Development of a Transposon Based Activation Tagged Mutant Population in Tomato for Functional Genomic Analysis

Jared Daniel Carter

Thesis submitted to the Faculty of the
Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science
in
Horticulture

Richard E. Veilleux
Andy Pereira
Bingyu Zhao

April 17, 2012
Blacksburg, VA

Keywords: Solanaceae, *AcDs*, *Solanum lycopersicum*, transformation, TAIL-PCR

Development of a Transposon Based Activation Tagged Mutant Population in Tomato for Functional Genomic Analysis

Jared Daniel Carter

Abstract

Tomato serves as an important model organism for Solanaceae in both molecular and agronomic research. With whole genome sequencing in progress, there is a need to study functional genetics through mutant lines that exceed the practical limitations imposed by the popular research cultivar, Micro-Tom. This project utilized *Agrobacterium* transformation and the transposon tagging construct, *Ac-Ds*ATag-Bar_gosGFP, to produce activation tagged and knockout mutants in the processing tomato variety, M82. The construct contained hygromycin resistance (*hyg*), green fluorescent protein (*GFP*), and maize transposase (*TPase*) on the stable *Ac* element, along with a 35*S* enhancer tetramer and glufosinate herbicide resistance (*BAR*) on the mobile *Ds* element. An *in vitro* propagation strategy was used to produce a population of 25 T_0 plants from a single transformed plant regenerated in tissue culture. A T_1 population of 10,568 selfed and M82 backcross progeny was produced from the functional T_0 line. This population was screened by spraying with 0.05% Liberty® herbicide, followed by a 100 mg/L hygromycin leaf painting procedure to select for *Ds*-only (herbicide tolerant and hygromycin sensitive) individuals. The T-DNA genotype of *Ds*-only plants was confirmed through multiplex PCR and the location of insertions within the genome was determined through TAIL-PCR. Resulting product sequences were blasted against the pre-publication tomato genome browser to determine insertion sites. A population of 309 independent transposants dispersed to all twelve chromosomes from the original insertion site on chromosome five has been developed. The transposon tagged lines are currently being immortalized in seed stocks.

Contents

List of Tables iv
List of Figures v
Chapter 1: Literature Review and Background 1
 Introduction 1
 Functional Genomics as Applied to Plant Breeding 2
 The Use of Mutant Lines in Functional Genomics 4
 Ac/Ds Transposon Tagging 7
 Objectives 9
Chapter 2: Generation of a Mutant Population and Progeny Screening 10
 Abstract 10
 Introduction 11
Chapter 3: Materials and Methods 13
 Plant Material Selection and Transformation 13
 T_1 Population Generation 14
 T_1 Screening 14
 DNA Extraction and PCR Screening 14
 Identification of *Ds* Insertion Flanking Regions 15
Chapter 4: Results 17
 T_0 Line Generation and Analysis 17
 Analysis T_1 Families 18
 T_1 Progeny Screening 19
 Characteristics of Transposition 20
 Phenotypic Mutants 21
Chapter 5: Discussion 22
 Supplemental Materials and Methods 27
  CTAB DNA Extraction Protocol 27
  Extraction of PCR Bands from Agarose Gel 28
References 29
Appendix A. Tissue Culture Media Formulations 40
Appendix B. Primer Sequences 41

List of Tables

Table 1. Evaluation of finalized T_1 progeny screening protocol applied to progeny derived by self-pollination of eight T_0 families and cross-pollination to WT M82. All T_0 plants contained a non-mobile *Ac* element and a mobile *Ds* element. The progeny were screened for resistance to Liberty® herbicide to identify those with the *Ds* element, followed by hygromycin leaf painting of herbicide resistant seedlings to identify those with the *Ac* element. Screening efficiency measured the percentage of seedlings identified as putative transposants (herbicide resistant, hygromycin sensitive) that were verified as *Ds*-only by multiplex PCR. Totals do not represent all T_1 progeny screened during project. 32
Table 2. Distribution of transposed *Ds* elements to tomato pseudochromosomes (Tomato WGS Chromosomes SL2.40), sorted by T_0 family. Includes T_0 self and WT × T_0 progeny. Underlined numbers indicate the presence of an identified common (somatic) transposition event on the chromosome in the respective cutting family. Common insertions are counted as one independent transposition event.. 33
Table 3. Insertion site classification of 309 identified *Ds* transposition events based on Tomato WGS Chromosomes SL2.40. Insertion site preference displayed as a percentage of the 309 total insertions. 34
Table 4. Tissue culture media formulations for Agrobacterium mediated transformation of M82 tomato primary leaf explants. Modified from Khoudi et al., (2009). All media formulations were adjusted to pH 5.7-5.8. 40
Table 5. Primer sequences.. 41

List of Figures

Figure 1: The *Ac-Ds*ATag-Bar_gosGFP construct. Elements are as follows; LB = Left Border, Hyg = hygromycin resistance, 35SPr = Cauliflower Mosaic Virus 35*S* promoter, GFP = Green Fluorescent Protein, GosPr = *Zea mays* promoter, AcTPase = *Zea mays* Transposase, AcPr = *Zea mays* transposase native promoter, IR = Inverted Repeat, 35S 4 Enh = Four tandem copies of Cauliflower Mosaic Virus 35*S* enhancer, BAR = glufosinate resistance, UbiPr = Ubiquitin promoter, IR = Inverted Repeat, RB = Right Border. 35

Figure 2. Phenotypic Screening of T_1 Progeny. (A) Herbicide resistant seedling (left) next to sensitive seeding (right) 7 days after spraying. (B) Close up of herbicide resistant leaf (left) next to sensitive leaf (right) removed from previous seedlings. (C) Adaxial side of four leaves detached from different seedlings sensitive to hygromycin 6 days after painting. Reaction to hygromycin localized to exact region of leaf that was painted. (D) Abaxial side of a leaf sensitive to hygromycin displaying the diagnostic necrotic spots that distinguish hygromycin sensitivity from common leaf damage due to pathogens and mechanical injury. 36

Figure 3. Analysis of T_0 transgenic line 7a by rtPCR, screening of T_1 progeny through a combined PCR multiplex and zygosity test, and TAIL-PCR analysis of *Ds*-only T_1 progeny. (A) Expression of transposase in whole inflorescence RNA. PCR product spans transposase intron 2 (DNA product = 476 bp, cDNA product = 405 bp). Positive control DNA and cDNA extracted from transgenic potato inflorescence. (B) Visualization of primary TAIL-PCR products. (C) Screening of T_1 progeny by multiplex PCR and zygosity testing. SQE = squalene epoxidase (distinguishes wild types from failed reactions), CI#1 = common insertion one from a somatic transposition event early in development (defines T_0 group 2), BAR = glufosinate herbicide resistance gene (*Ds* element), HPT = hygromycin resistance gene (*Ac* element), CI#2 = common insertion two from somatic transposition event unique to T_0 7m. 37

Figure 4. Graphic distribution of 309 transposed *Ds* elements to tomato pseudochromosomes (Tomato WGS Chromosomes SL2.40). Includes self and WT × T_0 progeny from every T_0 family. Both common and unique insertions are represented by a single band. Original T-DNA insertion site during transformation (*Ac* element) denoted by ←**Ac** on Chromosome 5............................ 38

Figure 5. Phenotypes for seven T_1 activation tagged mutants. (A) 7c-79, rolled leaves, no lateral branching, weak stem. (B) 7p-99, dwarf, warped leaves, variegation. (C) 7g-300 crumpled leaves. (D) 7i-271, twisted leaves, extreme dwarfism. (E) 7g-25, tightly rolled, leathery leaves, dwarf, aborted flowers. (F) WT7i-48, extensive yellow variegation, aborted flowers. (G) 7p-3 speckled variegation.. 39

Chapter 1: Literature Review and Background

Introduction

The selective breeding of tomato (*Solanum lycopersicum* L.) has transformed the species from a weedy perennial in the Peruvian highlands to a globally cultivated, horticultural food crop. Tomato fruit currently stands as the world's eleventh largest agricultural product (FAOSTAT, 2011) and has been incorporated into the cuisine of many cultures. Characteristics such as the rich antioxidant content of the fruit hold promise that tomato will remain an important crop in the foreseeable future.

Tomato breeding has focused heavily on fruit characteristics, giving the modern world an extensive list of cultivars that offer varying shape, water content, and flavor profiles. Much attention has also been given to the plant's morphology as part of efforts to develop cultivars for growing conditions ranging from mechanically harvested field rows to hydroponic operations. Disease and pest resistance has become a common concern in tomato breeding, as crops are challenged by a large array of pathogens, some of which rapidly evolve means of circumventing plant and chemical defenses.

More recent advances in tomato breeding have focused on identifying and making use of disease resistance genes found in wild tomatoes and closely related *Solanum* accessions. Research into recombinant DNA technology has provided both an additional means to introduce this genomic diversity into cultivated tomato and access to genes far beyond the reach of conventional breeding. A greater understanding of genetics has also enabled tomato to serve as a model organism for other crops in the Solanaceae family, including the staple food crop, potato (*Solanum tuberosum* L.).

Current advances in genomics are providing new possibilities for tomato as a crop. A complete sequence of the cultivated tomato genome is in progress (Bombarely et al., 2011;

Mueller et al., 2009) and will provide the basis for advances in molecular research. Translating this data into agronomic progress by tomato breeders requires an understanding of how the components of the genome function in both the cell and the overall plant. The development of mutant populations has proven to be a versatile tool for understanding gene function. An activation tagged mutant population in a commercial tomato cutivar, M82, will provide a useful resource for the study of functional genomics in Solanaceae.

Functional Genomics as Applied to Plant Breeding

Functional genomics can be described as an attempt to bring together bodies of data relating to gene structure and function into an understanding of complex interactions in molecular biology. Sources of data can include whole genome sequences, transcriptome profiling, and protein characterization. As applied to plant breeding, functional genomics reveals the genes and interactions thereof that underlie important agronomic traits, as well as mechanisms by which these traits can be manipulated in the field (Barone et al., 2009).

Publication of a first whole genome sequence of a species opens up many avenues of research and possibilities for functional genomics projects. The relatively small genome (125 Mb) of the research model species, *Arabidopsis thaliana*, was the first plant to be completed (Arabidopsis, 2000). Next generation sequencing technology has allowed the sequencing of larger genomes, such as *Solanum tuberosum* (700 Mb) (Xu et al., 2011), while the genomes of major cereal crops are in progress (Feuillet, 2012). The first whole genome sequence of tomato (950 Mb) is currently in progress (Bombarely et al., 2011; Mueller et al., 2009) and a draft sequence is already available for utilization by researchers.

