An Assessment of Genomic Sequence Restoration in *Arabidopsis thaliana*

by

Aaron Munir Khalid

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Science
in
Biology

Waterloo, Ontario, Canada, 2009

AUTHOR'S DECLARATION

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

A fundamental tenet of classical Mendelian genetics is that allelic information is stably transmitted from parent to progeny. Work in our laboratory has revealed a novel exception to this law where *Arabidopsis thaliana* plants homozygous for the recessive organ-fusion mutation *hothead* (*hth*) gave rise to phenotypically and genotypically wild-type (*HTH*) progeny at high frequencies. We have coined the term *restoration* to describe this phenomenon, since the reverted *HTH* allele was not detectable in the parental genome but was present in a recent ancestor (the grandparent). Recent work in our laboratory has demonstrated that 45-94 bp insertions and deletions (indels) can also restore, irrespective of their genomic location. The work described in this thesis expands our understanding of restoration by characterizing previously identified non-parental loci at the molecular level, and monitoring the inheritance of native and transgenic alleles in *hth* mutant and wild-type genetic backgrounds.

Two – eight hundred bp genomic intervals containing non-parental loci were cloned and sequenced. This revealed that the tracts of sequence which had been reinstated were identical in phase and sequence composition to the corresponding grandparental sequences. Furthermore, molecular markers flanking non-parental loci were profiled across 80-90 kb chromosomal regions. In all cases, the flanking markers reverted concurrently, suggesting that restoration can affect comparatively large genomic regions. However, it is not clear if flanking markers revert as a result of multiple independent events or, alternatively, are the result of one continuous restoration event.

A number of individuals studied in this thesis are genetic mosaics, wherein the restoration events are localized within a single individual. Genetic mosaicism cannot be attributed to pollen contamination, and provides the strongest evidence to-date that restoration is a genuine and novel biological phenomenon.

The inheritance of a transgenic allele and two native alleles was monitored in pedigrees compromising a number of distinct ancestries in *hth-4*, *hth-8*, and wild-type genetic backgrounds. Although none of the F3 progeny exhibited atypical segregation of the investigated alleles, molecular screening may have revealed localized (mosaic) restoration of the transgenic marker. However, these results remain inconclusive based on results obtained in a negative control experiment.

Several significant conclusions can be derived from the work described in this thesis: (1) Restoration is a highly specific template-directed process. The template is likely of ancestral origin, although the nature of the template and the precise mechanism of restoration remain unclear. (2) Restoration frequently gives rise to individuals that are genetic mosaics, a finding that cannot be attributed to outcrossing. (3) Restored sequences are more readily identified by molecular genotyping than phenotypic screening. Possible mechanisms and recommended future studies are discussed.

Acknowledgements

I would like to express my sincere gratitude to both the present and former members of our laboratory who have contributed to this project. Technical assistance provided by Meghan Doerr, Emily Manning, and Bounmy Inthavong significantly contributed to the work which is presented here. Work conducted by Dulcie Lai several years ago provided a great contribution to the work which is done in our laboratory on a daily basis. Ryan Lee, a former student of Purdue University, provided exceptional data which was fundamental to the interpretation of my own. I am grateful for the many helpful discussions which I have shared with my colleagues Pearl Chang and Rebecca 'Becci' Rooke. Most importantly, I would like to thank all of my labmates for the friendship and personal support they have provided throughout my career as a grad student.

I cannot begin to express my gratitude for the assistance and friendship which I have received from Dr. Marianne Hopkins, who was a Post-Doctoral Fellow in our laboratory during the time I spent there. All of the plant material which formed the basis of Chapter 2 of this thesis was the fruit of her labor – this project truly would not have been possible without her.

Of course, I would also like to thank my supervisor Dr. Susan Lolle for accepting me as a student, mentoring me throughout this unique and challenging project, and tirelessly editing my thesis. Her exceptional skills as a scientist and teacher, and her unwavering dedication to her students and research truly distinguish her as a great asset to the University of Waterloo's Department of Biology.

My committee members Dr. Bruce Reed and Dr. John Semple were of the utmost help throughout this project, and I am very thankful for their time and efforts. In particular, I would like to thank Dr. Semple for his support during my early days as a 'wandering' graduate student, and for his continued friendship throughout the years.

I am grateful for generous funding which was provided in the form of Ontario Graduate Scholarships granted by the Ontario Ministry of Training, Colleges and Universities, President's Graduate Scholarships granted by the University of Waterloo, and project funding provided by the Natural Sciences and Engineering Research Council of Canada.

Finally, I would like to thank my friends and family for all of your help and support throughout the years. Thank you for bearing with me throughout the final months of my graduate work, when I often found myself distracted from the things that really matter.

Dedication

I dedicate this thesis to the memory of my grandmother Janina (Jane) Feskun who passed December 12th, 2008, after 85 wonderful years.

Table of Contents

List of Figures

List of Tables

Chapter 1
Introduction

1.1 Background

According to Mendelian law, allelic information is stably inherited from parent to progeny (Mendel, 1866; Druery and Bateson, 1901). This tenet represents a cornerstone of genetics which, since its inception nearly 150 years ago, has proven to be universally applicable to practically all of life on earth. This is a testament to the brilliance of Mendel's work, and also to the basic similarities shared by all organisms at the molecular level. New technological advances have allowed us to continuously expand our studies of genetics, with this improved resolution, it has become clear that some exceptional mechanisms exist which require a broader view of inheritance.

In 2005, Lolle *et al.* reported that *Arabidopsis thaliana* plants homozygous for alleles of the recessive *hothead (hth)* organ fusion mutation segregate wild-type progeny at high frequencies (Lolle *et al.*, 2005a; **Figure 1.1**). These phenotypically wild-type 'revertants' were shown to be heterozygous with respect to the *HTH* gene (*HTH/hth*). Even more remarkably, *HTH/HTH* embryos were recovered from selfed *hth/hth* plants at low frequencies, and *hth/hth* plants were shown to be a source of wild-type pollen (Lolle *et al.*, 2005a). All of the eleven mutant *hth* alleles

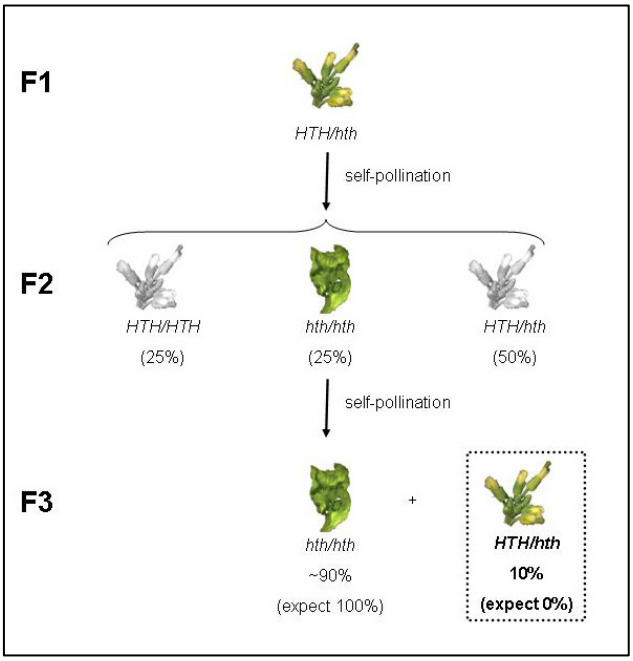


Figure 1.1: Non-mendelian segregation of *hth* allelic information in *A. thaliana*. The F3 progeny can inherit allelic information not detectable in the parental (F2) genome but present in the genome of an ancestor, such as the grandparent (F1).

that harbor G to A transition point mutations (*hth-1* through *hth-11*; Lolle *et al.*, 1998), were reported to show genetic instability (Lolle *et al.*, 2005a).

Arabidopsis plants homozygous for *hth* exhibit a characteristic organ fusion phenotype due to aberrant epidermal cell interactions (Krolikowski *et al.*, 2003; **Figure 1.2**). Although the biochemical function of the Hth protein has not been demonstrated directly, phylogenetic analyses suggest that *HTH* shares similarity to mandelonitrile lyase (Krolikowski *et al.*, 2003). Alternatively, results obtained from the metabolic profiling of *hth*-12 mutants lead Kurdyukov *et al.* (2006) to hypothesize that Hth is involved in the synthesis of long-chain fatty acids that are required for ensuring the integrity of the epidermal cuticle. Our laboratory is currently conducting biochemical assays with recombinant Hth proteins to test the hypothesis that Hth functions as a mandelonitrile lyase.

Figure 1.2: Inflorescences of wild-type (*HTH/HTH*) and mutant (*hth-8/hth-8*) *Arabidopsis* plants. The *hth* fusion phenotype is characterized by tightly fused floral organs with exerted stigmas, limited self-pollination, and pollen hydration on non-reproductive surfaces (Lolle *et al.,* 1998).

Genetic instability is not unique to the *HTH* locus but occurs on a genome-wide scale (Lolle et al, 2005a). In the past few years, ongoing research our laboratory has confirmed and expanded upon these original findings and shown that instability occurs at loci with insertions and deletions. These more recent studies have demonstrated that insertions and deletions (indels) ranging in size from 45-94 bp revert at a frequency similar to that observed for *HTH* alleles harboring single point mutations (**Figure 1.3**). These events seem to occur irrespective of genome location and often independent of changes at the *HTH* locus (Hopkins *et al.*, unpublished results).

Since no currently known genetic processes can adequately explain these observations, it is likely that the underlying mechanism is novel. We have coined the term *restoration* to describe the re-appearance of sequences that were not detectable in the parental genome but were present in an ancestral genome (such as the grandparent).

The single nucleotide restoration events reported by Lolle *et al.* (2005a) were shown to reinstate the same nucleotide that was present in the grandparent, suggesting that the mechanism is not the result of random mutation. Furthermore, the restoration of 45-94 bp insertions provides compelling evidence that restoration is template-directed. Since a template source has not been detectable by DNA-based methods, Lolle *et al.* (2005a) proposed the existence of a stable and inheritable RNA "cache" of extra-genomic information that had been sequestered in earlier generations.

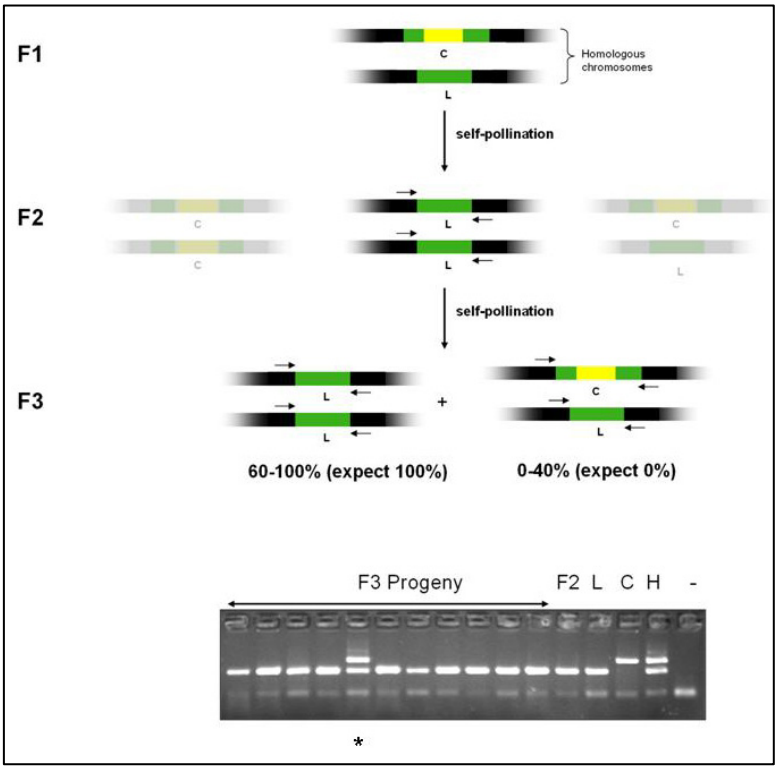


Figure 1.3: Genetic instability at 50-100 bp L*er* (*L*)/Col (*C*) indel polymorphisms. Previous work in our laboratory has demonstrated that F2 parents homozygous for a deletion polymorphism can give rise to F3 progeny harboring a restored insertion that is identical in size to that which was present in the F1 hybrid. Arrows represent allele-specific PCR primers. As shown in the electrophoretogram, indel polymorphisms are genotyped using a size-based PCR assay. By comparing the migration of the PCR products to that of a positive control, the presence or absence of an insertion can be determined. The asterisk indicates an F3 individual that is heterozygous for a restored insertion. Deletions have also been observed to restore (not shown).

Over the last decade and, increasingly in recent years, a number of remarkable reports have surfaced which suggest that RNA can function as a heritable source of genetic variation in a variety of organisms (Fire *et al.*, 1998; Rassoulzadegan *et al.*, 2006; Nowacki *et al.*, 2008), thus providing some indirect support for the RNA cache hypothesis. Early RNA interference work in *Caenorhabditis elegans* reported that gene silencing resulting from the introduction of double-stranded RNA could persist for several generations (Fire *et al*, 1998). The zygotic transfer of RNA molecules has been reported in mice (Rassoulzadegan *et al.*, 2006) and RNA templates have been demonstrated to guide genome rearrangements in the ciliate *Oxytricha trifallix* (Nowacki *et al*, 2008).

A number of alternative hypotheses that do not invoke a role for RNA have been put forth to explain the aberrant inheritance observed in *hth* mutants. These include a gene conversion-based model (Ray, 2005), a cryptic DNA cache model (Chaudhury, 2005), a toxic metabolite induced mutation-selection model (Comai and Cartwright, 2005) and a model based on chimerism (Krishnaswamy and Peterson, 2007). While some of these models build on well-defined genetic processes, the specific mechanisms have not been demonstrated and therefore the debate remains open. The most parsimonious and trivial hypothesis that wild-type pollen provided the *HTH* allele through outcrossing has also been proposed and tested (Peng *et al.*, 2006; Mercier *et al.*, 2008). Nonetheless, outcrossing alone cannot wholly explain a number of the observations reported in *hth* mutants (Lolle *et al.*, 2005a; Lolle *et al.*, 2006).

While the mechanism of restoration remains elusive, it would seem that the intrinsic capacity to rapidly confer genome-wide genetic variation would be of extreme adaptive utility, particularly to an inbreeding plant such as *Arabidopsis*. After long periods of inbreeding genomes of individuals should, in theory, approach complete homozygosity at all loci such that subsequent offspring are genetically identical to the parent, effectively nullifying the benefit afforded by the costly ability to reproduce sexually (that is, the maintenance of genetic diversity). This seems contradictory to the widely held belief that, when faced with the intimidating combination of sessility and food chain baseness, the best defense a

plant can have against diverse environmental insults is an equally diverse pallet of genetic resources (Linhart and Grant, 1996).

Interestingly, there exist a number of enigmatic and mechanistically undefined reports of non-mendelian inheritance in plants, in addition to the *hth* phenomenon (Roth *et al.*, 1989; Schneeberger *et al.*, 1991; Xu *et al.*, 2007). Cultivar-specific RFLP markers have been shown to spontaneously appear in tissue cultured soybean lines that were derived from cultivars that lacked the RFLP (Roth *et al*, 1989). It was hypothesized that the appearance of congruent markers across inbred cultivars could be the result of an undefined DNA rearrangement mechanism which can introduce genetic variation into inbred genomes as a response to environmental stress (Roth *et al.*, 1989). In flax (*Linum* sp.), several decades of work has revealed that environmental factors can induce stable genome changes in a single generation (Durrant, 1962; Evans *et al.*, 1966; Durrant and Jones, 1971; Cullis, 1973; Cullis, 1981; Schneeberger and Cullis, 1991; Cullis *et al.*, 1999). A variety of techniques have been applied to identify genomic variation in flax lines including quantitative DNA staining (Evans *et al.*, 1966, Durrant and Jones 1971) and Random Amplified Polymorphic DNA (RAPD) analysis (Cullis *et al.*, 1999). One particularly well studied complex insertion polymorphism has been reported to spontaneously appear in a number of flax lines and, interestingly, this same insertion has also been identified in a number of other flax and linseed varieties (Schneeberger and Cullis, 1992; Chen *et al.*, 2009). In rice (*Oryza sativa* L.) a similarly intriguing but enigmatic phenomenon has recently been described (Xu *et al.*, 2007). In this study, hybrids between phenotypically distinct diploid ♂ and triploid ♀ cultivars gave rise to viable diploid F1 progeny that were phenotypically uniform and homozygous for the paternal allele at 48/56 of the examined simple sequence repeat (SSR) markers, with nearly complete loss of heterozygosity manifested after only 8 generations of selfing (Xu *et al.*, 2007). When this aptly named Early Generation Stabilized Rice (EGSR) was reciprocally crossed with another inbred variety, Red-1, it was observed that the rapid loss of heterozygosity exhibited by EGSR was heritable, with 5 – 42% SSR homozygosity frequencies observed

in the F1 generation, depending on the F1 individual and direction of the parental cross (Xu *et al.*, 2007). Furthermore, since Red-1 exhibited 3 dominant traits (glume, auricle, and seed color) the F1 progeny were expected to exhibit these traits but, strikingly, the traits segregated. The authors attributed this phenotypic result to a loss of gene heterozygosity, analogous to loss of heterozygosity at the SSR alleles (Xu *et al.*, 2007), although this predicted genotypic change was not directly demonstrated.

