2008) Sustained retrotransposition is mediated by nucleotide deletions and interelement recombinations. Proc Natl Acad Sci U S A 105:15470-15474.
15. Du J, et al. (2010) Bifurcation and enhancement of autonomous-nonautonomous retrotransposon partnership through LTR swapping in soybean. Plant Cell 22:48-61.
16. Hou Y, Rajagopal J, Irwin PA, & Voytas DF (2010) Retrotransposon vectors for gene delivery in plants. Mob DNA 1:19.
17. Feng YX, Moore SP, Garfinkel DJ, & Rein A (2000) The genomic RNA in Ty1 virus-like particles is dimeric. J Virol 74:10819-10821.
18. Xu H & Boeke JD (1987) High-frequency deletion between homologous sequences during retrotransposition of Ty elements in *Saccharomyces cerevisiae*. Proc Natl

Acad Sci U S A 84:8553-8557.
19. Stuhlmann H & Berg P (1992) Homologous recombination of copackaged retrovirus RNAs during reverse transcription. J Virol 66:2378-2388.
20. Onafuwa-Nuga A & Telesnitsky A (2009) The remarkable frequency of human immunodeficiency virus type 1 genetic recombination. Microbiol Mol Biol Rev 73:451-480.
21. Grandbastien MA, Spielmann A, & Caboche M (1989) Tnt1, a mobile retroviral-like transposable element of tobacco isolated by plant cell genetics. Nature 337:376-380.
22. Vernhettes S, Grandbastien MA, & Casacuberta JM (1998) The evolutionary analysis of the Tnt1 retrotransposon in *Nicotiana* species reveals the high variability of its regulatory sequences. Mol Biol Evol 15:827-836.
23. Hajdukiewicz P, Svab Z, & Maliga P (1994) The small, versatile pPZP family of *Agrobacterium* binary vectors for plant transformation. Plant Mol Biol 25:989-994.
24. Jones PA & Taylor SM (1980) Cellular differentiation, cytidine analogs and DNA methylation. Cell 20:85-93.
25. Boffa LC, Vidali G, Mann RS, & Allfrey VG (1978) Suppression of histone deacetylation in vivo and in vitro by sodium butyrate. J Biol Chem 253:3364-3366.
26. Lutz KA, Svab Z, & Maliga P (2006) Construction of marker-free transplastomic tobacco using the Cre-loxP site-specific recombination system. Nat Protoc 1:900-910.
27. Silver J & Keerikatte V (1989) Novel use of polymerase chain reaction to amplify cellular DNA adjacent to an integrated provirus. J Virol 63:1924-1928.
28. Casacuberta JM, Vernhettes S, & Grandbastien MA (1995) Sequence variability within the tobacco retrotransposon Tnt1 population. EMBO J 14:2670-2678.
29. Onafuwa-Nuga AA, King SR, & Telesnitsky A (2005) Nonrandom packaging of host RNAs in moloney murine leukemia virus. J Virol 79:13528-13537.
30. Lucas H, Feuerbach F, Kunert K, Grandbastien MA, & Caboche M (1995) RNA-mediated transposition of the tobacco retrotransposon Tnt1 in *Arabidopsis thaliana*. EMBO J 14:2364-2373.
31. Feuerbach F, Drouaud J, & Lucas H (1997) Retrovirus-like end processing of the tobacco Tnt1 retrotransposon linear intermediates of replication. J Virol 71:4005-4015.
32. Tamura K, et al. (MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731-2739.
33. Ma J & Bennetzen JL (2004) Rapid recent growth and divergence of rice nuclear genomes. Proc Natl Acad Sci U S A 101:12404-12410.
34. Swain A & Coffin JM (1992) Mechanism of transduction by retroviruses. Science 255:841-845.
35. Hernandez-Pinzon I, de Jesus E, Santiago N, & Casacuberta JM (2009) The frequent transcriptional readthrough of the tobacco Tnt1 retrotransposon and its possible implications for the control of resistance genes. J Mol Evol 68:269-278.

36. Svab Z, Hajdukiewicz P, & Maliga P (1995) Generation of Transgenic Tobacco Plants by Colcultivation of Leaf Disks with *Agrobacterium* pPZP Binary Vectors. Methods in plant molecular biology: a laboratory course manual, ed Maliga P (Cold Spring Harbor Laboratory Press), pp 55-77.
37. Murray MG & Thompson WF (1980) Rapid isolation of high molecular weight plant DNA. Nucleic Acids Res 8:4321-4325.

CURRICULUM VITAE for Gregory THYSSEN

Education

Rutgers, the State University of New Jersey, New Brunswick, NJ
2007-2012 Ph.D. Plant Biology
Dissertation: *Nicotiana sylvestris*, a Model Plant for Cell Biology: Organelle Movement and Retrotransposon Mutagenesis

Yale University, New Haven CT, USA
1998-2002 B.S. Molecular Biophysics & Biochemistry

Professional Experience

2006-2007 Chemistry Teacher**,** Jim Hill High, Jackson Public Schools, Jackson, MS
2003-2006 Life Science Research Assistant, Stanford University, Stanford, CA

Publications

Thyssen, Gregory, Zora Svab, & Pal Maliga, Cell-to-cell movement of plastids in plants. Proceedings of the National Academy of Sciences 2012 Feb, 14(109):2439-2443.

Thyssen, Gregory, Tzu-Huey Li, Lynn Lehmann, Ming Zhuo, Manju Sharma, Zijie Sun, LZTS2 Is a Novel b-catenin Interacting Protein and Regulates the Nuclear Export of b-catenin, Molecular and Cellular Biology 2006 Dec, 26(23):8857-8867.

Li, Xiomeng, Gregory Thyssen, Jason Beliakoff, Zijie Sun, The Novel PIAS-like protein hZimp10 Enhances Smad Transcriptional Activity, Journal of Biological Chemistry 2006 Aug, 18(33):23748-56.