With the publication of a first whole genome sequence for a model plant of a particular species, it is then possible to quickly and economically sequence the entire genome of other

individuals, using the first genome as an alignment template (Huang et al., 2009). The power of this technique has been demonstrated in Arabidopsis (Turner et al., 2010), in which many genes underlying the complex trait, such as acid soil tolerance, were identified through population resequencing. Application of this strategy to tomato could result in a better understanding of how genes relate to agronomic traits and reveal putative interactions for further study. A whole genome sequence also allows for the rapid positioning of T-DNA insertions and sequenced genes onto mapped chromosomes using search engines, as opposed to laborious mapping techniques. While these data are useful, they only provide a basis for studying functional genomes.

Moving forward from the genome, studying the transcriptome of a species is an essential step in understanding gene function. A transcriptome consists of mRNA sequences and variations thereof, complete with expression levels of mRNA species and their specific expression profiles in various tissues within the plant under different environmental conditions. Sequencing a transcriptome has been made possible by RNA-Seq (Wang et al., 2009), a recent technique gaining wide application. This technology also enables the study of gene expression levels in specific tissue types, though such applications may be limited by the difficulty in isolating samples at precise developmental stages. Mutant lines play an important role in researching the transcriptome, for they allow the study of individual genes with altered expression profiles.

The Use of Mutant Lines in Functional Genomics

The term "reverse genetics" is often used to describe the role mutant lines play in functional genomics. Reverse genetics focuses on researching phenotypes that arise when genes or their expression characteristics are altered by some means, with the intent to gain a better understanding of the role a gene plays in the overall development and phenotype of an organism. This strategy is of particular interest in crop plants, as it may reveal methods by which a gene can be manipulated for better agronomic production. Many techniques have been developed to alter genes for reverse genetic screening and advances have already been made through such studies.

A broad category of mutants used in reverse genetics is knockout, or "loss-of-function" mutants, which can be generated through a variety of techniques. Knockout mutants are those in which a gene's function has been completely disabled or reduced to below a detectable threshold. This disabling of individual genes can be accomplished through mutations in the sequence that alter final protein structure, insertion of large pieces of nucleotide sequence into the coding regions of genes, or through shutdown of gene translation by RNA interference. Knockouts via sequence mutations are identified through screening populations for variants that emerge from natural mutation, or generated through the introduction of mutagenic agents to model plant lines. There is currently a population of 13,000 tomato knockout mutants generated through ethyl methanesulfonate (EMS) exposure (Menda et al., 2004). Knockout lines generated through sequence insertions are also available (Meissner et al., 2000). An individual from the latter population was used to further the understanding of how plants acquire inorganic phosphate, through a comparative study between wild type tomato and a knockout line for a known phosphate transporter (Xu et al., 2007). While this work is impressive, it is often eclipsed by the shortcomings of knockout mutants and the strategies by which they are generated

(Emmanuel and Levy, 2002). Knockout mutants suffer from several inherent limitations due to the nature of plant genomes. There are often multiple copies of a given gene, especially in larger genomes, which reduce or eliminate the phenotype induced by a knockout mutation. This same reduction in phenotype can also be observed due to redundancy in biochemical pathways. Finally, a knockout mutation within a gene essential to plant survival may never be studied due to lethality of the mutation early in development (Hsing et al., 2007). Natural and EMS mutants also carry a serious disadvantage in that it is difficult to identify the location of point mutations without extensive resequencing of candidate genes or the entire genome. Due to these many limitations, attention has shifted towards another broad class of mutants.

Mutant phenotypes may also be induced by altering the expression pattern and expression level of individual genes. These "gain-of-function" mutants overcome the difficulties of genetic redundancy and lethal phenotypes, though they may be more challenging to produce in large numbers. Great alterations in gene expression are accomplished through inserting a powerful transcriptional promoter or activator element into the native regulatory region of a gene. The activation tag is a common and widely used element in mutant line generation. This segment of T-DNA consists of four tandem copies of the transcriptional enhancer sequence from the P35S promoter found in Cauliflower Mosaic Virus (CaMV) (Walden et al., 1994). The activation tag was used to generate mutant collections in Arabidopsis, in which the tag was capable of both inducing constitutive expression and/or increasing expression levels in patterns dictated by native promoters (Weigel et al., 2000). More recent applications have constructed activation tagged mutant populations of crop plants, notably rice *(*Oryza sativa) (Hsing et al., 2007).

Among the many publications resulting from analysis of these mutants, an example in tomato by Mathews et al. (2003) demonstrates the potential for translation of molecular research

into applied plant breeding. An activation tagged mutant in the tomato cv. Micro-Tom was observed to have purple coloration in most floral organs due the production and accumulation of far more anthocyanin than was found in wild type plants. The activation tag was located adjacent to a MYB transcription factor, which ultimately regulated the production of anthocyanin. This relationship of an upregulated gene and novel trait also highlighted the ability of an activation tag to reveal the function of regulatory elements such as transcription factors. There has been great interest in plant anthocyanins as antioxidants in the human diet (Castañeda-Ovando et al., 2009), making purple plants more desirable to consumers. Though conflicting evidence has been found as to the benefits of anthocyanin consumption, continued interest creates an incentive for breeding purple food crops. Discovery of this regulatory pathway provided a means by which to manipulate anthocyanin levels if the goal should arise in tomato breeding.

Though activation tagging has many advantages, the difficulty of inserting an enhancer element into the genome presents a challenge. A proven strategy to insert T-DNA into a plant has been transformation by *Agrobacterium tumefactions*, which was also the method of choice for early projects in Arabidopsis (Weigel et al., 2000). Large scale mutant population generation via *in planta* transformation may be an efficient route in Arabidopsis, but such an approach is not applicable to most crop plants. Hence, a protocol that can deliver multiple activation tagged mutants from each independently transformed line regenerated in tissue culture is required. Activation tagging by means of the *Ac/Ds* transposable element has proven to be a system by which to generate large mutant populations from a small number of transformed lines (Kuromori et al., 2004).

***Ac/Ds* Transposon Tagging**

Transposons are mobile and potentially self-replicating sequences within the genome that have important implications in expanding genome size, mutation, and evolution. They can mobilize and replicate through mechanisms involving transcription to RNA, or simply move about as segments of DNA. Transposons are partially defined as autonomous or non-autonomous, indicating whether or not they contain a functional copy of an enzyme responsible for transposition. The past decades have brought extensive research into transposons, ultimately leading to the elements being modified into T-DNA constructs for scientific use. Their valuable attributes make transposons an indispensable tool in the development of mutant populations.

In the paradigm of plant biology, transposon systems are based on the transposase enzyme, Tn5, and its corresponding inverted repeat sites native to *Zea mays* (Becker and Kunze, 1997). This mobile element was termed the *Activator* (*Ac*). In addition to *Ac*, other mobile elements were found in the maize genome. These elements, individually termed *Dissociator* (*Ds*) elements, were mobile due to the activity of *Ac*, but lacked a functional copy of the transposase gene. Further analysis revealed these elements were degraded copies of *Ac* that retained their terminal recognition sites (Rubin and Levy, 1997). The numerous potential applications of these elements to research were soon realized and the system was found to function in other species (Baker et al., 1986; Izawa et al., 1991). For research purposes, the inverted repeats surrounding *Ac* are removed to stabilize the element, while sequences within the *Ds* element are tailored to specific applications. Segments of T-DNA ranging from activation tags to entire native plant genes may be inserted into this *Ds* element, which transposase may then freely move throughout the genome. Selectable marker genes are commonly included to assist in screening for transposon tagged lines. The behavior of transposase is highly dependent on the promoter chosen to drive its expression in transformed cells. Promoter elements may be either constitutive or

tissue specific, directly governing the point in development in which transposase is expressed and indirectly affecting the number of times the *Ds* element can move during a plant's life cycle. T-DNA constructs may also be designed to inhibit multiple movements of the *Ds* element by removing promoters driving transposase upon excision (Suzuki et al., 2001). Other research efforts have focused on driving transposase expression with an inducible promoter, such as a one specific to heat response (Fladung and Polak, 2011). Analysis of large transposon tagged populations has shown global transposition to all chromosomes within a genome, as well as preference for transposition to distal chromosome ends or clustering of insertions into hot and cold regions along chromosome arms (Alonso et al., 2003; Jeong et al., 2006).

The *Ac/Ds* system has been utilized in many transposon tagging projects, including those in Arabidopsis (Kuromori et al., 2004), tomato (Meissner et al., 2000), and rice (Kolesnik et al., 2004). In addition to the system providing multiple tagged lines from the progeny of a single transformed plant, *Ds* tagged mutants carry other advantages. *Ds* elements may be designed to include herbicide or antibiotic resistance genes to enable efficient screening for tagged lines based on phenotype. Nested primer sites, restriction sites, and bacterial origins of replication (Suzuki et al., 2001) may also be included to simplify isolating flanking regions surrounding transposed *Ds* elements. Given these attributes, an *Ac/Ds* transposon based system would be an effective method to develop a population of activation tagged mutant lines.

Objectives

Activation tagged mutants produced during the project will be identified through progeny screening and characterized for the location and behavior of T-DNA insertions. Specific objectives include:

1. Construct a T_0 population of M82 tomato plants transformed with an available *Ac/Ds* activation tagging system.

2. Develop an efficient and economical protocol for screening segregating progeny of transformed plants, using green fluorescent protein (*GFP*), glufosinate herbicide (Liberty®), hygromycin resistance (*hpt*), and multiplex PCR.

3. Identify the insertion site of activation tags within mutant tomato lines by sequencing regions flanking transposed *Ds* elements using TAIL-PCR (Tsugeki et al., 1996) with hiTAIL-PCR primers (Liu et al., 1995).

4. Utilize nucleotide sequences of flanking regions to identify genes within range of the activation tags by BLAST searching against the pre-publication Tomato Genome Browser (Bombarely et al., 2011).

5. Immortalize activation tagged lines through seed stocks.

Chapter 2: Generation of a Mutant Population and Progeny Screening

Abstract

Primary leaf explants of M82 tomato were transformed with the *Ac/Ds* transposon based, activation tagging construct *Ac-Ds*ATag-Bar_gosGFP. A single transgenic shoot was regenerated in tissue culture, in which the entire T-DNA construct inserted into the intron of 60S ribosomal protein L29 (Solyc05g053440.2.1). *In vitro* propagation was used to develop a T_0 population of 25 clonal plants, designated as families, from the transformed shoot. A T_1 population was generated from these families by self-pollination and outcrosses to non–transgenic M82. A screening protocol was developed for use on over 10,000 T_1 seedlings, using glufosinate herbicide as a marker for the *Ds* element and hygromycin leaf painting as a marker for the *Ac* element. A finalized version of protocol was applied to 3,930 seedlings, demonstrating the technique to be 83% effective in eliminating non-target progeny prior to PCR analysis. A total of 309 independent insertions of the *Ds* element were identified through TAIL-PCR, dispersed to every chromosome in the genome. Insertion site analysis showed that 73% of *Ds* elements transposed to genomic regions where the T-DNA could induce a gain or loss of function phenotype (coding sequences, introns, promoters, 3' UTRs, < 2 kb from described genes). Fifteen T_1 progeny were observed to have obvious mutant phenotypes during initial screening, several of which included widespread morphological changes. Screening of the T_1 population revealed that the T_0 families divided into four distinct groups based on the presence of common insertions and progeny segregation data. Efforts to preserve unique *Ds*-only lines through seed stocks and *in virto* propagation of sterile mutants are currently underway.