Collectively, these studies suggest the existence of mechanisms capable of exerting pleiotropic effects upon the genome, often introducing novel genetic variants as a consequence. A feature common among these aforementioned studies is stress, whether an environmental stress, such as tissue culture, or a 'genome stress' such as diploid/triploid hybridization. The idea that mechanisms may exist that promote variation through genome reorganization is not a new one (McClintock, 1978).

Interestingly, enigmatic genetic phenomena like those described above are not limited to the plant kingdom. The cases of genetic reversion that most closely resemble the phenomenon of allelic restoration as reported by Lolle *et al.* (2005a) have arisen over the last two decades from studies of human recessive genetic disorders. Case studies focusing on three different human genetic disorders have uncovered genetic phenomena that remain poorly understood. These studies are examples of the fascinating and perplexing cases of 'natural gene therapy' reported in a variety of human genetic diseases (reviewed in detail in Hirschhorn, 2003). For example, the group of Eli Anne Kvittengen has shown that a fraction of patients affected by the autosomal recessive disease Tyrosinemia Type I exhibited mosaic patterns of fumarylacetoacetase (FAH) activity in their liver samples based on immunological screens (Kvittengen *et al.*, 1993; Kvittengen *et al.*, 1994). This was striking because the patients were all homozygous for FAH loss of function mutations. A molecular genetic analysis of a subset of patients exhibiting recovery of FAH activity demonstrated that these individuals were genetic mosaics and harbored a reverted wild-type allele on one homologue in some cell populations (Kvittengen *et al.*, 1994). Two years later, another study emerged reporting similar observations in a patient affected by adenosine deaminase (ADA) deficiency

(Hirschhorn *et al.,* 1996). ADA deficiency is an autosomal recessive disorder resulting from loss of function of the purine salvage enzyme ADA and in the absence of treatment is typically lethal prior to two years of age. Despite the fact that the patient in this study was identified early in life as having complete ADA deficiency, he gradually improved and was essentially of normal heath at age 12 (the time of publication). Detailed biochemical and molecular analyses revealed that the patient exhibited residual ADA activity resulting from the *in vivo* reversion to wild-type of the maternally inherited mutant *ADA* allele and that the patient was therefore a genetic mosaic. Similar cases reporting spontaneous genetic reversion in patients with Wiskott-Aldrich syndrome (WAS) have garnered enough attention to initiate the formation of an international consortium focused on WAS reversion (Stewart *et al.*, 2007). Clearly, genetic reversion is transcending its status as a rare anomaly towards being recognized as a tangible (albeit poorly understood) biological process with profound clinical implications.

Genetic variation within individuals can also arise via much less mysterious processes. The ubiquitous process of *de novo* somatic mutation leads to genetic variation within the individual, yet the heritability and therefore evolutionary significance of somatic mutation varies significantly between metazoans and plants. According to the widely accepted 'Weismann doctrine', somatic mutation is of no evolutionary significance since only gametic variation is heritable (Weismann, 1892). This idea still holds significance in most metazoans due to the fundamental separation between the germ line and the soma which is established in the early embryo (Buss, 1983). However, when considered in the context of plants and other clonal organisms, the relevance of this doctrine diminishes markedly. The modular and highly repetitive nature of plant development, the developmental plasticity of plant cells, and the ability to produce ramets by a variety of mechanisms (i.e. rhizomatry and layering) greatly increase the probability that a somatic mutation in any plant cell will propagate to give rise to a genetically distinct individual, and provides plants with the necessary redundancy to 'experiment' with new allelic variants. In fact, it has

been proposed that in plants, somatic mutations are of even greater evolutionary significance than mutations arising in the germ line (Whitham and Slobodchikoff, 1981).

When one considers the importance of heritable somatic mutation in shaping plant evolution, the concept of 'individual' takes on an entirely new meaning. A single plant can be considered an ecosystem unto itself, wherein an excess of individual units (i.e. branches) compete with one another for light and air to gain an upper hand in the final 'colony' of mature branches. This is highly analogous to the competition that can be observed in a recently cleared patch of forest, where an abundance of young saplings compete with each other to assert their dominance over what will become the mature canopy. And, like the saplings, the branches of a single plant have the capacity to evolve independently of one another, potentially contributing unique allelic variants to the next generation. This process of 'intra-organismal evolution' can be striking in long lived perennials which can be "mosaics of genetic variation" wherein one individual may have several branches of different genetic makeup (Whitham and Slobodchikoff, 1981). This concept is particularly well illustrated by 'bud sports' – a horticultural term used to describe phenotypically and genetically distinct branches which are thought to result from somatic mutations in the meristem. Bud sports have provided some striking examples of how somatic mutations can propagate beyond the individual, such as the example of the pink-fleshed grapefruit which was identified as one distinct branch on a tree in 1906 (Hartmann and Kester, 1975) and, via countless rounds of asexual division, has yielded all of the pink-fleshed grapefruits ever produced in the United States (Whitham and Slobodchikoff, 1981).

In *Arabidopsis*, the novel process we define as restoration appears to be a particularly striking, albeit poorly understood, example of an intrinsic generator of genetic variation. This thesis describes two projects which collectively sought to increase our descriptive and mechanistic understanding of this phenomenon. By characterizing previously identified restoration events at the DNA-sequence level, I have shown that sequences are reinstated coordinately across genomic intervals of revertant individuals,

and that these sequences are identical to those present in the grandparent. This supports the hypothesis of Lolle *et al.* (2005a) that restoration is a template-guided process and, furthermore, that the template appears to be of ancestral origin. Additionally, I have monitored a transgenic marker as well as native alleles segregating in a number of genotypically diverse ancestries both in the presence and absence of the *hth* mutation. My results indicate that these loci are stably inherited in the majority of lines. Complete restoration was not observed for any of these loci regardless of the presence of a *hth* allele. These projects are introduced in greater detail below.

1.2 Summary of the projects presented in this thesis

1.2.1 Project 1: Molecular characterization of loci with non-parental genotypes

Background

Two frequently used laboratory strains of *A. thaliana* are the highly inbred Landsberg *erecta* (L*er*, or *L*) and Columbia (Col, or *C*) accessions (ecotypes) which were originally collected in 1951 by Arabidopsis research pioneer Friedrich Laibach (Laibach, 1951). The Col-0 accession was the first plant genome to be completely sequenced and high-quality sequence data (approximately 1 error in 10,000 bp) is publicly available (*Arabidopsis* Genome Initiative, 2000). Within 2 years, the L*er* genome was sequenced using a low-coverage shotgun approach by Cereon Genomics (Jander *et al.*, 2002). Aligning the sequences of the Col and L*er* genomes has thus far revealed 56,670 polymorphisms of which 37,344 are single-nucleotide polymorphisms (SNPs; **Figure 1.4*a***) and 19,326

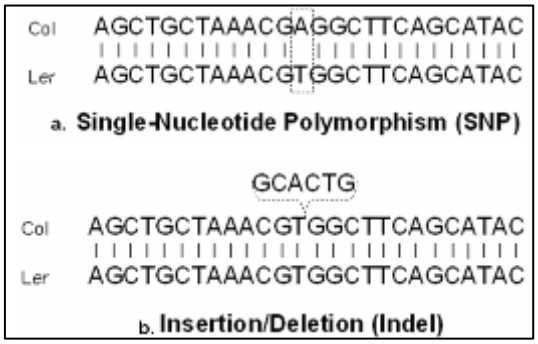


Figure 1.4: Examples of naturally occurring genomic sequence polymorphisms between the Col and L*er* accessions. **a.** SNPs occur when there is a single nucleotide mismatch in the alignments. **b.** A 6 bp indel polymorphism. Indels are characterized by a continuous gap in the alignment ≥1 bp in length.

are indel polymorphisms (**Figure 1.4*b***). These polymorphisms are catalogued in the publicly available Cereon *Arabidopsis* Polymorphism Collection (www.arabidopsis.org/cereon).

This natural genetic variation between accessions represents a valuable resource for generating co-dominant markers in L*er*/Col hybrids which can be used in genetics studies in *Arabidopsis*. Hybrid lines such as these are fundamental to the work described here. By using polymerase chain reaction (PCR)-based genotyping to profile a series of 16 genome-wide 50-100 bp indel polymorphisms in F2 and F3 L*er*/Col hybrids, F2 lines have been identified in our laboratory that segregate F3 individuals that exhibit non-parental genotypes. These lines provided the starting material for the experiments described in Chapter 2 of this thesis.

As was shown in **Figure 1.2**, we utilize a size-based PCR assay to score these indel polymorphisms and thereby determine the genotype. While this size-based assay is a simple and inexpensive approach for initially identifying changes in indel marker genotype profiles, it has certain limitations. When a non-parental indel is identified, only the relative size and genomic location can be ascertained. Three significant questions that cannot be answered from such PCR results are:

1. Do sequences flanking the indel marker change concurrently when indels revert? If so, what is the absolute extent of the restoration event?
2. What is the sequence of the non-parental loci? Do the sequences share identity with any known ancestral sequences?
3. Assuming that restoration events do extend beyond the individual indel markers interrogated, is restoration confined to one homologous chromosome (*cis* events) or are both homologues affected (*trans* events)? In other words, what is the *phase* of sequences detected in restoration events?

The work described in Chapter 2 of this thesis is a first step towards addressing these pertinent questions.

Assessing the genomic extent of restoration

Genotype profiling using 16 indel markers has revealed that changes in markers often occur independently of one another (S.J. Lolle, personal communication). Since the average distance separating the 16 indel markers is over 7 Mb, little is known about how smaller genomic intervals are affected by restoration. This study describes the profiling of flanking molecular markers spanning approximately 80-90 kb intervals of chromosome 1 and 5, each of which contained a previously identified non-parental indel. Determining if flanking markers change concurrently with the indel restoration event can provide some insight into the physical size of intervals affected by the events. This provides a 'higher resolution' image of the genomic region containing the restoration event, which may also provide some important insight into the size and nature of the template.

Isolation and DNA sequencing of non-parental loci

Size-based PCR genotyping of indel markers is not informative with respect to the sequence composition of a given restoration event. To determine the DNA sequence of these novel indel markers, a subset of the non-parental markers identified in various lineages were subjected to cloning and DNA sequencing. Cloning and sequencing 200-800 bp homologous genomic intervals from heterozygous revertants reveals the exact sequence of *cis* chromosomal regions harboring reversion events. The data presented in this thesis suggest that reversion events predominantly occur in *cis*, may occur coordinately over at least 80-90 kb regions, and reinstate sequences identical to those which were present in a recent ancestor.

1.2.2 Project 2: Monitoring the segregation of native and transgenic alleles in the presence and absence of *hth*

Background

Prior to this work, all loci assayed were sequences native to the *Arabidopsis* genome. The primary goal of the work presented in Chapter 3 of this thesis was to monitor the inheritance of a transgenic marker (*PAP1-D*, described below) and a native allele (*HTH*) segregating in both the presence and absence of *hth*. To test the segregation of *PAP1-D* and *HTH*, pedigrees were constructed in both the presence and absence of a mutant *hth* allele. The pedigrees consist of 5 'Tiers' with the collective goal of introducing the transgenic marker into a number of diverse genetic backgrounds via sexual crosses (Tiers 1 and 2), removing the marker by segregation (Tier 3 and 4), and monitoring the progeny for stable inheritance of the parental alleles (Tier 5). These pedigrees mimic the ancestry in which Lolle *et al.* (2005a) observed restoration of the *HTH* allele, and expand upon it by exploring how a number of different ancestries might affect segregation patterns. The ancestries are 'different' with respect to the presence or absence of *hth*, the zygosity of *hth* in the F1 (Tier 3) hybrid, and the direction in which the crosses were conducted in Tiers 1 and 2. Determining if these variables affect the stable inheritance of *HTH* and *PAP-D* was a primary goal of this project. A more detailed rationale of the design of these pedigrees is available in **Appendix A**. The results presented in this study contribute to our understanding of this atypical genetic phenomenon by providing information about target sequences which can become unstable, the putative role of *hth* in template generation and retrieval, and in revealing how different ancestries may influence genetic instability of both native and non-native sequences.

Selection of an appropriate transgenic marker: PAP1-D

When selecting an appropriate transgenic marker for the purpose of this project, several criteria had to be met: (1) The marker should have at least one distinct phenotype to facilitate monitoring its segregation

in a large population without resorting to tedious and error-prone molecular or cytological assays; (2) The marker must display complete dominance so that restoration of a single copy will produce a phenotype, as is the case with restoration of the *HTH* allele; (3) It should be stably inherited in the wild-type *Arabidopsis* genetic background and be present in a single copy; (4) Plants harboring the marker should be healthy and viable; (5) The presence of the transgene should not in any way affect the expression of *hothead* or its downstream metabolic processes since this could have an unforeseen effect on the restoration phenomenon; (6) The marker should contain sequences wholly foreign to the *Arabidopsis* genome.

With these considerations in mind, an appropriate transgenic line was selected from the *Arabidopsis* Biological Resource Center (http://www.biosci.ohio-state.edu/pcmb/Facilities/abrc/abrchome.htm). Selecting a previously generated transgenic seed stock saved several months of time over transfecting and stabilizing a *de novo* transgenic line.

The transgenic line selected for this project was seed stock number CS3884, a phenotypically distinct activation-tagged line generated in the Col-0 genetic background. This line was produced by screening several thousand mutants generated by random T-DNA insertion of an activation-tagging construct pSKI015 (Borevitz *et al,* 2000), the right border of which contains a constitutive 35S transcript promoter (35Se) from the cauliflower mosaic virus (**Figure 1.5**, Weigel *et al,* 2000). In this particular line, the T-DNA insertion lies approximately 4kb downstream of the *PAP1* locus (*P*roduction of *A*nthocyanin *P*igment), the product of which is a regulator of phenylpropanoid synthesis (Borevitz *et al,* 2000). The influence of the enhancer construct on the *PAP1* gene leads to its over-expression and the production of higher than normal quantities of anthocyanin pigments giving the plant an intense purple pigmentation as well as a hirsute phenotype due to increased trichome production. This completely dominant enhancer-trap allele has been named *PAP1-D* (Borevitz *et al.,* 2000), where '*D*' designates the dominant nature of the allele. The *PAP1-D* phenotype is visible in the first true leaves approximately one week following

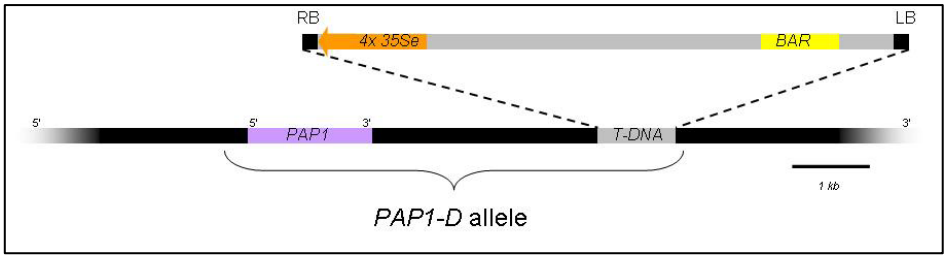


Figure 1.5: A schematic representation of the *PAP1-D* allele, resulting from the insertion of an activation tagging construct ~4kb downstream of the *PAP1* locus on chromosome 1. RB/LB: T-DNA left and right borders, respectively; 4x35Se: 4 tandem copies of the cauliflower mosaic virus 35S transcript promoter; *BAR*: gene conferring resistance to hygromycin (basta).

germination, and becomes increasingly distinct in the leaves and inflorescence as the plant matures (**Figure 1.6**).

The enhancer trap construct contains the *BAR* gene which confers resistance to the herbicide bialaphos (commercial name Basta). Bialaphos is a tripeptide containing phosphinothricin which when released upon proteolysis acts as a glutamate analogue, inhibiting glutamate synthesis and impairing nitrogen metabolism, effectively killing the plant (Thomspon *et al.,* 1987). *BAR* encodes a product which acetylates phosphinothricin, inactivating the herbicide (Thompson *et al.,* 1987). The *BAR* gene was originally isolated from the gram-positive bacterium *Streptomyces hygroscopicus* which produces bialaphos as a secondary metabolite (Thompson *et al.,* 1987).

Transgenic *PAP1-D* individuals can therefore be identified in three ways: (1) by the pigmentation phenotype, (2) by resistance to Basta, and (3) by PCR-based molecular genotyping using primers specific for the transgenic construct. The ease with which this transgenic construct can be identified made it conducive to the studies described here.

Molecular characterization and segregation analysis of the original *PAP1-D* line verified that it contains a single, simple insertion and that the mutant phenotype is determined by a single, dominant

allele (Borevitz *et al.,* 2000). Segregation analysis has also confirmed that the T-DNA insertion is tightly linked to the *PAP1* locus and that all individuals showing the purple phenotype are also Basta-resistant (Borevitz *et al.,* 2000). *PAP1* is located approximately 5 Mb proximal to the *HTH* locus on chromosome 1 (see **Figure 1.7**).

These preceding considerations, as well as the fact that the growth and fertility of *PAP1-D* plants is comparable to the wild-type Col-0 accession, made *PAP1-D* an ideal transgenic marker for the purposes of this project.

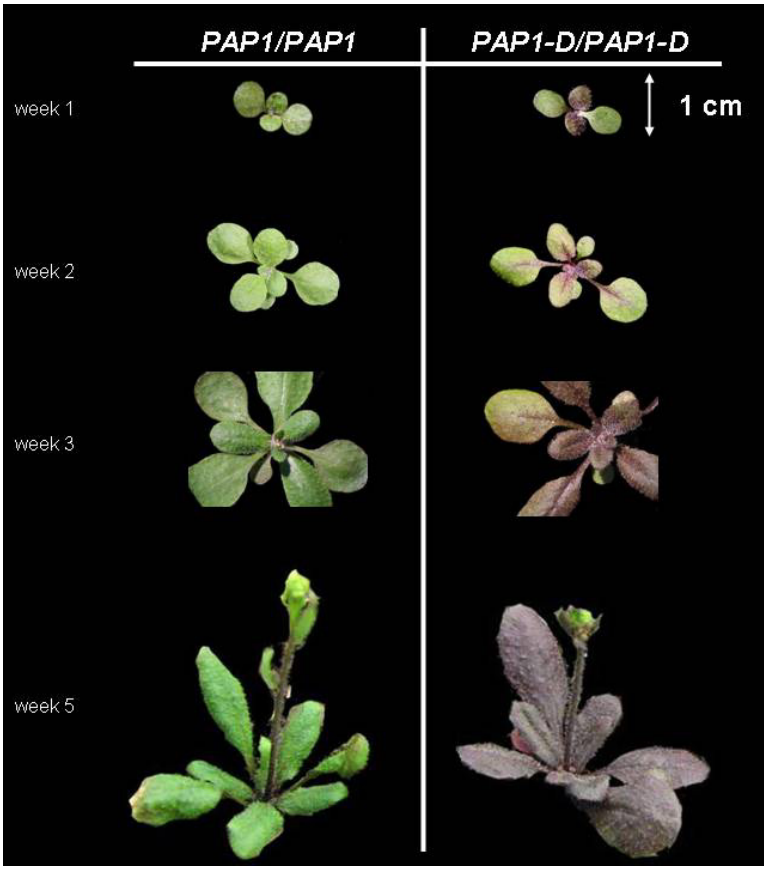


Figure 1.6: Individuals harboring the *PAP1-D* allele are easily distinguished from wild-type (*PAP1/PAP1*) individuals as soon as the first true leaves emerge.

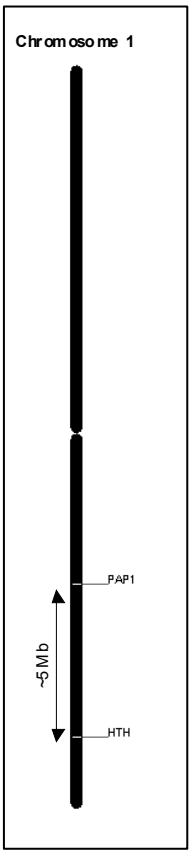


Figure 1.7: Proximity of the *PAP1* locus to *HTH* on chromosome 1.

Chapter 2
Molecular characterization of loci with non-parental genotypes

2.1 Materials and Methods

2.1.1 Plant material harboring non-parental loci

Three hybrid *Arabidopsis* lines segregating non-parental markers were identified in a parallel study aimed at determining the frequency of marker changes in the wild-type and various *hth* mutant backgrounds (Hopkins *et al.,* unpublished data). These lines, described below, provided the plant material used in this study.

194E9PCL10 hybrid line: hth-7 mutant background

In the process of screening for highly unstable *hth* lines an F2 line designated 194E9PCL10 was identified that segregated a high frequency of non-parental genotypes among the F3 progeny (Hopkins *et al.* unpublished data). To produce this line, F1 hybrids were generated by crossing *hth-7/hth-7* mutants in the Landsberg *erecta* (L*er*) background with the Columbia (Col) ecotype (L*er*♂ x Col♀), thus introducing thousands of co-dominant L*er*/Col polymorphisms into the F1 progeny. The resulting F1 hybrids were allowed to self pollinate and F2 seed was collected. The genotype profiles of F2 individuals were determined using PCR primers that amplified molecular markers polymorphic between L*er*/Col using the methods described below (section *2.1.4*). Select F2 fusion mutants (*hth-7/hth-7*) were allowed to self-pollinate to generate F3 seed. For the 194E9PCL10 line, F3 seed was collected from individual inflorescences (branches) and studied in cohorts (branches #1-20). The F3 progeny were either planted on soil or surface sterilized and plated on ½ strength Murashige and Skoog basal salt agar without sucrose (½ MS agar). Individuals were screened for the reappearance of non-parental indel markers using the PCR genotyping methods described in section *2.1.4*. The high frequency of individuals with non-parental

genotypes identified in the progeny of 194E9PCL10 prompted the selection of this lineage for a more detailed analysis of flanking markers (see section *2.1.7*). A subset of F3 individuals from this line were also selected to undergo DNA sequence analysis as described in section *2.1.6*.

PL113H12C2_11_12 hybrid line: hth-3 mutant background

The PL113H12C2_11_12 line was generated in the *hth-3* mutant background in the same manner as the 194E9PCL10 line described above, except that Col was used as the ♀ parent. The F3 progeny of PL113H12C2_11_12 were plated on ½ MS agar and the roots and shoots of approximately seven day old seedlings were bisected at the hypocotyls and collected separately. Roots and shoots were assayed for the reappearance of non-parental indel markers. A subset of individuals demonstrating non-parental genotypes in either the root or the shoot were selected for further analyses by DNA sequencing.

CL11B1 Hybrid Line: a wild-type HTH line

Ongoing work in our laboratory has demonstrated that, in the absence of the *hth* mutation, non-parental genotypes can be detected in a Col/L*er* hybrid lines at low frequencies (Hopkins *et al*, unpublished data). The line designated CL11B1 was generated in the same manner as the 194E9PCL10 line described above, except that the L*er* individual used in the F1 hybridization was *HTH/HTH*. The F3 seed obtained from this plant was surface sterilized and plated on ½ MS agar. Approximately 7 day old seedlings were bisected at the hypocotyls to separate the roots and shoots. Root and shoot samples were assayed independently for the reappearance of non-parental indel markers. Of the F3 progeny, 2/100 individuals exhibited distinct genotypes in the root. These two F3 progeny were selected for further analyses by DNA sequencing.

2.1.2 Plant growth conditions

Seeds were sown at a density of approximately 5-9 seeds per 36 cm^2 pot in moistened potting mix (4:3 mixture of LC1:LG3 Sungro Sunshine potting mixes, Sungro Horticulture, Seba Beach, AB) and vernalized at 4˚C for 2-5 days. Plants were maintained in growth chambers (Econoair AC60, Ecological Chambers Inc., Winnipeg, MB; GC8-VH/GCB-B, Environmental Growth Chambers, Chagrin Falls, Ohio; Conviron PGW36/E15, Controlled Environments Ltd., Winnipeg, MB) and illuminated with a mixture of incandescent and fluorescent lights (140 - 170µmol m^{-2} sec^{-1} at pot level) with a 24 h photoperiod . The environment in the growth chambers was maintained at 20 ± 1˚C and 40 – 60% relative humidity. Plants were watered as needed.

To control for the spurious introduction of exogenous pollen and thereby limit outcrossing events when generating F3 seed, the F2 lines PL113H12C2_11_12 and CL11B1, were maintained in isolation from potentially contaminating pollen sources. Following the onset of inflorescence elongation (bolting) each plant was grown in individual pots and surrounded by 55 cm tall rolled-up tubes of transparent plastic (**Figure 2.1**).

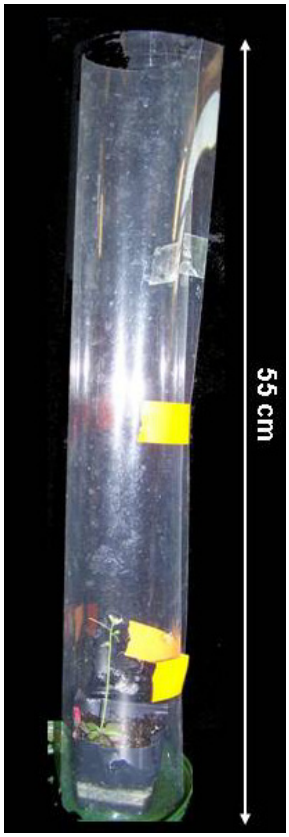


Figure 2.1: An example of the isolated growth conditions, where transparent plastic tubes were used to minimize the introduction of exogenous pollen.

F3 seeds were surface sterilized and grown on ½ MS agar using the following protocol: the seed was placed in sterile 15 mL blue-capped conical tubes and 10 mL of 15% bleach prepared in sterile water and supplemented with 1% Tween. The tubes were vortexed for approximately 10 seconds (s) and left lying on their sides for 10-20 min with gentle agitation. The seeds were allowed to settle to the bottom of the tubes and the bleach solution was poured off. The seeds were rinsed once in sterile water and resuspended in 10

mL 70% ethanol. The tubes were shaken continuously for approximately 30 s and the ethanol was poured off. The seeds were then rinsed five times with 10 mL sterile water. Thirty to fifty freshly sterilized seeds were immediately sown on the ½ MS agar, and vernalized at 4˚C for 2-5 days. The plates were oriented vertically in growth chambers to promote directional growth and seedlings were harvested approximately seven days following germination.

2.1.3 Tissue sampling and DNA extraction

For each DNA extraction, either a single cauline leaf or an approximately 9 mm^2 piece of rosette leaf tissue was removed with clean forceps and scissors and immediately placed in a prechilled 1.5 mL microcentrifuge tube and kept on ice. For seedlings grown on ½ MS, whole seedlings or root/shoot samples from bisected seedlings were used instead of rosette or cauline leaf tissue. Samples that were not processed immediately for DNA purification were stored at -20˚C. Additional tissue compromising 2-4 pieces of leaf tissue and 2-4 inflorescences was taken from F2 parents and stored at -20˚C..

DNA was purified from plant tissue samples according to the method of Edwards *et al.* (1991). Briefly, each tissue sample was ground in a microcentrifuge tube using a disposable plastic pestle. 400µL of extraction buffer (200mM Tris-HCl pH 7.5, 250mM NaCl, 25mM EDTA, 0.5% SDS) was added to the tubes and the samples were vortexed on high speed for approximately 5 s. Following centrifugation at 13 000 revolutions per minute (rpm) for 1 minute (min), 300 µL of the supernatant was transferred to a fresh tube and an equal volume of isopropanol added. Samples were vortexed and incubated at room temperature (RT) for 2 min to allow nucleic acid precipitation. Following a 5 min centrifugation at 13 000 rpm, the supernatant was discarded and tubes were inverted on a paper towel to air dry pellets for 15-20 min. The pellets were resuspended in 50-100 µL of TE buffer (10 mM Tris-HCl pH 8.0, 1 mM EDTA) and stored at 4˚C, -20˚C, or -80˚C until use.

2.1.4 Screening of co-dominant molecular markers

Segregation of alleles in the F3 progeny of the lines described above was monitored using size-based PCR genotyping assays of co-dominant indel polymorphisms present in our Col/L*er* hybrid lines. Briefly, 16 sets of PCR primers were designed to amplify 150-300 base pair (bp) genomic regions, each containing a single 45-94 bp indel marker that is polymorphic between the Col/L*er* genome sequences (**Figure 2.2a**). For the 16 markers, all L*er* alleles are deletions while all Col alleles are insertions. By size separating the PCR products on an agarose gel, an individuals' genotype at a particular indel marker can rapidly be determined (**Figure 2.2b**).

Of the 16 available indel markers, four (F15H11_35.66, MSA6_50.84, F8D20_68.14, and MNJ8_44.33) were chosen for further characterization based on their atypical segregation in the lines described above. The sequence of the primer sets used to initially screen these markers are provided in **Table 2.1**.

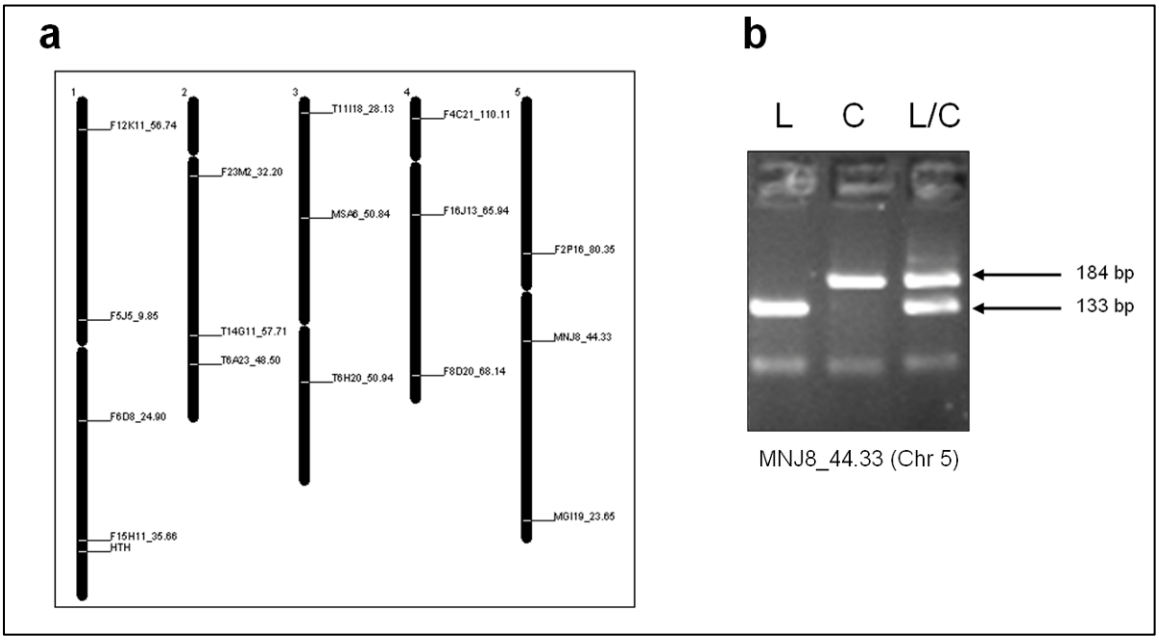


Figure 2.2: Genome-wide L*er*/Col indel polymorphisms. **a:** A haploid representation of the diploid *A. thaliana* genome showing the chromosomal location of the 16 indel markers used to initially screen the 194E9PCL10, PL113H12C2_11_12, and CL11B1 L*er*/Col hybrid lines. The indels are named according to their approximate location in kb on bacterial artificial chromosome (BAC) contigs (for example, the MNJ8_44.33 indel is located at approximately 44.33 kb on the 98.13 kb MNJ8 BAC. **b:** A sample electrophoretogram (4% agarose) showing the difference in migration of PCR products derived from individuals that are homozygous L*er* (*L* = deletion), homozygous Col (*C* = insertion), and heterozygous (*C/L*) at the MNJ8_44.33 indel (chromosome 5).