Introduction

Activation tagged mutant populations have proven to be a useful resource in functional genomics, with topics of research including auxin biosynthesis (Kondou et al., 2010), iron accumulation (Lee et al., 2009), and fruit ripening (Giménez et al., 2010). The development of many populations has involved a T-DNA tag containing four tandem copies of the enhancer region taken from the promoter within *Cauliflower mosaic virus* (Walden et al., 1994). Insertion of this element into gene flanking regions offers both the possibility of constitutive increases in transcription and amplification of expression profiles dictated by native promoters (Weigel et al., 2000). Though mutants can be generated through plant transformation, the use of the *Activator (Ac)/Dissociator (Ds)* transposon tagging system has proven to be an efficient method for mutant population construction. Mutant populations numbering in the thousands have been produced using the *AcDs* system in Arabidopsis (Kuromori et al., 2004) and rice (Jeong et al., 2002; 2006). Use of the *AcDs* system enables the recovery of many independent mutants from a single transformed line regenerated in tissue culture and facilitates rapid identification of insertion flanking regions (Kuromori et al., 2004).

In addition to being an important food crop unto itself, tomato serves as a functional model organism for other crops in Solanaceae and plant molecular biology. With efforts underway to publish the first whole genome sequence of tomato in the cultivar, Heinz 1706-BG (Bombarely et al., 2011), the role of tomato as a research organism will continue to grow. Transposon tagging has already proven useful in tomato (Meissner et al., 2000) and research on activation tagged mutant lines has been published (Giménez et al., 2010; Mathews et al., 2003). However, tomato still lacks the large, readily available populations of T-DNA tagged mutants that have accelerated research in other model crops. Many previous efforts have focused on the

popular research cultivar, Micro-Tom. While economical to maintain, the diminutive size and yield of Micro-Tom make it less useful in studying traits that directly relate to agronomic performance under field conditions. A proven alternative to Micro-Tom is the determinate processing cultivar, M82, a selection of UC-82 (Stevens et al., 1982). Advantages of using this cultivar include prolific seed production, a mature size conducive to greenhouse cultivation, and the ability to correlate research findings with mechanized tomato production in commercial settings. The cultivar has also seen wide use in research, including collections of EMS mutants (Menda et al., 2004) and introgression lines (Eshed and Zamir, 1995).

Chapter 3: Materials and Methods

Plant Material Selection and Transformation

The activation tagging construct *Ac-Ds*ATag-Bar_gosGFP (Figure 1) in the vector plasmid PMOG22 was selected for use in tomato transformation based on its active transposition in rice (Trijatmiko, 2005). Primary leaf explants taken from seedlings grown for 21 days in vitro in Magenta boxes (Magenta Plastics, Chicago, IL) (4-5 seedlings per box) in a growth chamber were selected and transformed according to protocols published by Khoudi et al. (2009), with several modifications made during transformation and regeneration. Modified media formulations are listed in Table 4. Selection by hygromycin was not initiated for the first 3 weeks following transformation. The concentration of zeatin in shoot regeneration and shoot elongation media was doubled in all cases. The concentration of cefotaxime in regeneration and elongation media was increased to 500 mg/L, while carbenicillin was excluded. As a further modification, regenerating shoot masses were kept on shoot regeneration media until a readily identifiable apical meristem was formed. Only masses with identifiable meristems were transferred to elongation media. Once a shoot had regenerated from a putatively transformed callus and elongated to have multiple nodes, the apical portion was removed to rooting medium and the base with a single node was retained until an axillary branch had grown sufficiently to repeat the process. By this method of vegetative propagation in tissue culture, each transgenic plant was represented by a population of consecutively lettered shoots. Plantlets were transferred to rooting media (MS + vitamins 4.43 g/L, sucrose 30 g/L, IAA 2 mg/L, cefotaxime 250 mg/L, Phytagel 2 g/L) for 7 days or until roots emerged. Rooted cuttings were transferred to hydrated peat plugs in Magenta boxes for 7 days to promote acclimation to potting media before transfer to a greenhouse facility. All tissue cultures were incubated in a growth chamber under 16 h days at 25°C under 110 μE *m*-2 s-1 fluorescent light and 8 h nights at 20°C.

T_1 Population Generation

T_0 plants were grown to maturity in soilless mix (equal volumes peat moss, course vermiculite, and perlite + 30 kg/m^3 gypsum) and allowed to self-pollinate with the assistance of daily flower vibration by a pollinator wand. Crosses were made between non-transformed M82 plants and transformed T_0 lines, with the transgenic acting as the pollinator in all cases. Seeds and their accompanying ovary flesh were cut from whole ripened fruit and further cleaned by soaking in approximately 100 ml 1M HCl in a 125 ml flask, under 20 min of constant blending by a stir bar. Seeds were then dried for 24 h at 37°C prior to storing in 5 ml capsules with silica bead desiccant.

T_1 Screening

All seedlings were grown in greenhouses at 25°C under 16 h photoperiod using natural and fluorescent light. T_1 seeds obtained from T_0 selfs or WT × T_0 crosses were soaked in tap water for 12 h and sown in soilless mix into flats of 48 or 96 cell packs. Emerging seedlings were screened with glufosinate herbicide after 3-4 weeks by spraying with an aqueous 0.05% v/v Liberty® solution. Reactions to the herbicide were evaluated after 5-7 days and sensitive seedlings were discarded. Glufosinate resistant seedlings were screened for hygromycin resistance by painting a primary or secondary leaf on each plant with a solution of 100 mg/L hygromycin + two drops polysorbate 20. Reactions to the antibiotic were evaluated after 5-7 days based on the presence of characteristic necrotic regions indicative of sensitivity to hygromycin. Hygromycin sensitive seedlings were identified as putative transposants.

DNA Extraction and PCR Screening

A single expanded leaf was collected from each glufosinate resistant, hygromycin sensitive seedling and frozen with liquid nitrogen in a 2 ml microcentrifuge tube containing a

single #2 steel shot pellet (Ballistic Products Inc., Corcoran, MN). The tissue was ground using a Geno/Grinder 2000 at cryogenic temperature. The DNA was isolated from ground tissue using a modified CTAB protocol (Doyle and Doyle, 1987) (Supplemental Materials). Isolated DNA was diluted tenfold for PCR analysis. The genotype of putative transposants was confirmed in a multiplex PCR reaction (Figure 3), which amplified the hygromycin resistance gene from the *Ac* element (HPT5F, HPT5R primers, Table 5), the glufosinate resistance gene from the *Ds* element (G38BARF, G38BARR primers, Table 5), and a sequence from squalene epoxidase within the tomato genome (Sl_SQEF, Sl_SQER primers, Table 5). For families containing *Ds* insertions common to many T_1 seedlings, a single zygosity primer and the sequencing primer, *Ds*5'-3 (Table 5), from within the *Ds* element were included to identify transposants harboring the common insertion (Figure 3C). Seedlings for which only the BAR gene and squalene epoxidase gene amplified were identified as putative unique transposants.

Identification of *Ds* Insertion Flanking Regions

Sequences flanking *Ds* insertions in unique transposants were amplified using primers and thermocycler protocols according to Tsugeki et al. (1996). The protocol was modified to include improvements according to Liu and Chen (2007), namely the use of long arbitrary degenerate primers and the addition of a 21-nucleotide sequence corresponding to the AC1 primer used in the primary and secondary TAIL-PCR reactions. Only the pre-amplification and primary TAIL-PCR reactions were carried out. Primary products were run on 1% agarose/1× TAE gel with 0.1% EtBr (Figure 3**B**). The brightest band(s) in each lane were excised and isolated (Supplemental Materials). Extracted bands were sequenced using the respective TAIL primer originally intended for use in secondary TAIL-PCR reactions (*Ds*5'-3, *D*s3'-3, Table 5). Resulting sequences were trimmed of T-DNA and aligned to the pre-publication tomato genome

(Bombarely et al., 2011) using a BLAST engine. Insertions sites were assumed to correspond to single alignment hits with a score of less than e^{-30}, or to the location with the lowest e-value in cases of multiple alignments. The identity and orientation of gene models up to 10 kb upstream and downstream from *Ds* insertion sites were recorded. Insertions for which sequencing produced multiple, near identical hits and those that did not align to the current genome (Tomato WGS Chromosomes SL2.40) were not counted as unique mutant lines during further analysis.

Chapter 4: Results

T_0 Line Generation and Analysis

A laborious transformation protocol involving hundreds of primary leaf explants regenerated three M82 tomato lines harboring the *Ac*-*Ds*ATag-Bar_gosGFP construct. Of these, one line did not set fruit after transfer to the greenhouse and another harbored only the *Ac* portion of the construct, as determined by multiplex PCR (data not shown). The remaining line was found to contain all the elements of the T-DNA construct. TAIL-PCR was performed on this intact line using primers situated at the right border of the *Ds* element and the resulting product was sequenced. BLAST searching this sequence against the draft tomato genome (Nov, 2011) indicated that the T-DNA had inserted into the single intron of 60S ribosomal protein L29 (Solyc05g053440.2.1) on chromosome 5 (Bombarely et al., 2011). The presence of the T-DNA insertion in this gene had no obvious effect on the overall phenotype compared to wild type M82 plants. Later analysis demonstrated that T_1 progeny homozygous for the insertion did not have a lethal or deleterious phenotype. Further investigation was made into the expression of 60S ribosomal protein L29 and the nearest gene on either strand, DNA-directed RNA polymerase subunit (Solyc05g053430.2.1) and late embryogenesis abundant protein 1 (Solyc05g053450.2.1), respectively. Analysis of 60S ribosomal protein L29 expression was predicted by comparison to orthologs in *Solanum phureja* (PGSC0003DMG400027161) and Arabidopsis (At3g06700). RNA-seq data from *Solanum* indicated widespread transcription of the native ortholog in every tissue type and treatment measured (Massa et al., 2011). Graphic expression of microarray data for At3g06700 revealed high levels of expression in all shoot apex tissue, in comparison to other tissue types (Winter et al., 2007). Cis-regulatory elements corresponding to pollen specific expression were identified (Higo et al., 1999) within the promoter regions of both neighboring genes, each of which closely flank Solyc05g053440.2.1. In addition, non-quantitative reverse

transcriptase PCR demonstrated readily detectable expression of the transposase gene (Figure 3, AcTPase) in whole inflorescence tissue.