Table 2.1: PCR primer sets used to screen for reversion events at the four indel loci characterized in this study.

Primer Set	Sequence (5' → 3')	Target Indel	Product Size (bp)
F15H11-L/R	Fwd: ATTTGCGGCTGAAAGACAAG Rev: TGAGTGTGTCATGAGTGTTTGTTT	F15H11_35.66	229
MSA6-L/R	Fwd: CTGGGGTGTTCTCACAGGAT Rev: CGTTGGAGGTGGTCTTAGGT	MSA6_50.84	199
F8D20-L/R	Fwd: CACCAGACGGTGATGAAGAG Rev: CATTCGCGCATTTATTGTTG	F8D20_68.14	202
MNJ8-L/R	Fwd: CATGGATCAAAGATGATCTCCA Rev: TTCGCTTTTCGTGTTTCTGA	MNJ8_44.33	184

2.1.5 PCR and gel electrophoresis

For genotyping single nucleotide polymorphisms (SNPs), the derived cleaved amplified polymorphisms (dCAPS) PCR assay was used (Neff *et al.*, 1998). In this method, single nucleotide mismatches in PCR primers are used to introduce novel restriction enzyme recognition sites into the amplicons based on the presence of a given SNP in the DNA template. For the size-based PCR genotyping of co-dominant indel markers and dCAPS assays, amplification was conducted in NH buffer (10x NH Buffer: 200 mM Tris-HCl pH 8.8, 50 mM $MgCl_2$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$) using a Taq DNA polymerase that was purified from recombinant *E. coli* stocks. The PCR was conducted using the following temperature cycles: 1 cycle of 2 min at 95˚C, 30 s at 55˚C, 30 s at 72˚C, followed by 34 cycles of 30 s at 94˚C, 30 s at 55˚C, 30 s at 72 ˚C. For generating PCR products for cloning, amplification was conducted using GoTaq in GoTaq buffer (Promega, Madison, WI). The PCR was conducted using the following temperature cycles: 1 cycle of 2 min at 94˚C, 15 s at 55˚C, 30 s at 72˚C, followed by 39 cycles of 15 s at 94˚C, 15 s at 55˚C, 30 s at 72 ˚C. Custom DNA oligonucleotide primers were obtained from Sigma Aldrich (Mississauga, ON). Deoxynucleotide triphosphates (dNTPs) were purchased from Bio Basic Inc. (Markham, ON). Each 20 microlitre (µL) PCR reaction contained 1 µL of template DNA, 200 µM dNTPs, 0.5U Taq polymerase, 5 pmol of each primer, and 1x buffer. The dCAPS PCR products were digested directly with the appropriate restriction enzyme.

PCR products were size-separated by Tris-borate-EDTA buffered agarose gel electrophoresis. Gels were stained with 10µg/mL ethidium bromide and DNA bands visualized by ultraviolet illumination.

2.1.6 Isolation and sequencing of *cis* chromosomal regions

A subset of non-parental indel markers in the F3 progeny of the 194E9PCL10, PL113H12C2_11_12, and CL11B1 lines were chosen for further analysis by DNA sequencing. Primers flanking the loci of

interest were used to PCR amplify a 200-800 bp region containing the non-parental marker in both the F3 progeny and the F2 parent. The primers used are shown in **Table 2.2**.

Table 2.2: Primer sets used for PCR amplification and subsequent cloning of genomic sequences among the progeny of F2 lineages 194E9PCL10, PL113H12C2_11_12, and CL11B1. The primer sets used to clone the MSA6_50.84 and F8D20_68.14 indels were the same primer sets used for the initial screening of the restoration events.

Primer Set	Sequence (5' → 3')	Target Indel	Product Size (bp)
F15H11_IndMid-L/R	Fwd: CTCCACTAACTCCCGTTATTCC Rev: GAACAATCGGGCCACATATAG	F15H11_35.66	701
MSA6-L/R	Fwd: CTGGGGTGTTCTCACAGGAT Rev: CGTTGGAGGTGGTCTTAGGT	MSA6_50.84	199
F8D20-L/R	Fwd: CACCAGACGGTGATGAAGAG Rev: CATTCGCGCATTTATTGTTG	F8D20_68.14	202
MNJ8_IndMid-L/R	Fwd: TGCACTGACTAATCGATCTCC Rev: GACCCAACCGCTTAAACCTC	MNJ8_44.33	799

The individual amplicons were ligated directly into the pGEM T-Easy Vector System (Promega, Madison, WI) and the resulting recombinant plasmids were transformed into chemically competent *E. coli* DH5α cells.

Chemically competent cells were produced as follows: 5 mL of Luria Broth (LB) was inoculated with a single colony of *E. coli* DH5α and grown overnight at 37˚C with agitation. A 200 mL flask of LB was inoculated with 2.5 mL of the overnight culture and grown to an optical density at 600 nm (OD_{600}) of 0.4 – 0.6. The cultures were chilled on ice, aliquoted into sterile 50 mL centrifuge tubes, and centrifuged at 4,000 rpm for 10 min. The pellet was resuspended in 15 mL ice-cold 0.1 M $CaCl_2$ and chilled on ice for 15 min. Cells were harvested by centrifugation at 5,000 rpm for 1 min, resuspended in 0.1 M $CaCl_2$ supplemented with 15% glycerol, and left on ice overnight. The cells were divided into 200 µL aliquots and stored at -80˚C until use.

Bacterial transformants harboring plasmids with the PCR-generated clones were identified by blue/white colony screening on LB agar supplemented with 80 µg/mL 5-bromo-4-chloro-3-indolyl-β-D-

galactoside (X-Gal), 0.5 mM isopropyl β-D-1-thiogalactopyranoside (IPTG) and 100 μg/mL ampicillin. Bacterial transformants were genotyped for Col/L*er* indel sequences by touching the colony with a sterile 20 μL disposable pipette tip and rinsing the tip into 20 μL of PCR mastermix. The primers used for plasmid genotyping were the same as those used for the initial screening of individuals for the non-parental genotypes. This procedure is summarized in **Figure 2.3**.

Single colonies which contained plasmids harboring a fragment of interest were used to inoculate 5 mL of LB broth supplemented with 100 μg/mL ampicillin and grown overnight at 37˚C with agitation. Recombinant plasmids were purified from 1-2 mL of the stationary phase cultures using the GenElute Plasmid Miniprep Kit (Sigma Aldrich, Mississauga, ON). Plasmid inserts were sequenced in both directions using oligonucleotide primers flanking the multiple cloning site of the pGEM T-Easy vector (forward primer T7L 5'-TAA TAC GAC TCA CTA TAG GG-3'; reverse primer SP6R 5'-ATT TAG GTG ACA CTA TAG-3'). Sequences were manually inspected and edited using BioEdit (Ibis Biosciences, Carlsbad, CA) and alignments were constructed using CLC Workbench (CLC Bio, Aarhus, Denmark).

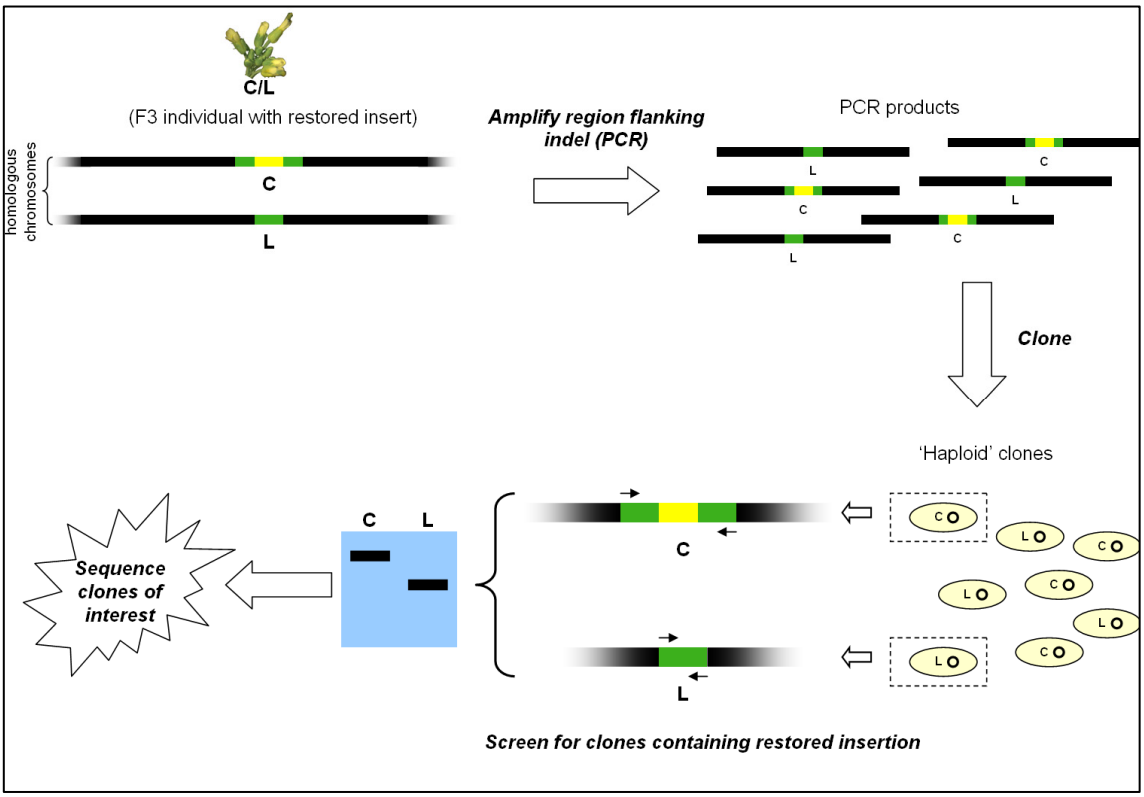


Figure 2.3: Schematic representation summarizing the procedure used to isolate *cis* chromosomal regions containing the indel marker from revertant individuals. In this diagram, an F3 individual was previously identified as being heterozygous for a restored insertion at a particular indel marker (revertant genotype *C/L*). PCR amplification of the region surrounding the indel marker yields *C* and *L* amplicons derived from both homologous chromosomes. To isolate the individual amplicons, the PCR products are cloned directly into a bacterial cloning vector. Each clone contains a single insert which can be genotyped as either *C* or *L* using the PCR primers that were initially used to identify the restoration event. Once genotyped, clones of interest can be sequenced.

2.1.7 Designing PCR-based markers to assay flanking loci

One hundred F3 progeny of 194E9PCL10 were selected for a more detailed analysis of markers flanking the unstable F15H11_35.66 (chromosome 1) and MNJ8_44.33 (chromosome 5) indel markers. Four co-dominant L*er*/Col polymorphisms flanking each of the indel markers were identified from the Cereon *Arabidopsis* Polymorphism Collection (www.arabidopsis.org/cereon) and exist as either SNPs or indel polymorphisms between the genome sequences of the L*er* and Col ecotypes. PCR primers were designed to differentiate genotypes using either size-based assay for indels or a dCAPS method for SNPs (as previously described in section *2.1.5*).

The primer sequences and approximate location of the flanking markers relative to F15H11_35.66 and MNJ8 44.33 are shown in **Figures 2.4** and **2.5** respectively.

a.

Primer Set	Sequence (5' › 3')	Approximate Location on F15H11 BAC (kb)	Enzyme (dCAPS)
F15H11_11.76_IND-L/R	Fwd: GAGTTGACCAAAAATGTGTTCC Rev: TGGGCCGAATGACTAAAAAC	11.76	n/a
F15H11_28.03_MspI-L/R	Fwd: TATTCCTCGGCCAAAAGATG Rev: TGCAGATAGAGAGGCAAAGG	28.03	MspI
F15H11_47.48_IND-L/R	Fwd: GCCGCAGCTGATAATAAAAAG Rev: GCAAGATAAAGTGATCAATGACG	47.48	n/a
F15H11_101.57_HindIII-L/R	Fwd: GCTATTACGTGGGGAGAAGC Rev: AATTCCCGGGATCCATACTG	101.57	HindIII

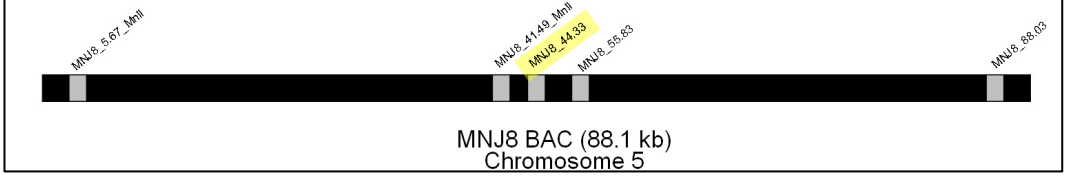
b.

Figure 2.4: Flanking markers on the F15H11 BAC. **a:** Primer sequences for genotyping indel (IND) and dCAPS polymorphisms flanking to the F15H11_35.66 indel marker. **b:** A diagrammatic representation (not to scale) of the proximity of the flanking markers relative to the F15H11_35.66 indel marker (highlighted).

a.

Primer Set	Sequence (5' › 3')	Approximate Location on MNJ8 BAC (kb)	Enzyme (dCAPS)
MNJ8_5.67_MnlI-L/R	Fwd: CAGACTGACCAGGAGTTCATTG Rev: CTGCAGCTCCAAAGAGAGTG	5.67	MnlI
MNJ8_41.49_MnlI-L/R	Fwd: TCGTTCTCAAGGACATCATACC Rev: GTGGATCAACGGGCTTATTC	41.49	MnlI
MNJ8_55.83_AluI-L/R	Fwd: AGCCTTTGGGATTGTAGATCA Rev: TGAGTCCACAGAGCCAACTAAA	55.83	AluI
MNJ8_88.03_IND-L/R	Fwd: GGAATCTTTATGACCTCAGCAG Rev: GAAAGAAGCTGTGCAAGTAGAGG	88.03	n/a

b.



MNJ8 BAC (88.1 kb)
Chromosome 5

Figure 2.5: Flanking markers on the MNJ8 BAC. **a:** Primer sequences for genotyping indel (IND) and dCAPS polymorphisms flanking the MNJ8_44.33 indel marker. **b:** A diagrammatic representation (not to scale) of the proximity of the flanking markers relative to the MNJ8_44.33 indel marker (highlighted).

2.2 Results

2.2.1 Assessment of changes in molecular markers flanking unstable indel markers in 100 F3 progeny of 194E9PCL10

The F2 parent 194E9PCL10 was determined to be homozygous L*er* (*L/L*) at both the F15H11_35.66 (chromosome 1) and MNJ8_44.33 (chromosome 5) indel markers. Nevertheless, this individual segregated F3 progeny that were heterozygous (*C/L*) at these loci (approximately 30% revertant progeny at F15H11_35.66 and 28% at MNJ8_44.33; Hopkins *et al.*, unpublished data). Of the 100 F3 progeny selected for further analysis at these loci, 29 were revertant at F15H11_35.66 and 33 were revertant at MNJ8_44.33. Seventy-one and 67 of the 100 F3 individuals had retained the parental genotype at F15H11_35.66 and MNJ8_44.33, respectively.

The genomic interval queried with the four flanking markers corresponds to approximately 89.82 kb on chromosome 1 (corresponding to BAC F15H11) and 82.36 kb on chromosome 5 (corresponding to BAC MNJ8). For all of the F3 progeny initially identified as being *C/L* at F15H11_35.66 and MNJ8_44.55, all of the flanking molecular markers tested were also *C/L* (29/29 and 33/33, respectively). Similarly, for all F3 individuals scored *L/L* (parental) at F15H11_35.66 and MNJ8_44.55, the flanking markers also retained the parental genotype (71/71 and 67/67, respectively). These results are summarized diagrammatically in **Figure 2.6** and the detailed results provided in **Appendix B**.

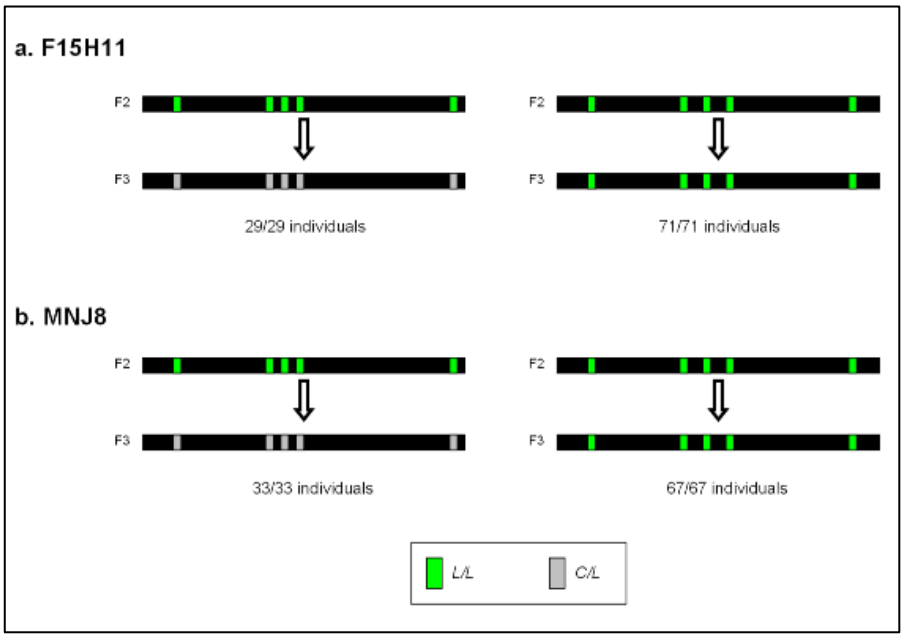


Figure 2.6: A summary of the results from the linked marker analysis of 100 F3 progeny of 194E9PCL10. **a:** 29/29 F3 individuals initially identified to have a heterozygous non-parental genotype at F15H11_35.66 (centre marker) also exhibited heterozygosity at the four flanking markers. The 71 individuals that were non-revertant at F15H11_35.66 exhibited the parental genotype at all of the flanking markers. **b:** Similar results were observed at the MNJ8 BAC

2.2.2 Isolation and DNA sequence analysis of regions containing non-parental indel markers

The following sections describe the results of DNA sequencing of a subset of cloned fragments derived from individuals showing non-parental indel genotypes. These reversion events were identified in the F3 progeny of 3 distinct Col/L*er* hybrid lines (194E9PCL10, PL113H12C2_11_12, and CL11B1). The results from each are presented separately below.

2.2.2.1 Molecular characterization of 194E9PCL10 progeny

Isolation and DNA sequence analysis of an unstable indel marker was conducted in each of two F3 individuals that descended from 194E9PCL10, designated F3#53 and F3#41. DNA flanking

F15H11_35.66 was PCR amplified from F3#53, cloned, and sequenced. DNA flanking MNJ8_44.33 was cloned from F3#41 and sequenced. The results are presented below.

Molecular characterization of the F15H11_35.66 locus from 194E9PCL10 F3#53

194E9PCL10 was *L/L* at F15H11_35.66, yet the seedling F3#53 was determined to be *C/L* at this marker. Following PCR amplification of a 701 bp region of F3#53 with the F15H11_IndMid-L/R primer set and subsequent cloning, 40 transformants were genotyped using the F15H11-L/R primer set. Of the 40 genotyped clones, 13 were determined to carry a single insert exhibiting the non-parental (*C*) allele and 24 carried a single insert exhibiting the parental (*L*) allele. The 3 remaining clones exhibited a heterozygous (*C/L*) genotype (see **Figure 2.7**). One clone containing the smaller putative *L* insert (F3#53CLN-9) and one containing the larger putative *C* insert (F3#53CLN-11) was subjected to DNA sequence analysis.

DNA encompassing the F15H11_35.66 locus was also amplified from the parent 194E9PCL10 using the same primer set. Ligation of PCR amplicons and subsequent screening of 10 of the resulting bacterial transformants with the F15H11-L/R primer set gave results consistent with all plasmid clones containing a single *L* allele insert. A single 194E9PCL10 recombinant plasmid containing an *L* fragment (F2PCLN-1) was subjected to DNA sequencing analysis.

The alignment of the obtained sequences with the corresponding region from the Col-0 reference genome is shown in **Figure 2.8** and is summarized diagrammatically in **Figure 2.9**. Alignment of this 498 bp sequence reveals a 76 bp insertion in clone F3#53CLN-11 which is identical to the insertion present in the Col-0 reference genome (i.e. that which was present in the F1 hybrid) with respect to length, sequence, and genomic location. Clone F3#53CLN-9 is identical in sequence to the parental clone F2PCLN-1, both of which lack the 76 bp insertion.

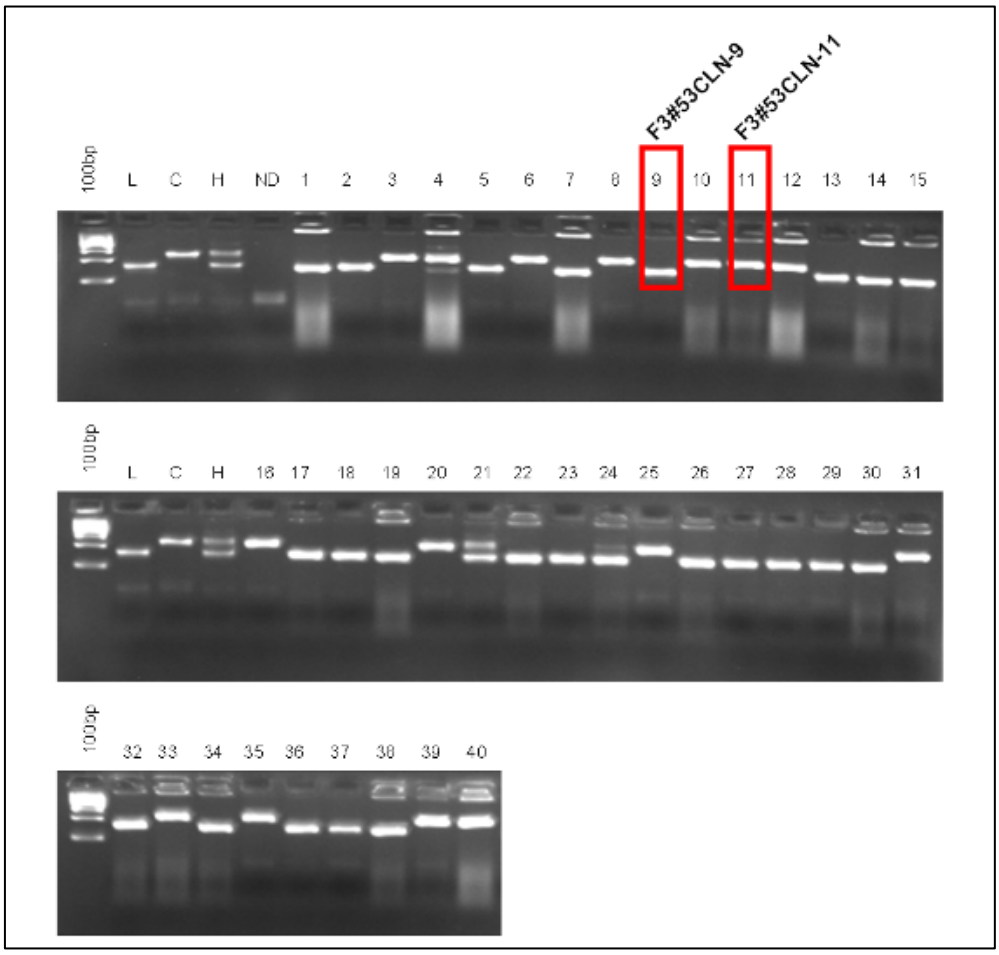


Figure 2.7: PCR genotyping F15H11_35.66 bacterial DNA samples cloned from the heterozygous *C/L* revertant F3#53. The smaller (153 bp) PCR products are consistent in size with the parental *L* allele. The large (229 bp) PCR products are consistent in size with the non-parental *C* allele. The clones F3#53CLN-9 and F3#53CLN-11 were selected for DNA sequencing analysis. 100 bp = 100 bp DNA ladder; L, C, H = L*er*, Col, Heterozygote (L*er*/Col) control DNA samples, respectively; ND = no DNA template (negative control). 4% agarose, F15H11-L/R primer set.

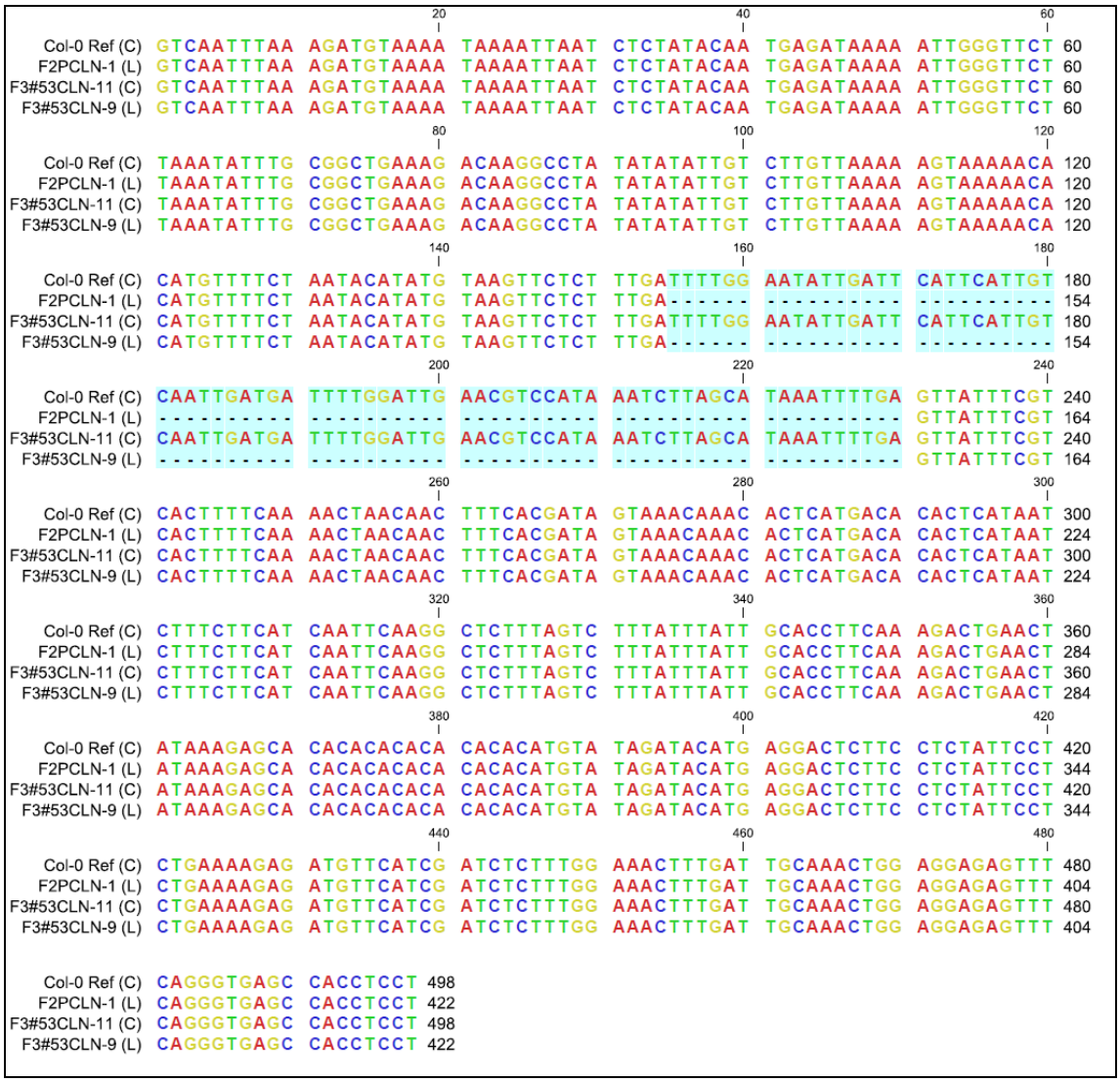


Figure 2.8: 194E9PCL10 F3#53 DNA sequence alignment showing a 498 bp alignment of the genomic region which contains the revertant F15H11_35.66 indel marker. The subjects are Col-0 Ref (Col-0 reference sequence = F1), F2PCLN-1 (sequence derived from an *L* clone isolated from the F2 parent 194E9PCL10), and homologous sequences cloned from the heterozygous revertant F3#53 (F3#53CLN-11 = revertant [*C*] homologue; F3#CLN-9 = non-revertant [*L*] homologue). The interval containing the F15H11_35.66 indel polymorphism (bases 155 – 230) is highlighted in blue. Ambiguous nucleotides were cropped from the ends of the sequences to produce the final alignment.

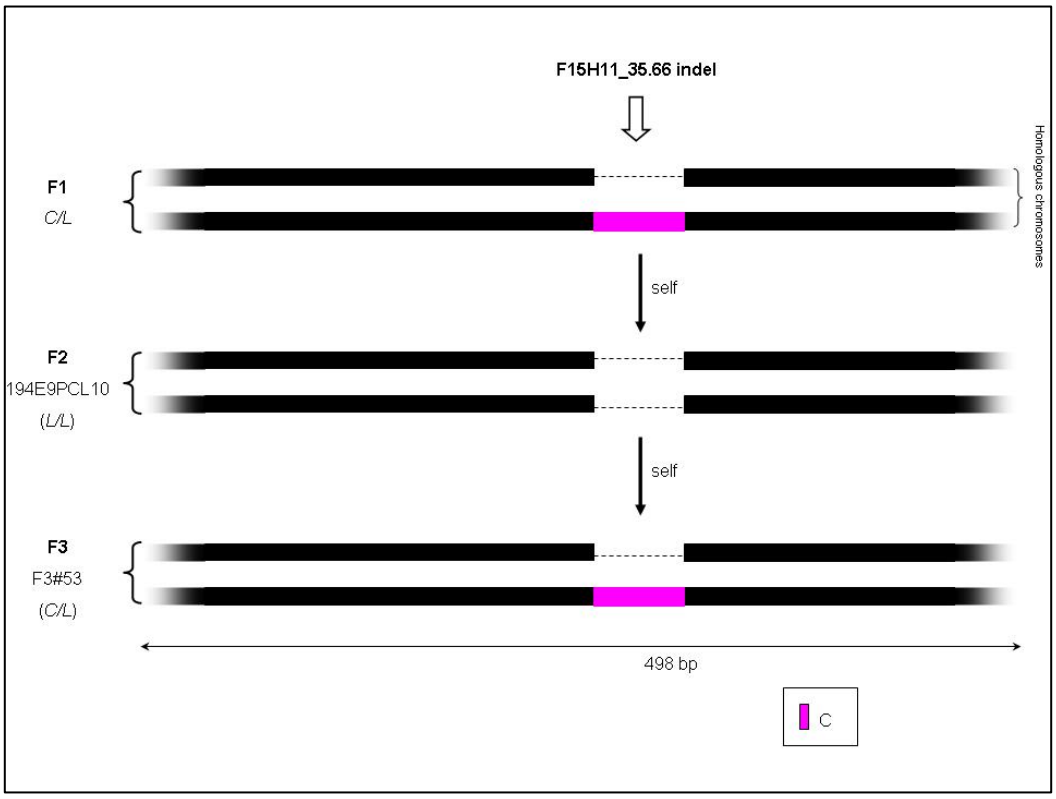


Figure 2.9: Diagrammatic summary of the alignment from **Figure 2.8** showing the presence of a 76 bp non-parental insertion at F15H11_35.66 on one of the homologous chromosomes of F3#53. The reinstated insertion is identical to the insertion which was present in the F1 hybrid.

Molecular characterization of the MNJ8_44.33 locus in 194E9PCL10 F3#41

F3#41 had phenotypically wild-type flowers whereas the parent, 194E9PCL10, exhibited floral fusion. Furthermore, the F3#41 was determined to be heterozygous (*C/L)* at the MNJ8_44.33 locus while the parent plant was homozygous *L/L*.