Analysis T_1 Families

Using axillary shoots as a means of propagation, the original, single transformant with a complete launch pad was micropropagated into a population of 25 clonal T_0 plants. Progeny of these plants were segregated into families based on T_0 parent. Analysis of these individuals and their progeny revealed four distinct groups of plants based on the frequency of transposition. The first group (7d, e, h, k, n, o, q, v, w) demonstrated no incidences of transposition from the original insertion site. This was confirmed by both a lack of segregation between the *Ac* and *Ds* elements in 1,363 T_1 progeny analyzed, and sequencing of the non-transposed *Ds* element through TAIL-PCR of the same progeny. The presence of the non-transposed *Ds* element was not identified within any progeny from families belonging to other groups. The second group (7a, b, f, i, p, r, s, t, u, z) was characterized by the recovery of many independent transpositions of *Ds* among the T_1 progeny, with some common insertions, presumed to derive from somatic transposition events prior to meiosis, unique among the progeny of a single family within the group. Multiple phenotypic mutants were identified among the T_1 progeny of this group. The third group (7c, g, m, x, y) was defined by the presence of a single common insertion that was prevalent in the progeny of all the families within the group. Additional somatic and many germinal transposition events were also identified in the progeny of this group, included several phenotypic mutants. In addition, a fourth trend was observed in a single line (7l), which produced no self-progeny with a transposed or non-transposed *Ds* element. This apparent loss of the *Ds* element during excision was not unique to 7l, having been observed in one of two additional independently transformed T_0 lines generated during the course of the project.

T_1 Progeny Screening

A functional progeny screening protocol was developed based on selectable marker genes within the T-DNA construct. GFP expression was insufficient for selection at every stage of development observed during the course of tissue culture and T_1 progeny analysis. Spray applications of Liberty® herbicide were highly effective in selecting against wild type and *Ac* only seedlings (Figure 2A, B). The segregation ratio of resistant to sensitive genotypes varied with the family and ranged from 1.3:1 to 52:1 for selfed progeny of primary transgenic lines, and between 0.5:1 to 1.2:1 for T_1 progeny from WT × T_0 crosses (Table 1). Chi-square tests for the expected 3:1 segregation in T_0 selfs, or 1:1 in crosses to WT revealed that the expected segregation ratio only occurred in two families, 7k and 7o, including crosses made to wild type using 7o. Families 7 c, g, and m exhibited an excess of resistant progeny, whereas families 7b, h, and i demonstrated an excess of sensitive progeny. These trends closely match previously established grouping of the families, as those with no incidence of transposition followed expected segregation patterns, with the exception of 7h. No transpositions of *Ds* were identified among the progeny of 7h, but a significant excess of wild type progeny was found during herbicide screening. Families with an abundance of resistant phenotypes (7c, g, m) were previously grouped together due to a shared somatic transposition of *Ds*, whereas those families characterized by both active transposition and an abundance of sensitive phenotypes (7b and 7i) were grouped based on a high incidence of independent transpositions of *Ds* recovered among the T_1 progeny.

Due to the loss of GFP as a functional marker, a hygromycin leaf painting procedure was developed and refined to distinguish *Ds*-only individuals from *AcDs* and *Ac* + *Ds* seedlings. Necrotic spots visible on both the upper and lower leaf surfaces 5-7 days after painting revealed

the absence of the *Ac* element and its hygromycin resistance gene (Figure 2C, D). Leaf damage was strictly limited to areas painted with hygromycin solution, with no damage appearing on other parts of tested seedlings. Necrotic regions did not change in size after screening and would remain readily identifiable until the point of natural leaf senescence. The putative *Ds*-only transposants (herbicide resistant, hygromycin sensitive) were verified by PCR. The percentage of *Ds*-only T_1 progeny recovered from the total progeny of each family varied from 7 to 21% (Table 1). After common insertions were eliminated by zygosity PCR, the remaining independent *Ds*-only transpositions per family ranged from 3 to 9% (Table 1). The screening protocol was found to be 83% effective in identifying *Ds*-only individuals using the combination of herbicide spraying and hygromycin leaf painting. The protocol of combining zygosity primers with the PCR multiplex to identify common insertions prior to performing more expensive TAIL-PCR procedures was found to vary in efficiency by family and the reliability of zygosity primer design for the common insertion site (data not shown). Results from progeny screenings that did not follow the protocol specified in this paper were not considered in efficiency calculations.

Characteristics of Transposition

Amplification of genomic regions flanking insertions of the *Ds* element via transposition revealed incidences of insertion onto every pseudochromosome in the tomato genome (Figure 4) (Bombarely et al., 2011). When the distribution of insertions was broken down by T_0 family, the presence of a common insertion of *Ds* from a somatic transposition event on a particular chromosome often occurred in conjunction with higher incidences of independent insertions identified on that same chromosome (Table 2). As a whole, the construct engendered global transposition by distributing *Ds* elements throughout the genome. Analysis of genomic regions

flanking these insertion sites revealed a higher incidence of insertion into genes and their associated regulatory regions (60%) rather than into intergenic sequences (40%) (Table 3). Insertions into intergenic regions have been divided into those occurring greater than or less than 2 kb from the nearest described gene or putative gene model. Insertions greater than 2 kb from the nearest gene have been considered less likely to enhance transcription based on current understanding of activation tag kinetics (Kondou et al., 2010).

Phenotypic Mutants

Fifteen T_1 progeny from several actively transposing families were identified as having unique, dominant mutant phenotypes during screening**.** Most of these phenotypes involved leaf structural changes, yellow variegation, and changes in plant size. Loss of fertility due to flower abortion was also observed in several mutant lines. Among the most interesting of these lines was 7c-69 (Figure 5A), exhibiting elongated internodes, no lateral branching, crumpled leaves, and a weak stem. An equally complex phenotype was identified in 7i-271 (Figure 5D), which demonstrated short internodes, reduced leaves, and slow growth. A third phenotype cataloged was that of 7g-25 (Figure 5E), which was observed to have tightly rolled, pale leaves with a leathery texture, a shortened stature, and aborted flowers. Further analysis of these lines is required to verify how activation tags on transposed *Ds* elements are interacting with neighboring genes. Efforts to preserve sterile phenotypic mutants through *in vitro* propagation and cryogenic tissue stocks have been successful in most cases. In addition to these obvious mutants, the T_1 population exhibited many potential metabolic or more subtle mutant phenotypes that may be selected for detailed analysis based on insertion site. The population also contains knock-out mutants that may exhibit altered phenotypes once *Ds* insertions reach a homozygous state in T_2 progeny.

Chapter 5: Discussion

Although the *Agrobacterium*-mediated transformation efficiency was only a fraction of a percent using established protocols on M82 leaf pieces, it was possible to exploit the transposition frequency of a single T_0 line to initiate the development of a functional resource for activation tagging in tomato. The practice of using micropropagation to produce many clonal plants from a single tissue culture regenerant proved valuable, as it multiplied T_1 seed production by up to 25 times. This strategy also capitalized on the behavior of transposase in *Ac*-*Ds*ATag-Bar_gosGFP, by isolating chimeric tissue from the original transformant into separate plantlets, allowing germinal transposition from multiple sites of *Ds* integration. The selection of a self-fertile, true breeding tomato cultivar allowed crossing to non-transgenic M82, thus maximizing T_1 seed production. Pollen could be collected from transgenic flowers and distributed to multiple non-transgenic plants, all while still obtaining transgenic self-progeny.

Modifications made to the transformation protocol have yielded greater regeneration frequencies in ongoing research with more recently developed T-DNA constructs. These current efforts have also shown the modified protocol to be effective for other tomato cultivars. In addition to greater transformation frequencies, the modified protocol shortened the time transformed cell lines remain in tissue culture from a full year, as was the case with the transformed plants in the present study, to less than 6 months. The long callus phase subsequent to transformation likely resulted in undesirable somaclonal variation, as evidenced by two out of three regenerated T_0 lines being useless due to sterility and loss of the transposable element.

The lack of observable GFP expression proved to be a shortcoming of the *Ac*-*Ds*ATag-Bar_gosGFP construct in M82 tomato. This was compensated by the development of a hygromycin leaf painting procedure, for which a direct precedence in the literature could not be found. With refinement, this procedure provided effective screening when used in conjunction

with well-established glufosinate herbicide screening protocols. Even though the effectiveness of hygromycin leaf painting has not been evaluated in other crops, the unique and easily identifiable sensitivity reactions demonstrated in tomato give it the potential to be a widely applicable screening procedure. While the described herbicide and hygromycin screening protocol proved effective, both were sensitive to improper application. Herbicide sprays were easily rendered ineffective if solutions were not sufficiently absorbed by leaves prior to regular watering. More importantly, application of herbicide to older seedlings (>6 weeks, as pictured in Figure 2) often resulted in less obvious sensitive phenotypes, from which plants could recover to the point of being indistinguishable from resistant seedlings if phenotypic data were not collected within a week of treatment. Application of herbicide to 3-5 week old seedlings consistently resulted in complete necrosis of sensitive individuals, which rendered screening easier and more efficient. While hygromycin leaf painting proved less influenced by watering and evaluation time, sensitive phenotypes could be easily confused with other forms of leaf damage that may occur in a greenhouse setting. Screening results from T_1 progeny in which complications arose that confounded effective data collection due to the aforementioned factors were excluded from efficiency calculations (Table 1).

The dispersal of transposed *Ds* elements in the T_1 population produced during the course of this project is encouraging for the implementation of a high-throughput system intended to produce larger numbers of activation tagged mutants. Most importantly, the *Ds* element has been shown to globally transpose to all twelve tomato chromosomes from its initial launch pad site on chromosome 5 (Figure 4). When the distribution of these insertions was taken into consideration with other data from T_1 progeny screening, further insights into the nature of transposition in these lines was revealed. Given that no incidences of the *Ds* element remaining in the original

T-DNA insertion site were detected within thousands of progeny from actively transposing families, it can be deduced that unique transposition events were the result of multiple movements of the *Ds* element during the lifespan of the T_0 plant. This scenario is further supported by the family specific detection of disproportionately more unique insertions onto chromosomes also harboring a common insertion (Table 2). The implications of these data are that a given *Ds* element insertion site could be effectively utilized as a new launch pad for future insertions, simply by reintroducing the *Ac* element and its embedded transposase gene through cross-pollination of *Ds*-only plants with *Ac*-only plants. With careful selection, lines containing a *Ds* element on a target chromosome or euchromatic region could be utilized to produce more unique mutants. These mutants would contain *Ds* activation tags preferentially distributed to the targeted genomic region neighboring the originally selected *Ds* insertion.

The reliable generation of unique transposition events in this mutant population may in part be a function of a fortuitous insertion of the original T-DNA vector harboring the conjoined *Ac* and *Ds* elements. In theory, transposase expression should be consistent with the *Ac* promoter driving transcription of its coding sequence. In the case of the *Ac*-*Ds*ATag-Bar_gosGFP construct, the promoter for transposase was the native *Zea mays* promoter, which has long been characterized to drive expression throughout embryo development (Scofield et al., 1992). It is also known that regions flanking a T-DNA insertion site may influence transgene expression, as observed with promoter and enhancer trapping schemes (Topping et al., 1994; Wu et al., 2003). Expression profiling data for the Arabidopsis ortholog of the 60S ribosomal protein L29 into which the T-DNA inserted dictated widespread transcription of the gene, with the greatest expression occurring in shoot tips. The influence of a promoter region driving somatic expression of transposase is consistent with the frequent somatic transpositions of *Ds* detected in

each T_0 family. Furthermore, the frequency of germinal transposition may have been enhanced by pollen specific cis-regulatory elements on adjacent genes. Interestingly, orientation of the T-DNA with respect to the neighboring late embryogenesis abundant protein places the native *Ac* promoter within 2 kb of these regulatory elements following an initial excision and transposition of *Ds*. The potential influence of these elements is consistent with the clear tendency for frequent germinal transposition of *Ds* to occur in T_0 families where an initial transposition event had already taken place. These putative interactions between genomic elements and transposase expression are supported by the current body of evidence concerning the influence of genomic elements gene expression within T-DNA insertions (Filipecki and Malepszy, 2006; Mirza, 2005).

Classification of *Ds* insertions by their flanking regions demonstrated insertion site preferences reasonably consistent with results from screens of over 10,000 tagged lines in Arabidopsis, favoring insertion into genes and euchromatin (Kuromori et al., 2004). Of the 309 identified *Ds* insertions, 31% were in regions suitable for functional activation tagging (promoter, 3'UTR, <2 kb intergenic) (Table 3). An additional 28% of the *Ds* insertions were within described or putative coding regions. While such locations may be less accommodating to activation tagging, these lines provide a potential source of knock-out mutants for further study. The *Ds* insertions into introns are of unknown value, as such insertions may result in reduced function or activation tagged phenotypes. The remaining 27% of transposed *Ds* elements are those inserted into intergenic regions that were greater than 2 kb from the nearest described gene, rendering these lines of unknown value. Some insertions under this classification may be sufficiently close to one or more genes to influence expression, whereas others are embedded into repetitive regions, limiting opportunity for interaction with native genes. The implication of these insertion site preferences is that up to 73% of the unique, activation tagged progeny

produced by this population are potential phenotypic mutants. In addition, the tendency for *Ds* to transpose to new locations from a previous insertion site on a given chromosome holds promise for the development of new transgenic lines. These new lines that recombine *Ac* and *Ds* into the same genome through crossing could produce their own progeny, in which transpositions of *Ds* preferentially saturate a targeted region of the genome. Continued research is required to evaluate the feasibility and limitations of this strategy. Finally, the development of an effective progeny screening protocol makes this population and its derivatives capable of functioning in a high-throughput system. These factors make the transgenic population developed during the course of this project an effective resource for functional genomics.

Supplemental Materials and Methods

CTAB DNA Extraction Protocol

Steel shot pellets used for tissue grinding were added to 2 ml micro centrifuge tubes prior to tissue collection. A single expanded terminal leaflet was collected from each plant selected for DNA extraction and samples were immediately placed into liquid nitrogen. Samples were ground by oscillating tubes loaded with shot at 400 rpm for 30 s. Steel shot pellets were removed and samples were stored at -80°C prior to DNA extraction. For isolation of DNA, 500 µl of extraction buffer (100 mM TRIS-HCl pH 8.9, 1.4 M NaCl, 20 mM EDTA pH 8.0, 2.0% w/v CTAB, 1.0% w/v PVP-40, 0.2% v/v β-mercaptoethanol) was added to each sample and samples were incubated at 65°C for 1 h with regular agitation. Following incubation, 500 µl of 24:1 chloroform:iso-amyl alcohol were added to each sample. Phase separation was accomplished through centrifugation at 13,000 rpm for 5 min. 400 µl of the top aqueous phase was removed to a new 2 ml micro centrifuge tube for further use. DNA was precipitated by adding 800 µl of chilled 100% ethanol and 200 µl of 5 M NaCl, followed by centrifugation at 14,000 rpm for 10 m at 4°C. Liquid components were then discarded without dislodging precipitated DNA pellets. DNA pellets were then washed by adding 500 µl of chilled 70% ethanol and centrifuged at 14,000 rpm for 5 min at 4°C. The ethanol was then poured off, while the precipitated DNA was dehydrated at room temperature overnight. Dried DNA pellets were then suspended in 100 µl TE/RNase (10 mM TRIS-HCl pH 8, 1 mM EDTA pH 8, 20 mg/L RNase A) prior to PCR analysis. Total DNA yield varied from 10 to 150 µg.

Extraction of PCR Bands from Agarose Gel

TAIL-PCR products were run on agarose gels (1% agarose/1× TAE gel with 0.1% EtBr) for 70 min at 150 V/400 A. Bands were visualized under UV light and selected for sequencing based on apparent magnitude and by comparison to banding patterns produced by other samples. Selected bands were cut from gels in 300 mg slices in preparation for DNA isolation. Gel slices were dissolved by incubating in 300 µl Buffer QG (5.5 M guanidine thiocyanate, 20 mM TRIS-HCl, pH 6.6) for 10 min at 50°C, followed by the addition of 100 µl of isopropyl alcohol after samples had cooled. Samples were then applied to an Econospin™ DNA Spin Column (Epoch Life Science Inc., Sugar Land, TX) and centrifuged at 13,000 rpm for 1 min. Samples were washed by adding an additional 500 µl of Buffer QG and centrifuged under the same parameters. A second wash using 700 µl of Buffer PE (80% ethanol, 20 mM NaCl, 2 mM TRIS-HCl, pH 7.5) was applied to each sample, followed by centrifugation at 13,000 rpm for 1 min. Samples were dried overnight in their respective spin columns at room temperature. DNA was eluted from spin column membranes by adding 20 µl TE Buffer (10 mM TRIS-HCl pH 8, 1 mM EDTA pH 8) and centrifuging under the previous parameters. Band yield varied from between 180 to 600 ng.

References

Alonso J.M., Stepanova A.N., Leisse T.J., Kim C.J., Chen H., Shinn P., Stevenson D.K., Zimmerman J., Barajas P., Cheuk R. (2003) Genome-Wide Insertional Mutagenesis of *Arabidopsis thaliana*. Science 301:653.

Arabidopsis G.I. (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796.

Baker B., Schell J., Lörz H., Fedoroff N. (1986) Transposition of the maize controlling element "Activator" in tobacco. P.N.A.S. 83:4844.

Barone A., Di Matteo A., Carputo D., Frusciante L. (2009) High-Throughput Genomics Enhances Tomato Breeding Gfficiency. Curr. Genomics 10:1.

Becker H.A., Kunze R. (1997) Maize *Activator* transposase has a bipartite DNA binding domain that recognizes subterminal sequences and the terminal inverted repeats. Mol. Gen. Genet. 254:219-230.

Bombarely A., Menda N., Tecle I.Y., Buels R.M., Strickler S., Fischer-York T., Pujar A., Leto J., Gosselin J., Mueller L.A. (2011) The Sol Genomics Network (solgenomics.net): growing tomatoes using Perl. Nucleic Acids Res. 39:D1149.

Castañeda-Ovando A., Pacheco-Hernández M.L., Páez-Hernández M.E., Rodríguez J.A., Galán-Vidal C.A. (2009) Chemical studies of anthocyanins: A review. Food Chem. 113:859-871.

Doyle J.J., Doyle J.L. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19:11-15.

Emmanuel E., Levy A.A. (2002) Tomato mutants as tools for functional genomics. Curr. Opin. Plant Biol. 5:112-117.

Eshed Y., Zamir D. (1995) An Introgression Line Population of *Lycopersicon pennellii* in the Cultivated Tomato Enables the Identification and Fine Mapping of Yield-Associated QTL. Genetics 141:1147.

FAOSTAT. (2011) Food and Agriculture Organization of the United Nations

Chief, Publishing Policy and Support Branch, Office of Knowledge Exchange, Research and Extension, FAO, Viale delle Terme di Caracalla, 00153 Rome, Italy.

Feuillet C. (2012) The international effort to sequence the 17Gb wheat genome: Yes, Wheat can!, Plant and Animal Genome.

Filipecki M., Malepszy S. (2006) Unintended consequences of plant transformation: a molecular insight. J. Appl. Genet. 47:277-286.

Fladung M., Polak O. (2011) Activation tagging in poplar by using an inducible Ac/Ds transposon system, BioMed Central Ltd. pp. O51.

Giménez E., Pineda B., Capel J., Antón M.T., Atarés A., Pérez-Martín F., García-Sogo B., Angosto T., Moreno V., Lozano R. (2010) Functional analysis of the *Arlequin* Mutant Corroborates the Essential Role of the *Arlequin/TAGL1* Gene during Reproductive Development of Tomato. PLoS ONE 5:e14427.

Higo K., Ugawa Y., Iwamoto M., Korenaga T. (1999) Plant cis-acting regulatory DNA elements (PLACE) database: 1999, Nucleic Acids Res., National Institute of Agrobiological Sciences. pp. 297-300.

Hsing Y.I., Chern C.G., Fan M.J., Lu P.C., Chen K.T., Lo S.F., Sun P.K., Ho S.L., Lee K.W., Wang Y.C. (2007) A rice gene activation/knockout mutant resource for high throughput functional genomics. Plant Mol. Biol. 63:351-364.

Huang X., Feng Q., Qian Q., Zhao Q., Wang L., Wang A., Guan J., Fan D., Weng Q., Huang T. (2009) High-throughput genotyping by whole-genome resequencing. Genome Res. 19:1068-1076.

Izawa T., Miyazaki C., Yamamoto M., Terada R., Iida S., Shimamoto K. (1991) Introduction and transposition of the maize transposable element *Ac* in rice (*Oryza sativa* L.). Molecular and General Genetics MGG 227:391-396.

Jeong D.H., An S., Kang H.G., Moon S., Han J.J., Park S., Lee H.S., An K., An G. (2002) T-DNA Insertional Mutagenesis for Activation Tagging in Rice. Plant Physiol. 130:1636-1644.

Jeong D.H., An S., Park S., Kang H.G., Park G.G., Kim S.R., Sim J., Kim Y.O., Kim M.K. (2006) Generation of a flanking sequence-tag database for activation-tagging lines in japonica rice. Plant J. 45:123-132.