A 799 bp PCR product containing the MNJ8_44.33 indel was amplified from F3#41 with the MNJ8_IndMid-L/R primer set and cloned directly. Of the 14 bacterial transformants genotyped with the MNJ8-L/R primer sets, 13 were determined to carry single inserts exhibiting the parental (*L*) allele, while one clone carried a single insertion exhibiting the non-parental (*C*) allele. One representative clone of

each the *L* and *C* genotype (clones F3#41CLN-4 and F3#41CLN-14, respectively) was subjected to DNA sequencing analysis. DNA encompassing the MNJ8_44.33 indel was also cloned from the parent 194E9PCL10. All clones (8/8) exhibited the *L* allele. A single clone (F2PCLN-7) was selected for DNA sequence analysis.

The DNA sequences obtained from these isolates were aligned with the corresponding region of the Col-0 genome, as shown in **Figure 2.10** and summarized diagrammatically in **Figure 2.11**. The sequences shown in the alignment span 775 bp and reveal that, in addition to the 51 bp MNJ8_44.33 indel, 33 Col/L*er* polymorphisms not previously documented are present in the sequenced region. In total, six small indels (1-2 bp in length) and 27 SNPs were identified in the 775 bp sequence alignment. All DNA samples were sequenced in both directions and the 33 polymorphisms were verified in both sequencing runs.

The sequence alignment also reveals that the non-parental 51 bp *C* insertion at MNJ8_44.33 in F3#41 is identical in sequence to the insertion that is present in the Col-0 reference genome. Furthermore, the 33 newly identified polymorphisms also reverted to produce a phase that is identical to that which was present in the F1 hybrid.

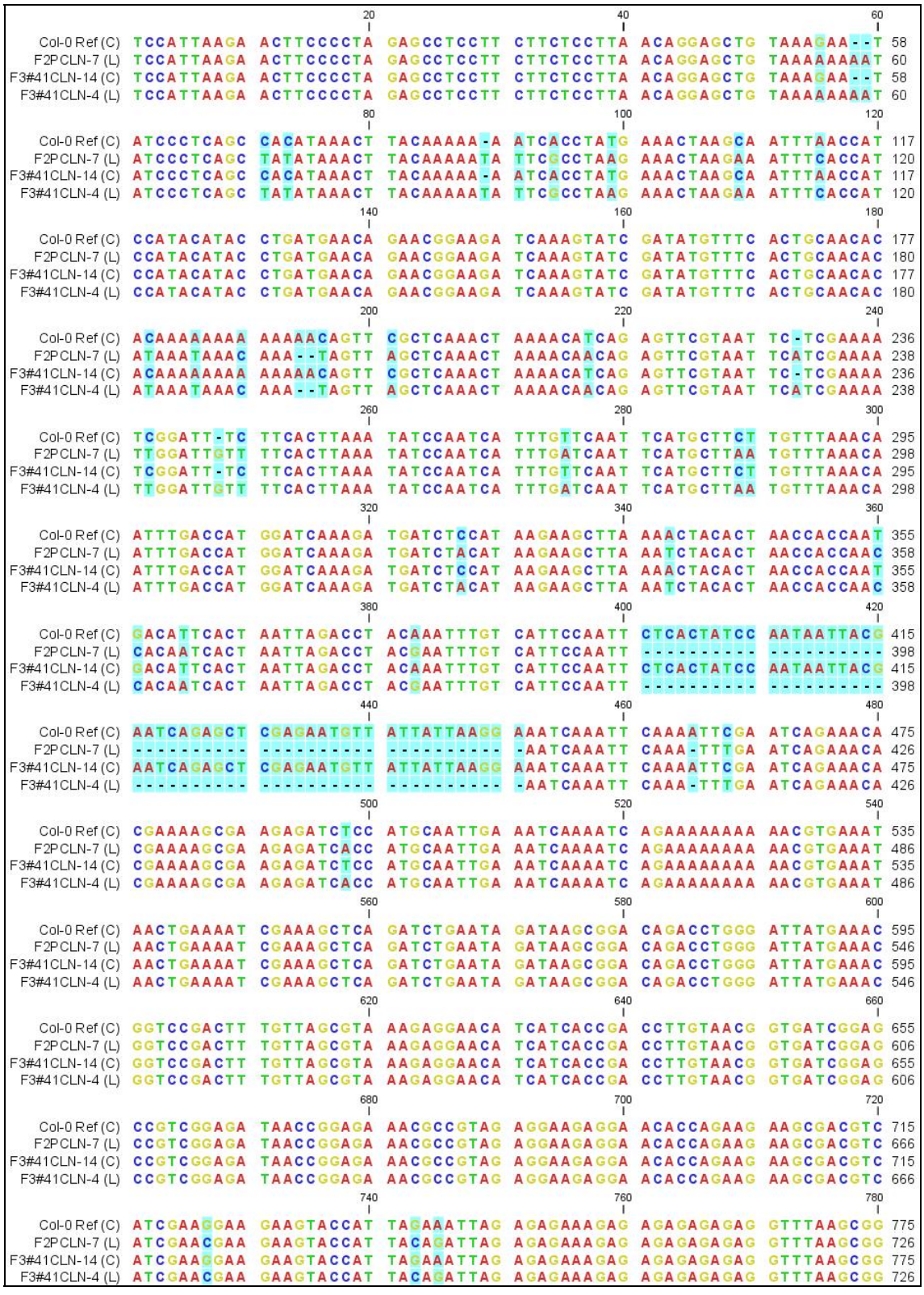

Col-0 Ref (C) TCCATTAAGA ACTTCCCCTA GAGCCTCCTT CTTCTCCTTA ACAGGAGCTG TAAAGAA--T 58
F2PCLN-7 (L) TCCATTAAGA ACTTCCCCTA GAGCCTCCTT CTTCTCCTTA ACAGGAGCTG TAAAAAAAAT 60
F3#41CLN-14 (C) TCCATTAAGA ACTTCCCCTA GAGCCTCCTT CTTCTCCTTA ACAGGAGCTG TAAAGAA--T 58
F3#41CLN-4 (L) TCCATTAAGA ACTTCCCCTA GAGCCTCCTT CTTCTCCTTA ACAGGAGCTG TAAAAAAAAT 60
Col-0 Ref (C) ATCCCTCAGC CACATAAACT TACAAAAA-A ATCACCTATG AAACTAAGCA ATTTAACCAT 117
F2PCLN-7 (L) ATCCCTCAGC TATATAAACT TACAAAAATA TTCGCCTAAG AAACTAAGAA ATTTCACCAT 120
F3#41CLN-14 (C) ATCCCTCAGC CACATAAACT TACAAAAA-A ATCACCTATG AAACTAAGCA ATTTAACCAT 117
F3#41CLN-4 (L) ATCCCTCAGC TATATAAACT TACAAAAATA TTCGCCTAAG AAACTAAGAA ATTTCACCAT 120
Col-0 Ref (C) CCATACATAC CTGATGAACA GAACGGAAGA TCAAAGTATC GATATGTTTC ACTGCAACAC 177
F2PCLN-7 (L) CCATACATAC CTGATGAACA GAACGGAAGA TCAAAGTATC GATATGTTTC ACTGCAACAC 180
F3#41CLN-14 (C) CCATACATAC CTGATGAACA GAACGGAAGA TCAAAGTATC GATATGTTTC ACTGCAACAC 177
F3#41CLN-4 (L) CCATACATAC CTGATGAACA GAACGGAAGA TCAAAGTATC GATATGTTTC ACTGCAACAC 180
Col-0 Ref (C) ACAAAAAAAA AAAAACAGTT CGCTCAAACT AAAACATCAG AGTTCGTAAT TC-TCGAAAA 236
F2PCLN-7 (L) ATAAATAAAC AAA--TAGTT AGCTCAAACT AAAACAACAG AGTTCGTAAT TCATCGAAAA 238
F3#41CLN-14 (C) ACAAAAAAAA AAAAACAGTT CGCTCAAACT AAAACATCAG AGTTCGTAAT TC-TCGAAAA 236
F3#41CLN-4 (L) ATAAATAAAC AAA--TAGTT AGCTCAAACT AAAACAACAG AGTTCGTAAT TCATCGAAAA 238
Col-0 Ref (C) TCGGATT-TC TTCACTTAAA TATCCAATCA TTTGTTCAAT TCATGCTTCT TGTTTAAACA 295
F2PCLN-7 (L) TTGGATTGTT TTCACTTAAA TATCCAATCA TTTGATCAAT TCATGCTTAA TGTTTAAACA 298
F3#41CLN-14 (C) TCGGATT-TC TTCACTTAAA TATCCAATCA TTTGTTCAAT TCATGCTTCT TGTTTAAACA 295
F3#41CLN-4 (L) TTGGATTGTT TTCACTTAAA TATCCAATCA TTTGATCAAT TCATGCTTAA TGTTTAAACA 298
Col-0 Ref (C) ATTTGACCAT GGATCAAAGA TGATCTCCAT AAGAAGCTTA AAACTACACT AACCACCAAT 355
F2PCLN-7 (L) ATTTGACCAT GGATCAAAGA TGATCTACAT AAGAAGCTTA AATCTACACT AACCACCAAC 358
F3#41CLN-14 (C) ATTTGACCAT GGATCAAAGA TGATCTCCAT AAGAAGCTTA AAACTACACT AACCACCAAT 355
F3#41CLN-4 (L) ATTTGACCAT GGATCAAAGA TGATCTACAT AAGAAGCTTA AATCTACACT AACCACCAAC 358
Col-0 Ref (C) GACATTCACT AATTAGACCT ACAAATTTGT CATTCCAATT CTCACTATCC AATAATTACG 415
F2PCLN-7 (L) CACAATCACT AATTAGACCT ACGAATTTGT CATTCCAATT ---------- ---------- 398
F3#41CLN-14 (C) GACATTCACT AATTAGACCT ACAAATTTGT CATTCCAATT CTCACTATCC AATAATTACG 415
F3#41CLN-4 (L) CACAATCACT AATTAGACCT ACGAATTTGT CATTCCAATT ---------- ---------- 398
Col-0 Ref (C) AATCAGAGCT CGAGAATGTT ATTATTAAGG AAATCAAATT CAAAATTCGA ATCAGAAACA 475
F2PCLN-7 (L) ---------- ---------- ---------- -AATCAAATT CAAA-TTTGA ATCAGAAACA 426
F3#41CLN-14 (C) AATCAGAGCT CGAGAATGTT ATTATTAAGG AAATCAAATT CAAAATTCGA ATCAGAAACA 475
F3#41CLN-4 (L) ---------- ---------- ---------- -AATCAAATT CAAA-TTTGA ATCAGAAACA 426
Col-0 Ref (C) CGAAAAGCGA AGAGATCTCC ATGCAATTGA AATCAAAATC AGAAAAAAAA AACGTGAAAT 535
F2PCLN-7 (L) CGAAAAGCGA AGAGATCACC ATGCAATTGA AATCAAAATC AGAAAAAAAA AACGTGAAAT 486
F3#41CLN-14 (C) CGAAAAGCGA AGAGATCTCC ATGCAATTGA AATCAAAATC AGAAAAAAAA AACGTGAAAT 535
F3#41CLN-4 (L) CGAAAAGCGA AGAGATCACC ATGCAATTGA AATCAAAATC AGAAAAAAAA AACGTGAAAT 486
Col-0 Ref (C) AACTGAAAAT CGAAAGCTCA GATCTGAATA GATAAGCGGA CAGACCTGGG ATTATGAAAC 595
F2PCLN-7 (L) AACTGAAAAT CGAAAGCTCA GATCTGAATA GATAAGCGGA CAGACCTGGG ATTATGAAAC 546
F3#41CLN-14 (C) AACTGAAAAT CGAAAGCTCA GATCTGAATA GATAAGCGGA CAGACCTGGG ATTATGAAAC 595
F3#41CLN-4 (L) AACTGAAAAT CGAAAGCTCA GATCTGAATA GATAAGCGGA CAGACCTGGG ATTATGAAAC 546
Col-0 Ref (C) GGTCCGACTT TGTTAGCGTA AAGAGGAACA TCATCACCGA CCTTGTAACG GTGATCGGAG 655
F2PCLN-7 (L) GGTCCGACTT TGTTAGCGTA AAGAGGAACA TCATCACCGA CCTTGTAACG GTGATCGGAG 606
F3#41CLN-14 (C) GGTCCGACTT TGTTAGCGTA AAGAGGAACA TCATCACCGA CCTTGTAACG GTGATCGGAG 655
F3#41CLN-4 (L) GGTCCGACTT TGTTAGCGTA AAGAGGAACA TCATCACCGA CCTTGTAACG GTGATCGGAG 606
Col-0 Ref (C) CCGTCGGAGA TAACCGGAGA AACGCCGTAG AGGAAGAGGA ACACCAGAAG AAGCGACGTC 715
F2PCLN-7 (L) CCGTCGGAGA TAACCGGAGA AACGCCGTAG AGGAAGAGGA ACACCAGAAG AAGCGACGTC 666
F3#41CLN-14 (C) CCGTCGGAGA TAACCGGAGA AACGCCGTAG AGGAAGAGGA ACACCAGAAG AAGCGACGTC 715
F3#41CLN-4 (L) CCGTCGGAGA TAACCGGAGA AACGCCGTAG AGGAAGAGGA ACACCAGAAG AAGCGACGTC 666
Col-0 Ref (C) ATCGAAGGAA GAAGTACCAT TAGAAATTAG AGAGAAAGAG AGAGAGAGAG GTTTAAGCGG 775
F2PCLN-7 (L) ATCGAACGAA GAAGTACCAT TACAGATTAG AGAGAAAGAG AGAGAGAGAG GTTTAAGCGG 726
F3#41CLN-14 (C) ATCGAAGGAA GAAGTACCAT TAGAAATTAG AGAGAAAGAG AGAGAGAGAG GTTTAAGCGG 775
F3#41CLN-4 (L) ATCGAACGAA GAAGTACCAT TACAGATTAG AGAGAAAGAG AGAGAGAGAG GTTTAAGCGG 726

Figure 2.10: 194E9PCL10 F3#41 DNA sequence alignment showing 775 bp of the genomic region containing the MNJ8_44.33 indel marker. The subjects are Col-0 Ref (Col-0 reference sequence = F1), F2PCLN-7 (sequence derived from an *L* clone isolated from the F2 parent 194E9PCL10), and homologous sequences cloned from the heterozygous revertant F3#41 (F3#41_CLN-14 = revertant [*C*] homologue; F3#41_CLN-4 = non-revertant [*L*] homologue). L*er*/Col polymorphisms are highlighted in blue. The 51 bp indel polymorphism at bases 400-451 is the MNJ8_44.33 indel marker. Ambiguous nucleotides at the beginning and end of the sequences were cropped to produce the final alignment.

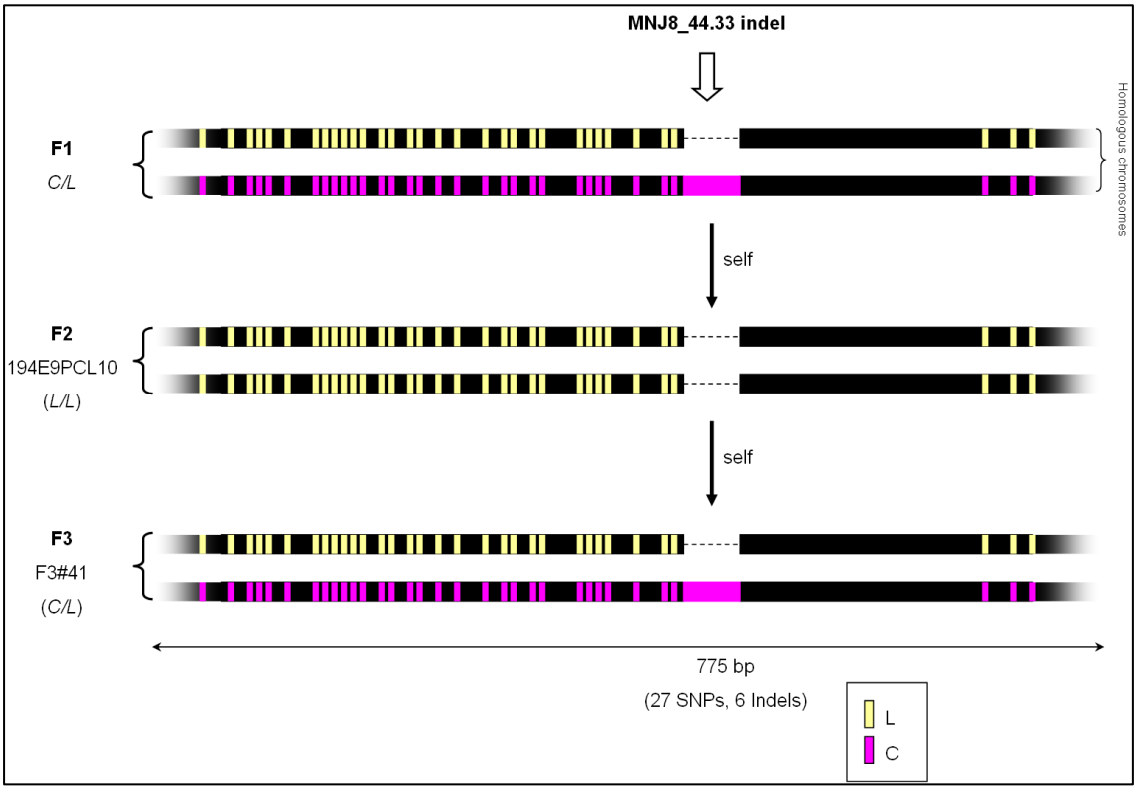


Figure 2.11: A diagrammatic summary of the 775 bp DNA sequence alignment from **Figure 2.10**, illustrating a *cis* reversion event at the indel marker MNJ8_44.33 in F3#41. The sequences are derived from the F2 parent 194E9PCL10 and the revertant progeny F3#41. Cloning and sequencing homologous chromosomal fragments from the heterozygous F3 revertant revealed that, in addition to the revertant 51 bp *C* insertion, 33 previously unidentified L*er*/Col polymorphisms had reverted, resulting in a 775 bp profile identical in phase and sequence composition to that of the F1 hybrid.