Khoudi H., Nouri-Khemakhem A., Gouiaa S., Masmoudi K. (2009) Optimization of regeneration and transformation parameters in tomato and improvement of its salinity and drought tolerance. Afr. J. Biotechnol. 8:6068-6076.

Kolesnik T., Szeverenyi I., Bachmann D., Kumar C.S., Jiang S., Ramamoorthy R., Cai M., Ma Z.G., Sundaresan V., Ramachandran S. (2004) Establishing an efficient *Ac/Ds* tagging system in rice: large-scale analysis of *Ds* flanking sequences. Plant J. 37:301-314.

Kondou Y., Higuchi M., Matsui M. (2010) High-Throughput Characterization of Plant Gene Functions by Using Gain-of-Function Technology. Annu. Rev. Plant Biol. 61:373-393.

Kuromori T., Hirayama T., Kiyosue Y., Takabe H., Mizukado S., Sakurai T., Akiyama K., Kamiya A., Ito T., Shinozaki K. (2004) A collection of 11 800 single-copy *Ds* transposon insertion lines in *Arabidopsis*. Plant J. 37:897-905.

Lee S., Jeon U.S., Lee S.J., Kim Y.K., Persson D.P., Husted S., Schjørring J.K., Kakei Y., Masuda H., Nishizawa N.K. (2009) Iron fortification of rice seeds through activation of the *nicotianamine synthase* gene. P. Natl. Acad. Sci. USA 106:22014.

Liu Y.G., Mitsukawa N., Oosumi T., Whittier R.F. (1995) Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. Plant J. 8:457-463.

Liu Y.G.C., Y. (2007) High-efficiency thermal asymmetric interlaced PCR for amplification of unknown flanking sequences. BioTechniques 43:649-650, 652, 654.

Massa A.N., Childs K.L., Lin H., Bryan G.J., Giuliano G., Buell C.R. (2011) The Transcriptome of the Reference Potato Genome Solanum tuberosum Group Phureja Clone DM1-3 516R44. PloS one 6:e26801.

Mathews H., Clendennen S.K., Caldwell C.G., Liu X.L., Connors K., Matheis N., Schuster D.K., Menasco D., Wagoner W., Lightner J. (2003) Activation Tagging in Tomato Identifies a Transcriptional Regulator of Anthocyanin Biosynthesis, Modification, and Transport. Plant Cell 15:1689-1703.

Meissner R., Chague V., Zhu Q., Emmanuel E., Elkind Y., Levy A.A. (2000) A high throughput system for transposon tagging and promoter trapping in tomato. Plant J. 22:265-274.

Menda N., Semel Y., Peled D., Eshed Y., Zamir D. (2004) *In silico* screening of a saturated mutation library of tomato. Plant J. 38:861-872.

Mirza B. (2005) Influence of the Nature of the T-DNA Insertion Region on Transgene Expression in *Arabidopsis thaliana*. Russ. J. Genet. 41:1322-1328.

Mueller L.A.L., Tanksley R.K., Giovannoni S.D., White J.J., Vrebalov R., Fei J., van Eck Z., Buels J., Mills R., Menda A.A. (2009) A Snapshot of the Emerging Tomato Genome Sequence. The Plant Genome 2:78.
Murashige T., Skoog F. (1962) A revised medium for rapid growth and bio assays with tobacco tissue cultures. Physiol. Plant. 15:473-497.
Rubin E., Levy A.A. (1997) Abortive Gap Repair: Underlying Mechanism for *Ds* Element Formation. Mol. Cell. Biol. 17:6294.
Scofield S.R., Harrison K., Nurrish S.J., Jones J.D.G. (1992) Promoter Fusions to the *Activator* Transposase Gene Cause Distinct Patterns of *Dissociation* Excision in Tobacco Cotyledons. Plant Cell 4:573-582.
Stevens M.A., Dickinson G.L., Aguirre M.S. (1982) UC 82 : a high yielding processing tomato University of California, Davis.
Suzuki Y., Uemura S., Saito Y., Murofushi N., Schmitz G., Theres K., Yamaguchi I. (2001) A novel transposon tagging element for obtaining gain-of-function mutants based on a self-stabilizing *Ac* derivative. Plant Mol. Biol. 45:123-131.
Topping J.F., Agyeman F., Henricot B., Lindsey K. (1994) Identification of molecular markers of embryogenesis in *Arabidopsis thaliana* by promoter trapping. Plant J. 5:895-903.
Trijatmiko K.R. (2005) Comparative analysis of drought resistance genes in Arabidopsis and rice.
Tsugeki R., Kochieva E.Z., Fedoroff N.V. (1996) A transposon insertion in the *Arabidopsis SSR16* gene causes an embryo-defective lethal mutation. Plant J. 10:479-489.
Turner T.L., Bourne E.C., Von Wettberg E.J., Hu T.T., Nuzhdin S.V. (2010) Population resequencing reveals local adaptation of *Arabidopsis lyrata* to serpentine soils. Nat. Genet. 42:260-263.
Walden R., Fritze K., Hayashi H., Miklashevichs E., Harling H., Schell J. (1994) Activation tagging: a means of isolating genes implicated as playing a role in plant growth and development. Plant Mol. Biol. 26:1521-1528.
Wang Z., Gerstein M., Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nat. Rev. Genet. 10:57-63.
Weigel D., Ahn J.H., Blázquez M.A., Borevitz J.O., Christensen S.K., Fankhauser C., Ferrándiz C., Kardailsky I., Malancharuvil E.J., Neff M.M. (2000) Activation Tagging in Arabidopsis. Plant Physiol. 122:1003-1014.
Winter D., Vinegar B., Nahal H., Ammar R., Wilson G.V., Provart N.J. (2007) An “Electronic Fluorescent Pictograph” Browser for Exploring and Analyzing Large-Scale Biological Data Sets, PLoS ONE, The Bio-Array Resource for Plant Biology. pp. e718.
Wu C., Li X., Yuan W., Chen G., Kilian A., Li J., Xu C., Zhou D.X., Wang S. (2003) Development of enhancer trap lines for functional analysis of the rice genome. Plant J. 35:418-427.
Xu G., Chague V., Melamed-Bessudo C., Kapulnik Y., Jain A., Raghothama K.G., Levy A.A., Silber A. (2007) Functional characterization of LePT4: a phosphate transporter in tomato with mycorrhiza-enhanced expression. J. Exp. Bot. 58:2491.
Xu X., Pan S., Cheng S., Zhang B., Mu D., Ni P., Zhang G., Yang S., Li R., Wang J. (2011) Genome sequence and analysis of the tuber crop potato. Nature 475:189-195.

Tables and Figures

Table 1. Evaluation of finalized T_1 progeny screening protocol applied to progeny derived by self-pollination of eight T_0 families and cross-pollination to WT M82. All T_0 plants contained a non-mobile *Ac* element and a mobile *Ds* element. The progeny were screened for resistance to Liberty® herbicide to identify those with the *Ds* element, followed by hygromycin leaf painting of herbicide resistant seedlings to identify those with the *Ac* element. Screening efficiency measured the percentage of seedlings identified as putative transposants (herbicide resistant, hygromycin sensitive) that were verified as *Ds*-only by multiplex PCR. Totals do not represent all T_1 progeny screened during project.

T_0 Family	BAR+	BAR-	Total	BAR+:1 BAR-	X^2	hyg-	*Ds*-only	*Ds*-only (%)	Independent *Ds*-only	Independent Ds-only (%)	Screening efficiency (%)
7b	159	78	237	2.0	7.5*	21	18	8	11	61	86
7c	176	21	197	8.4	20.9**	47	41	21	18	44	87
7g	234	22	256	10.6	35.9**	58	49	19	20	41	85
7h	172	128	300	1.3	49.0**	0	0	NA	0	NA	NA
7i	724	371	1095	2.0	45.6**	146	120	13	40	33	82
WT × 7i	240	400	640	0.6	39.5**	52	42	7	25	60	80
7k	253	65	318	3.9	3.3ns	0	0	NA	0	NA	NA
7m	208	4	212	52.0	59.2**	41	31	19	7	23	76
7o	81	34	115	2.4	1.0ns	0	0	NA	0	NA	NA
WT × 7o	267	293	560	0.9	1.1ns	0	0	NA	0	NA	NA
Total	2514	1416	3930			365	301		121		83

Chi square expect 3:1 BAR+/BAR- for T_0 selfs, 1:1 for WT × T_0 crosses
** = $p < 0.01$
* = $p < 0.05$
ns = not significant, $p > 0.05$
Ds-only (%) = *Ds*-only / Total *100
Screening Efficiency, (*Ds*-only / hyg-) *100
NA = Not Applicable

Table 2. Distribution of transposed *Ds* elements to tomato pseudochromosomes (Tomato WGS Chromosomes SL2.40), sorted by T_0 family. Includes T_0 self and WT × T_0 progeny. Underlined numbers indicate the presence of an identified common (somatic) transposition event on the chromosome in the respective cutting family. Common insertions are counted as one independent transposition event.

	Chromosome												
T_0 Family	1	2	3	4	5	6	7	8	9	10	11	12	Total
7a	8	0	2	3	0	1	2	0	5	1	0	3	25
7b	18	2	0	4	1	1	1	5	3	3	3	2	43
7c	3	2	2	1	1	3	0	11	1	1	10	1	36
7g	9	4	6	6	4	2	2	20	1	4	15	3	76
7i	58	7	5	5	5	4	5	4	4	3	2	4	106
7m	12	1	1	4	0	0	0	4	0	0	1	0	23
Total	108	16	16	23	11	11	10	44	14	12	31	13	309

Table 3. Insertion site classification of 309 identified *Ds* transposition events based on Tomato WGS Chromosomes SL2.40. Insertion site preference displayed as a percentage of the 309 total insertions.

	Coding	Intron	Promoter	3'UTR	Intergenic	
					<2kb	>2kb
Ds Insertions	86	46	38	15	42	82
Percentage (%)	28	15	12	5	13	27

Coding = exons, unigenes, cDNA clones, and unknown protein models
Promoter = up to one Kb upstream from nearest gene model
3'UTR = up to 500 bp downstream from nearest gene model
Intergenic = distance to nearest gene model, upstream or downstream

Figure 1: The *Ac-Ds*ATag-Bar_gosGFP construct. Elements are as follows; LB = Left Border, Hyg = hygromycin resistance, 35SPr = Cauliflower Mosaic Virus 35*S* promoter, GFP = Green Fluorescent Protein, GosPr = *Zea mays* promoter, AcTPase = *Zea mays* Transposase, AcPr = *Zea mays* transposase native promoter, IR = Inverted Repeat, 35S 4 Enh = Four tandem copies of Cauliflower Mosaic Virus 35*S* enhancer, BAR = glufosinate resistance, UbiPr = Ubiquitin promoter, IR = Inverted Repeat, RB = Right Border.

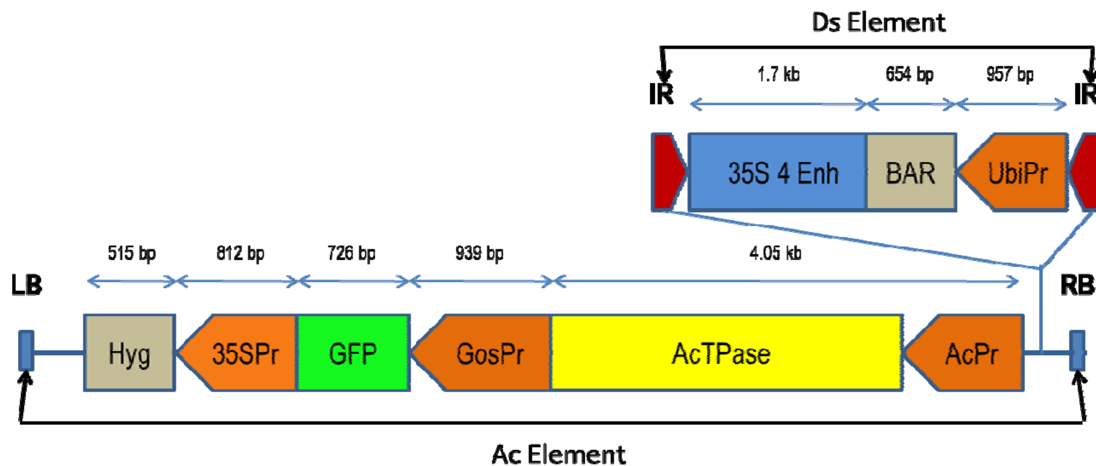


Figure 2. Phenotypic Screening of T_1 Progeny. (A) Herbicide resistant seedling (left) next to sensitive seeding (right) 7 days after spraying. (B) Close up of herbicide resistant leaf (left) next to sensitive leaf (right) removed from previous seedlings. (C) Adaxial side of four leaves detached from different seedlings sensitive to hygromycin 6 days after painting. Reaction to hygromycin localized to exact region of leaf that was painted. (D) Abaxial side of a leaf sensitive to hygromycin displaying the diagnostic necrotic spots that distinguish hygromycin sensitivity from common leaf damage due to pathogens and mechanical injury.

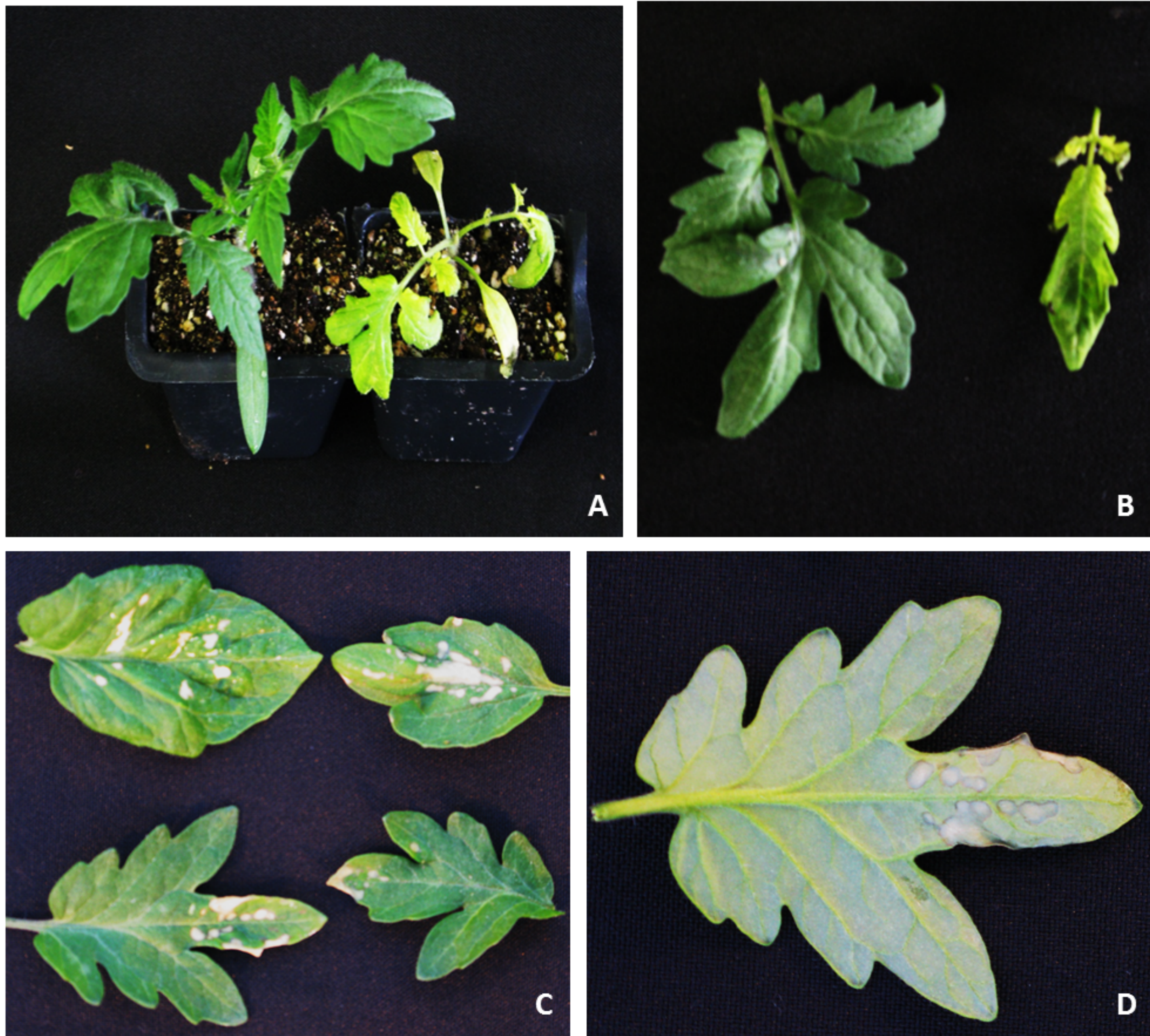


Figure 3. Analysis of T_0 transgenic line 7a by rtPCR, screening of T_1 progeny through a combined PCR multiplex and zygosity test, and TAIL-PCR analysis of *Ds*-only T_1 progeny. (A) Expression of transposase in whole inflorescence RNA. PCR product spans transposase intron 2 (DNA product = 476 bp, cDNA product = 405 bp). Positive control DNA and cDNA extracted from transgenic potato inflorescence. (B) Visualization of primary TAIL-PCR products. (C) Screening of T_1 progeny by multiplex PCR and zygosity testing. SQE = squalene epoxidase (distinguishes wild types from failed reactions), CI#1 = common insertion one from a somatic transposition event early in development (defines T_0 group 2), BAR = glufosinate herbicide resistance gene (*Ds* element), HPT = hygromycin resistance gene (*Ac* element), CI#2 = common insertion two from somatic transposition event unique to T_0 7m.

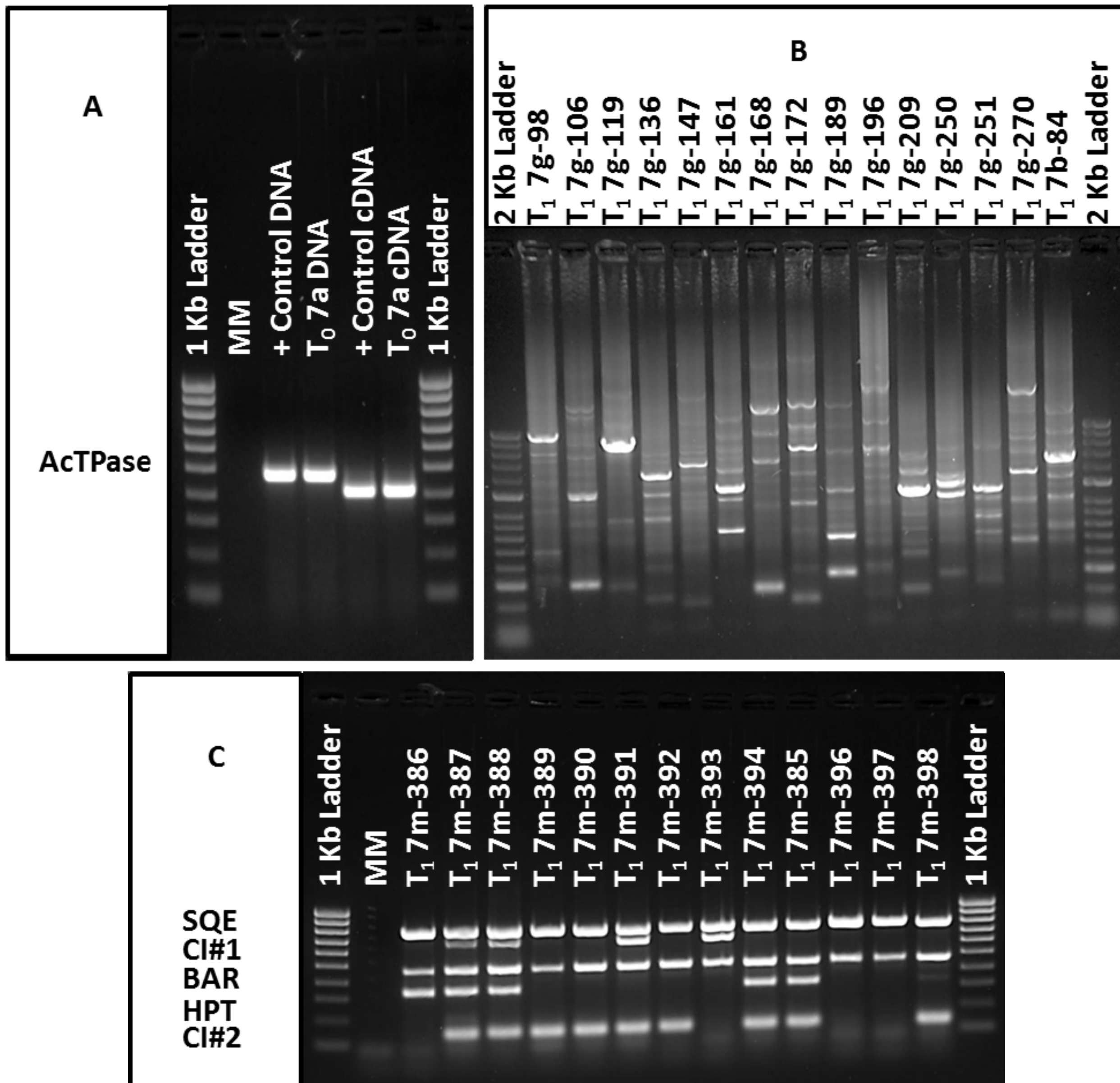


Figure 4. Graphic distribution of 309 transposed *Ds* elements to tomato pseudochromosomes (Tomato WGS Chromosomes SL2.40). Includes self and WT × T_0 progeny from every T_0 family. Both common and unique insertions are represented by a single band. Original T-DNA insertion site during transformation (*Ac* element) denoted by ←**Ac** on Chromosome 5.