2.2.2.2 Molecular characterization of PL113H12C2_11_12 progeny

The roots and shoots of F3 seedlings were separately interrogated for the appearance of non-parental alleles. PL113H12C2_11_12 was *C/C* at the indel marker F8D20_68.14 (chromosome 4) yet segregated a high frequency of progeny (approximately 50%) that were heterozygous for this molecular marker (*C/L)*. Furthermore, a subset of the progeny (approximately 14%) where shown to have distinct genotypes at F8D20_68.14 in the root (R) or the shoot (S) (Hopkins *et al.*, unpublished data). DNA samples from two of these individuals, designated F3#13-R/S and F3#24-R/S (both with non-parental alleles localized in the root), were subjected to further analysis by cloning and DNA sequencing of the F8D20_68.14 indel locus.

For the non-revertant shoot samples F3#13-S and F3#24-S, seven and six clones were genotyped, respectively. In both cases, all clones harbored the larger fragment that was consistent in size with the *C* allele that was present in the parent. One clone from each sample was selected for DNA sequence analysis (F3#13S_CLN-2 and F3#24S_CLN-67).

For the revertant root samples F3#13-R and F3#24-R, 13 and 27 clones were genotyped, respectively. Among the 13 F3#13R clones, five exhibited a fragment consistent in size with the parental *C* allele, six exhibited a non-parental *L* allele, and two exhibited a heterozygous (*C/L*) genotype. One *L* and one *C* clone were selected for DNA sequence analysis (clones F3#13R_CLN-10 and F3#13R_CLN-17, respectively). Among the 27 F3#24-R clones, 23 were *L* and 4 were *C*. Again, one representative clone of each the *L* and *C* genotype was selected for DNA sequence analysis (clones F3#24R_CLN-42 and F3#24R_CLN-61, respectively).

DNA corresponding to the F8D20_68.14 marker was also isolated from the parent PL113H12C2_11_12, and all genotyped clones were confirmed to be *C* (8/8). A single bacterial clone (F2PL113H12C2_CLN-1) was selected for DNA sequence analysis.

The alignment of the DNA sequences obtained from F3#13-R/S with the corresponding region of the L*er* genome is shown in **Figure 2.12**. Identical sequencing results were obtained from the F3#24-R/S

(data not shown). These results are summarized diagrammatically in **Figure 2.13**. The DNA interval shown for F3#13 and F3#24 spans 180 bp. Both alignments reveal the presence of three previously undocumented SNPs which are polymorphic between the L*er* and Col sequences. In addition to the *L*-specific 84 bp deletion having been reinstated, three polymorphisms also reverted to the *L* allele configuration in the roots of F3#13 and F3#24, resulting in a profile that is identical in phase and sequence composition to the corresponding region of the F1 hybrid.

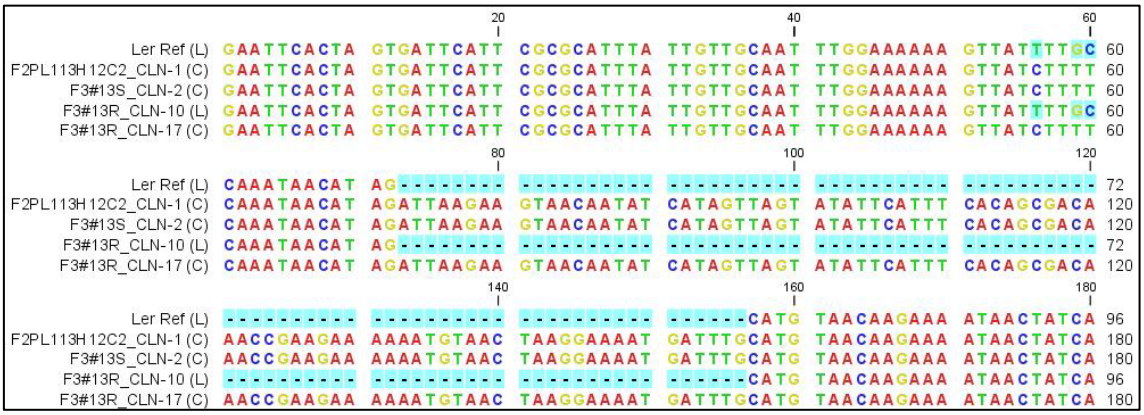


Figure 2.12: PL113H12C2_11_12 F3#13 DNA sequence alignments showing 180 bp of the genomic region containing the F8D20_68.14 indel marker. The subjects are L*er* Ref (sequence derived from a wild-type L*er* clone = F1), F2PL113H12C2_CLN-1 (sequence derived from a *C* clone from the F2 parent PL113H12C2_11_12), F3#13S_CLN-2 (sequence derived from a *C* clone from the shoot of F3#13), and homologous sequences cloned from the revertant root sample of F3#13 (F3#13R_CLN-10 = revertant [*L*] homologue; F3#13R_CLN-17 = non-revertant [*C*] homologue). L*er*/Col polymorphisms are highlighted in blue. The sequences were cropped at the beginning and end to remove ambiguous nucleotides and vector sequence. Identical results were observed in the corresponding clones from F3#24 R/S (not shown).

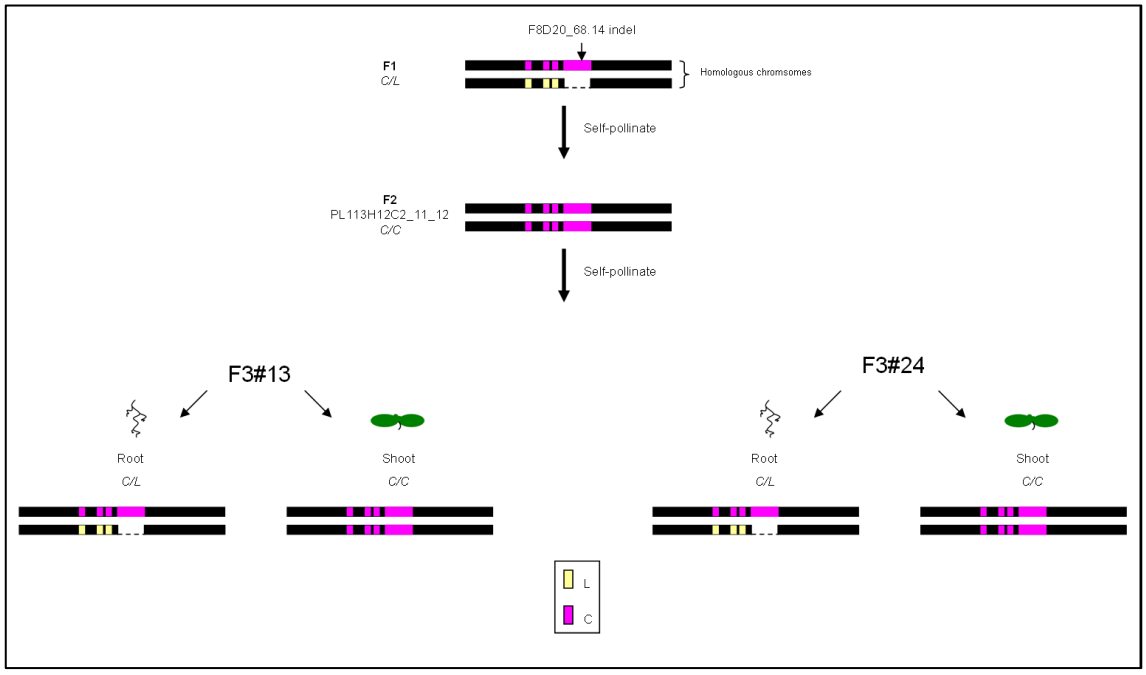


Figure 2.13: A diagrammatic summary of the results obtained from cloning and sequencing root-specific restoration events at F8D20_68.14 in two F3 progeny of PL113H12C2_11_12. In both F3#13 and F3#24, the reversion events in the root samples reinstated the *L* deletion and three *L* SNPs resulting in a profile which is identical to that of the F1 hybrid. The sequences illustrated here represent 180 bp.

2.2.2.3 Molecular characterization of CL11B1 progeny

Molecular genotyping of roots and shoots of F3 progeny descending from the wild-type (*HTH/HTH*) Col/L*er* hybrid F2 CL11B1 identified root- or shoot-specific genotypic changes in a subset of seven day old seedlings (Hopkins et al, unpublished data). CL11B1 was *L/L* at the indel marker MSA6_50.84 yet two of the F3 progeny exhibited a non-parental *C/L* genotype for this marker, localized to either the root or the shoot. These two individuals were the only revertants identified among the 100 progeny of CL11B1 that were screened. These progeny, designated F3#9 (reversion localized to the root) and F3#14 (reversion localized to the shoot) were selected for further analysis by DNA sequencing of the MSA6_50.84 indel locus in both the R and S samples.

Following amplification of a 199 bp PCR product containing the MSA6_50.84 indel and subsequent cloning, bacterial transformants were genotyped using the MSA6-L/R primer set. From each of the non-revertant samples F3#9-S and F3#14-R, 21 clones were genotyped. In both cases, 21/21 plasmid inserts exhibited the parental (*L*) genotype. A representative clone from each F3#9-S and F3#14-R was selected for DNA sequence analysis (clones F3#9S_CLN-2 and F3#14R_CLN-7, respectively).

For the revertant samples F3#9-R and F3#14-S, 21 and 20 clones were genotyped, respectively. Of the 21 F3#9-R clones, five were *L*, nine were *C*, and seven exhibited a heterozygous (*C/L*) genotype. One *L* clone and one *C* clone were selected for DNA sequence analysis (clones F3#9R_CLN-8 and F3#9R_CLN-17, respectively).

Of the 20 F3#14-S clones, twelve were *L*, three were *C*, and eight were *C/L*. Again, one *L* and one *C* clone were selected for DNA sequence analysis (clones F3#14S_CLN-20 and F3#14S_CLN-18, respectively).

The MSA6_50.84 indel was also cloned from the parent CL11B1. Genotyping 21 transformants revealed that all clones exhibited the *L* allele. A single clone designated F2CL11B1_CLN-18 was selected for DNA sequencing.

The alignments of the sequences obtained from DNA samples isolated from F3#9 and F3#14 are shown in **Figure 2.14** and **2.15**, respectively. These collective results are summarized diagrammatically in **Figure 2.16**. The sequence alignments of both F3#9 and F3#14 span a 198 bp DNA interval. The sequence data reveal that in the revertant samples F3#9-R and F3#14-S, the reinstated insertion is identical to the 54 bp insertion that was present in the F1 (Col-0 reference sequence) with respect to length, sequence, and genomic location. No additional Col/L*er* polymorphisms were identified within the sequenced regions.

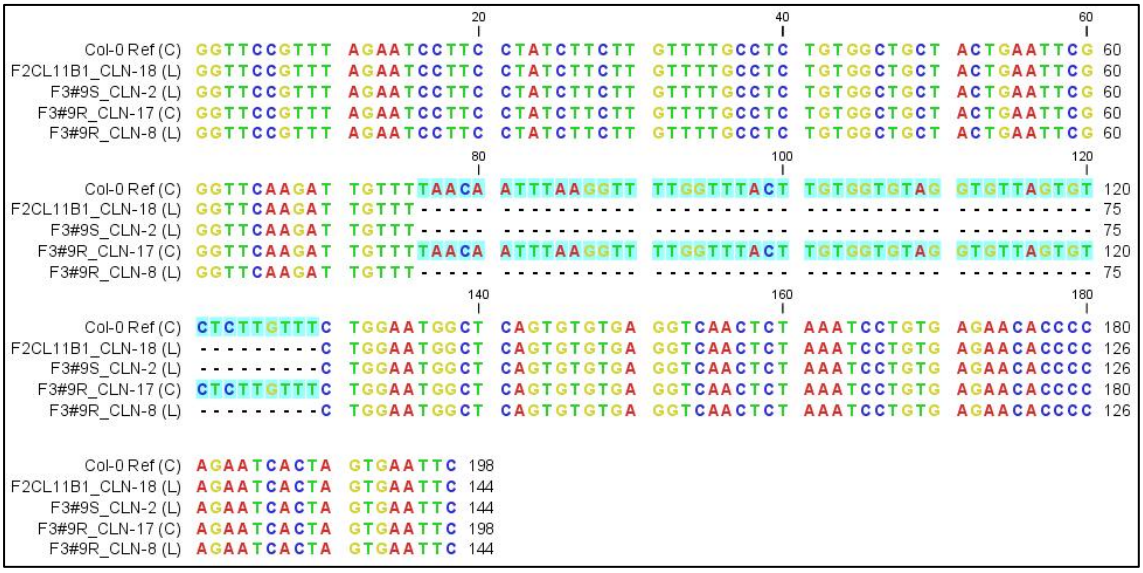


Figure 2.14: CL11B1 F3#9 DNA sequence alignments showing 198 bp of the genomic region containing the MSA6_50.84 indel marker. The subjects are Col Ref (Col-0 reference sequence = F1), F2CL11B1_CLN-18 (sequence derived from an *L* clone from the F2 parent PL113H12C2_11_12), F3#9S_CLN-2 (sequence derived from a *L* clone from the shoot of F3#9), and homologous sequences derived from the revertant root sample of F3#9 (F3#9R_CLN-17 = revertant [*C*] homologue; F3#9R_CLN-8 = non-revertant [*L*] homologue). The polymorphic 54 bp Col MSA6_50.84 insertion is highlighted in blue. The sequences were cropped at the beginning and end to remove ambiguous nucleotides and vector sequence.