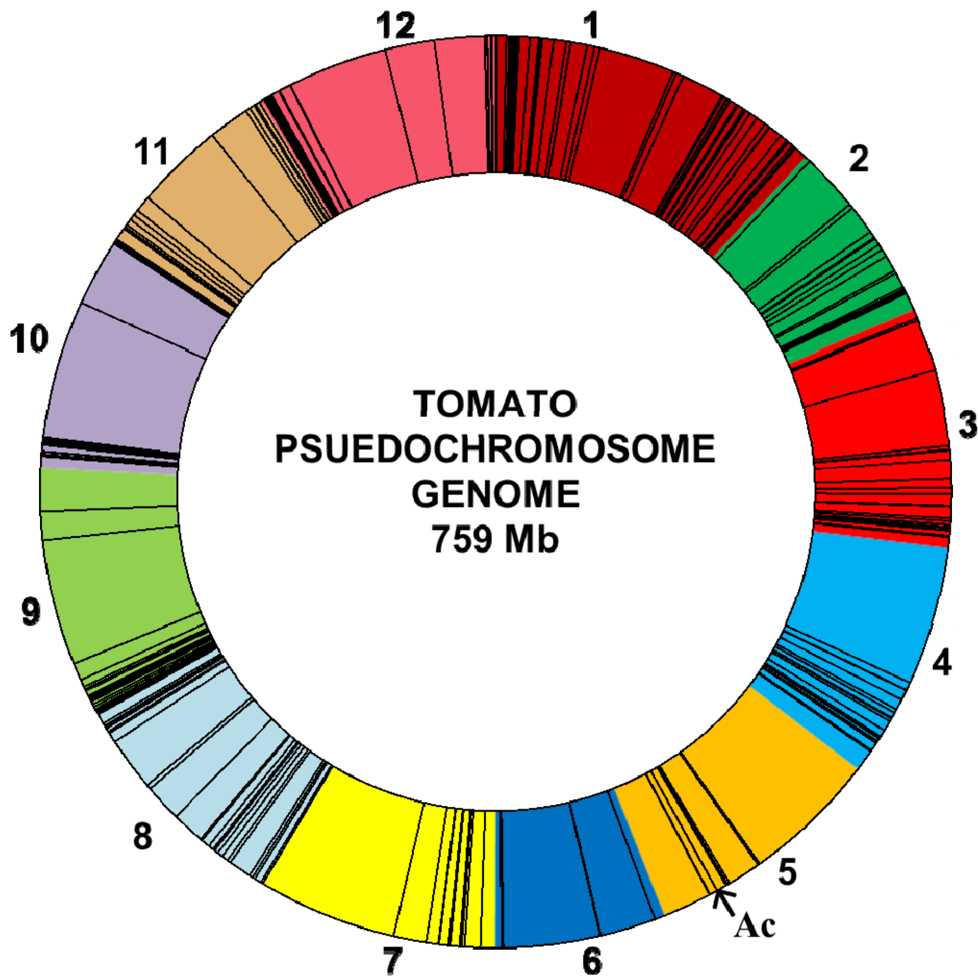
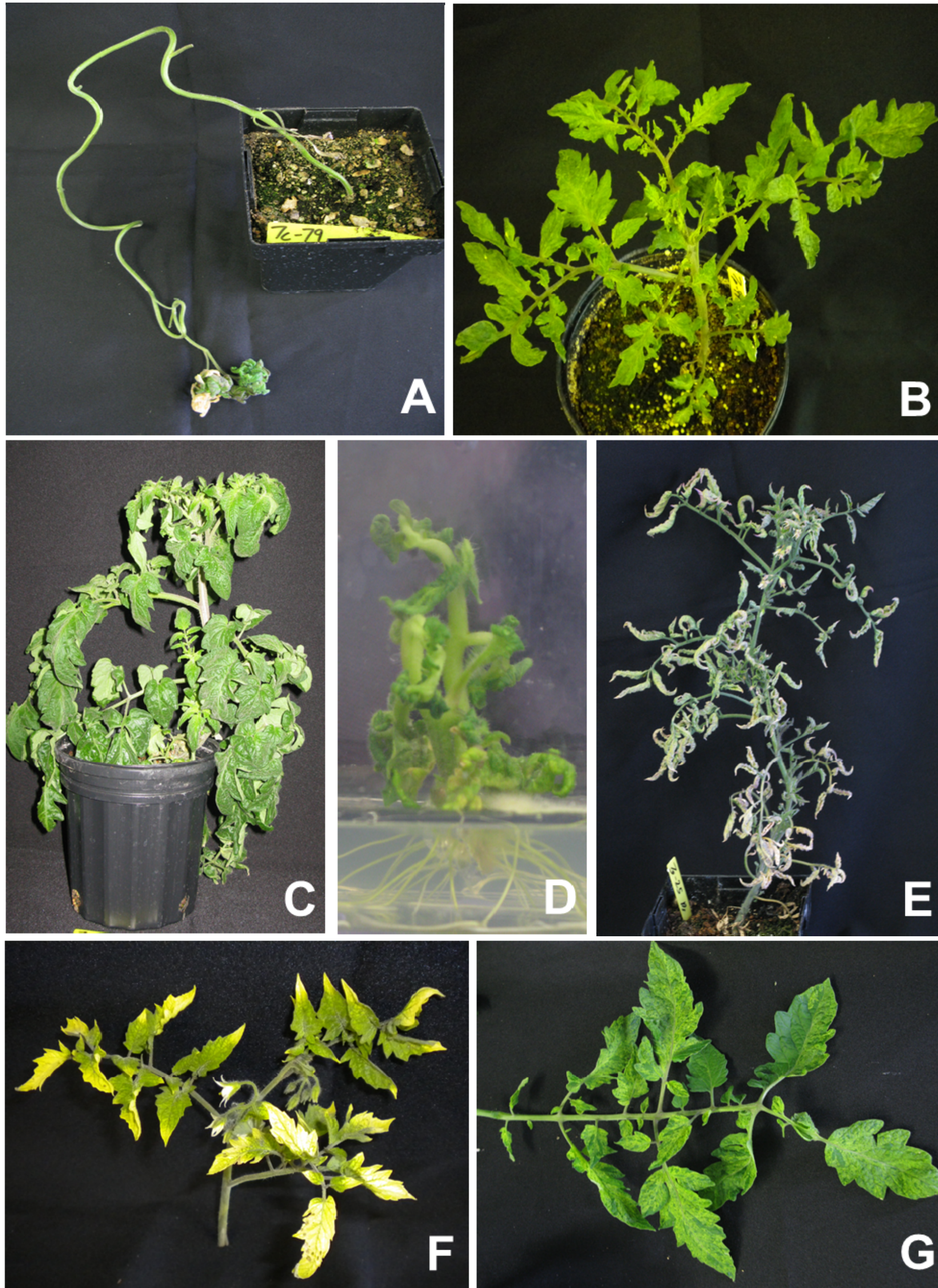


Figure 5. Phenotypes for seven T_1 activation tagged mutants. (A) 7c-79, rolled leaves, no lateral branching, weak stem. (B) 7p-99, dwarf, warped leaves, variegation. (C) 7g-300 crumpled leaves. (D) 7i-271, twisted leaves, extreme dwarfism. (E) 7g-25, tightly rolled, leathery leaves, dwarf, aborted flowers. (F) WT7i-48, extensive yellow variegation, aborted flowers. (G) 7p-3 speckled variegation.



Appendix A. Tissue Culture Media Formulations

Table 4. Tissue culture media formulations for Agrobacterium mediated transformation of M82 tomato primary leaf explants. Modified from Khoudi et al., (2009). All media formulations were adjusted to pH 5.7-5.8.

Ingredient	Unit	Pre-culture media	Dilution media	Wash off media	Shoot regeneration media	Shoot elongation media	Rooting media	Peat plug hydration media
MS+V	g/L	4.43	4.43	4.43	4.43	4.43	4.43	4.43
Sucrose	g/L	30	30	30	30	30	30	-
Thiamine	mg/L	0.9	0.9	0.9	0.9	0.9	0.9	0.9
Phytagel	g/L	4	-	-	4	4	4	-
BA	mg/L	1	-	-	-	-	-	-
NAA	mg/L	1	-	-	-	-	-	-
IAA	mg/L	-	-	-	0.1	0.1	1	-
Zeatin	mg/L	-	-	-	2	0.2	-	-
Hygromycin	mg/L	-	-	-	10	20	-	-
Cefotaxime	mg/L	-	-	500	500	500	500	-

MS+V = Murashige & Skoog Basal Medium (1962) with vitamins (*Phyto*Technology Laboratories, Shawnee Mission, KS)

Phytagel = Gellan Gum (Sigma-Aldrich, St. Louis, MO)

Appendix B. Primer Sequences

Table 5. Primer sequences.

Primer Name	Sequence
SL_SQEF	TGGGGTTCGTTGCAGTTTTC
SL_SQER	CGCGTTGAGCATCAATTTTCTC
G38BARF	CTGAAGTCCAGCTGCCAGAAACC
G38BARR	CTGCACCATCGTCAACCACTACAT
HPT5F	TCACAGTTTGCCAGTGATAC
HPT5R	ATCGTTATGTTTATCGGCAC
ubiTAIL2a	ACTTAGACATGCAATGCTCATTATCTC
*Ds*5'-2	ACGATGGACTCCAGTCCGGCCCCGTTTTGTATATCCCGTTTCCGT
*Ds*5'-3	TACCTCGGGTTCGAAATCGAT
*Ds*3’-1	ACCCGACCGGATCGTATCGGT
*Ds*3’-2	ACGATGGACTCCAGTCCGGCCCGATTACCGTATTTATCCCGTTC
*Ds*3’-3	GTATTTATCCCGTTCGTTTTCGT
AC1	ACGATGGACTCCAGAG
LAD1-1	ACGATGGACTCCAGAGCGGCCGCVNVNNNGGAA
LAD1-3	ACGATGGACTCCAGAGCGGCCGCVVNVNNNCCAA
AcTpaseRNAF	TTTCTTTCATGTGAGGTGTGCTTGTC
AcTpaseRNAR	GCAGTGGAATATTGAGTACCAGATAGGAG