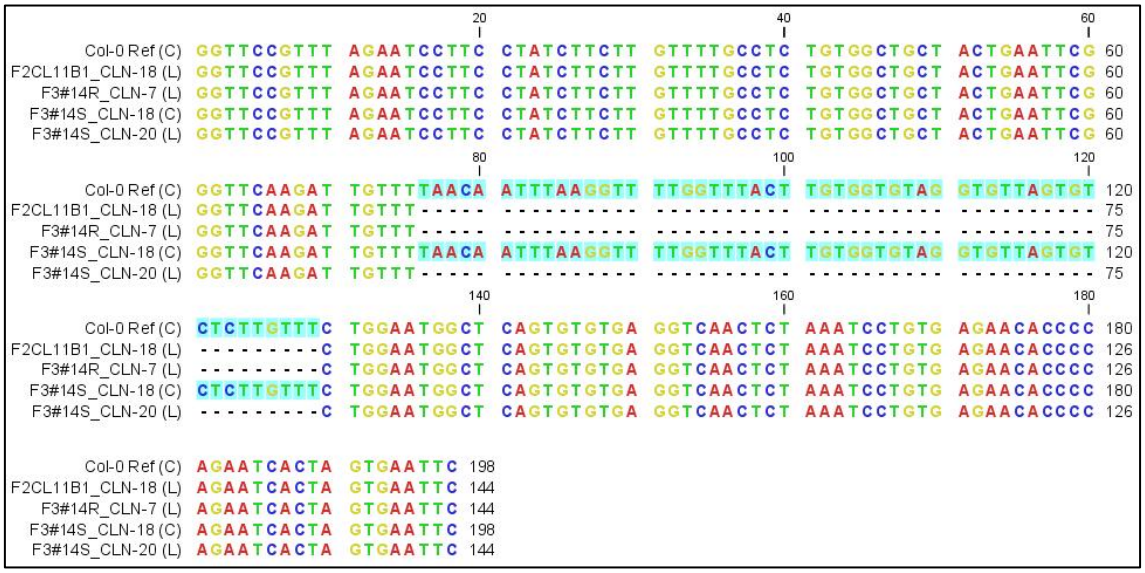


Figure 2.15: CL11B1 F3#14 DNA sequence alignments showing 198 bp of the genomic region containing the MSA6_50.84 indel marker. The subjects are Col Ref (Col-0 reference sequence = F1), F2CL11B1_CLN-18 (sequence derived from an *L* clone from the F2 parent PL113H12C2_11_12), F3#14R_CLN-7 (sequence derived from a *L* clone from the root of F3#14), and homologous sequences derived from the revertant shoot sample of F3#14 (F3#14S_CLN-18 = revertant [*C*] homologue; F3#14S_CLN-20 = non-revertant [*L*] homologue). The polymorphic 54 bp Col MSA6_50.84 insertion is highlighted in blue. The sequences were cropped at the beginning and end to remove ambiguous nucleotides and vector sequence.

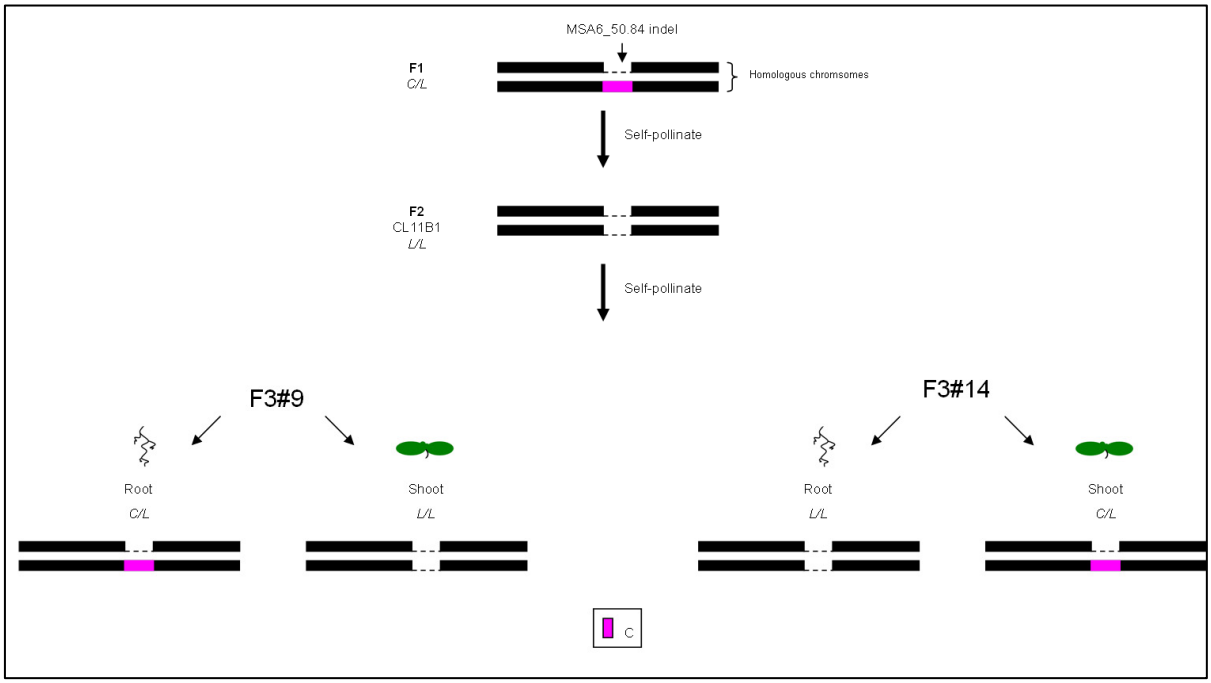


Figure 2.16: A diagrammatic summary of the results obtained from cloning and sequencing root/shoot-specific restoration events at MSA6_50.84 in two F3 progeny of the wild-type *HTH/HTH* hybrid CL11B1. In the revertant samples F3#9-R and F3#14-S, the 54 bp *C* insertion which was present in the F1 hybrid was exactly reinstated. The sequences illustrated here represent 198 bp in all samples.

Chapter 3
Monitoring the segregation of native and transgenic alleles in the presence and absence of *hth*

3.1 Materials and Methods

3.1.1 Plant growth conditions

Plants were grown as described in section *2.1.2.* To control for the spurious introduction of exogenous pollen, Tier 4 plants were grown in strict isolation from potentially contaminating *PAP1-D* and *HTH* pollen sources following the protocol described in section *2.1.2*.

3.1.2 Phenotypic analysis

Plants were scored for the presence of the *PAP1-D* allele 1 – 2 weeks following germination based on the presence of the purple hirsute phenotype which is evident in the first true leaves. *PAP1/PAP1* (green) individuals were scored based on the absence of the purple hirsute phenotype.

Plants were scored for the *erecta* mutant phenotype within 1 – 2 weeks of the onset of flowering. Individuals homozygous for a mutant *erecta* allele exhibit a short and highly ramified growth habit with compact inflorescences, blunt ended stout siliques and flowers with limited internode elongation. Individuals with the vigorous wild-type growth habit with highly elongated inflorescences and floral internodes were scored as *ER* (see **Figure 3.1**). *ER* individuals were not further distinguished as being *ER/ER* or *ER/er*.

Plants were scored for the *hth* mutant phenotype within 1 – 2 weeks of the onset of flowering since organ fusion in *hth* plants is generally limited to the floral organs (Lolle *et al.*, 1998). Plants were scored as fusion mutants (*hth/hth*) based on the presence of closed, tightly fused floral organs which adhered to one another.

PCR-based molecular genotyping was used to determine allele composition which could not be inferred from the phenotype (for example, distinguishing *HTH/HTH* from *HTH/hth*, or *PAP1-D/PAP1-D* from *PAP1-D/PAP1*) and, in many cases, to independently confirm that the genotype matched the phenotype.

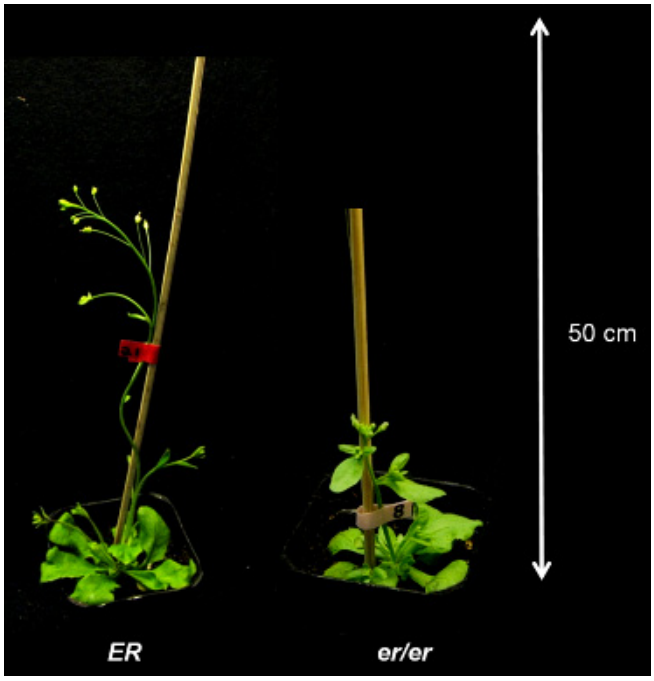


Figure 3.1: Individuals homozygous for *er* exhibit a distinct pleiotropic phenotype which is most readily identified as a compact growth habit and limited floral internode elongation.

3.1.3 Tissue sampling and DNA extraction

The tissue sampling and DNA extraction protocols used were described previously in section *2.1.3*.

3.1.4 PCR-based molecular genotyping

The zygosity of the *hth* alleles *hth-4* and *hth-8* was determined using a dCAPS assay described by Lolle *et al.* (2005a). In this way, all alleles (*hth-4, hth-8,* and *HTH*) could be distinguished. Two sets of dCAPS primers were available for genotyping each *hth* allele: one set which introduces a recognition site in the PCR product derived from the mutant sequence (leaving wild-type sequences uncleaved) and another set which introduces a restriction site into the PCR product derived from the wild-type sequence (leaving mutant sequences uncleaved). In some cases, both dCAPS assays were used on a given individual to independently confirm the presence or absence of a particular SNP and to reduce the possibility of scoring false negatives due to incomplete DNA cleavage or contaminating sequences. Information on the dCAPS primers and their corresponding restriction enzymes are provided in **Figure 3.2**. The PCR

reagents, reaction conditions, and electrophoresis methods used for dCAPS genotyping of the *hth* alleles were the same as those described previously for dCAPS amplification in section *2.1.5*.

a

dCAPS Primer Set	*hth* Allele	Sequence (5' › 3')	Restriction Enzyme	Comment
140C10a L/R	*hth-4*	Fwd: GAA GCT GGT GAA GGA GTC GT Rev: CTC CGC CGC GGT GTG TC	SalI	Cleaves WT
140C10b L/R	*hth-4*	Fwd: ATT TTC GAC AGA TTC GAA Rev: GGA ATT TCA TCG AAC ATT TGG	AclI	Cleaves mutant
208F658 L/R	*hth-8*	Fwd: TCA ATG GCT CAG TTC TGC AA Rev: TGC CCA ATA GAG GGT TAC ACA	BstNI	Cleaves WT
208F6m L/R	*hth-8*	Fwd: GTT CAA CGT TTG ACG AGT CTT Rev: AAA CAC CAG CTT TCT TTC CAA	MseI	Cleaves mutant

b

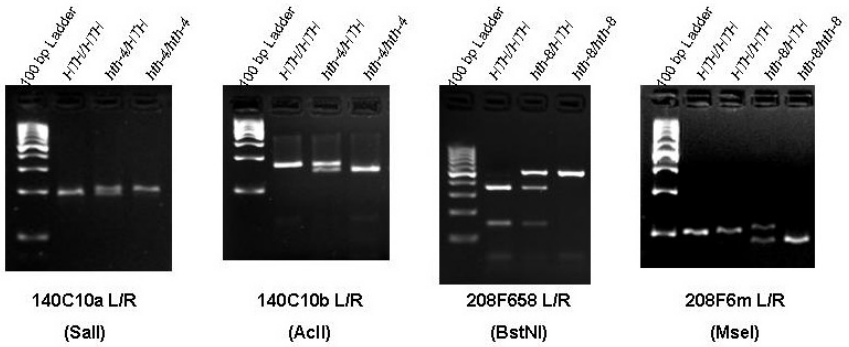


Figure 3.2: Genotyping *hth* alleles. **a**: A table summarizing relevant information pertaining to the dCAPS primer sets used to genotype individuals for *hth-4* and *hth-8*. **b:** Sample electrophoretograms (4% agarose) showing the separation of cleaved dCAPS PCR products. The size difference of dCAPS PCR products which have been subjected to cleavage is used differentiate the *hth* genotype as homozygous wild-type, heterozygous, and homozygous mutant.

To determine the zygosity of the *PAP1-D* allele in purple hirsute individuals, two separate PCR assays were used in combination. One assay used primers specific to the *BAR* gene which is present within the T-DNA insertion (forward primer BAR61L 5'-GTC TGC ACC ATC GTC GTC AAC C-3'; reverse primer BAR375R 5'-GAC AGC GAC CAC GCT CTT-3'). To distinguish heterozygotes, a second PCR assay utilized primers flanking the T-DNA insertion which, under the experimental conditions used, would only give a PCR product if the T-DNA insertion was absent (forward primer PAP1D4839L 5'-GAT TGG CTT

TGA TTG CTG GT-3'; reverse primer PAPD5258R 5'-GCT CTA ATG CTT GCT TAC GAA-3'). Based on the combined results of PCR amplification using the BAR61L/375R and PAP1D4839L/5258R primer sets, the zygosity of an individual with respect to *PAP1-D* could be determined (**Figure 3.3**). PCR with the BAR61L/375R primer set was conducted using GoTaq and GoTaq buffer (Promega, Madison, WI) using the reagents and conditions described in section *2.1.5*. PCR using the PAP1D4839L/5258R primer set was conducted using a Taq polymerase purified from recombinant *E. coli* stocks using reagents and conditions that were also described in section *2.1.5*.

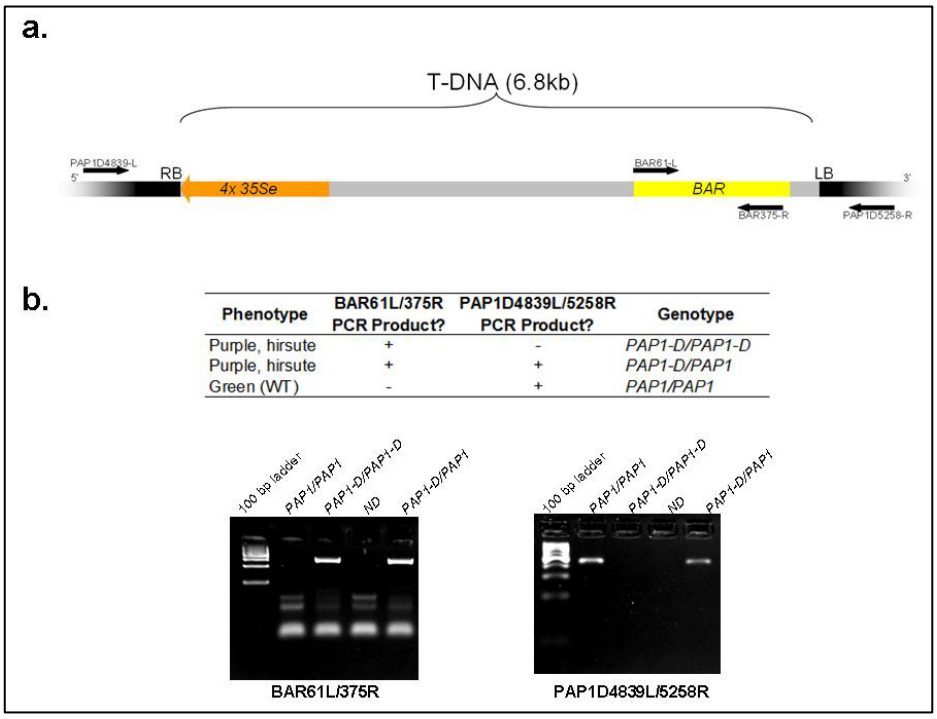


Figure 3.3: *PAP1-D* PCR-based genotyping. **a:** Location of PCR primers used to genotype individuals for the *PAP1-D* allele. RB/LB: T-DNA left and right borders, respectively; 4x35Se: 4 tandem copies of the cauliflower mosaic virus 35S transcript promoter; BAR: gene conferring resistance to hygromycin (basta) **b:** The combined results of the BAR61L/375R and PAP1D4839L/5258R PCR assays can be used to infer the zygosity of individuals with respect to *PAP1-D*. Sample electrophoretograms (4% agarose) using the two primer sets are shown for homozygous *PAP1*, homozygous *PAP1-D*, and heterozygous individuals. ND = no DNA template added to mastermix (negative control).

3.1.5 Plant hybridizations

Individuals of the desired genotypes were grown to flowering (approximately 3 weeks) in preparation for conducting crosses in Tier 1 and Tier 2. Crosses were carried out manually with the aid of a Zeiss Stemi 2000-C stereomicroscope. Prior to donor pollen application, anthers were removed from recipient *HTH/HTH* and *HTH/hth* flowers to reduce the chance of self-pollination. For each cross, mature dehiscent anthers were used to pollinate the stigma of 2 – 8 receptive female flowers. All crosses were carried out reciprocally. Plants generated in subsequent Tiers were allowed to self-pollinate.

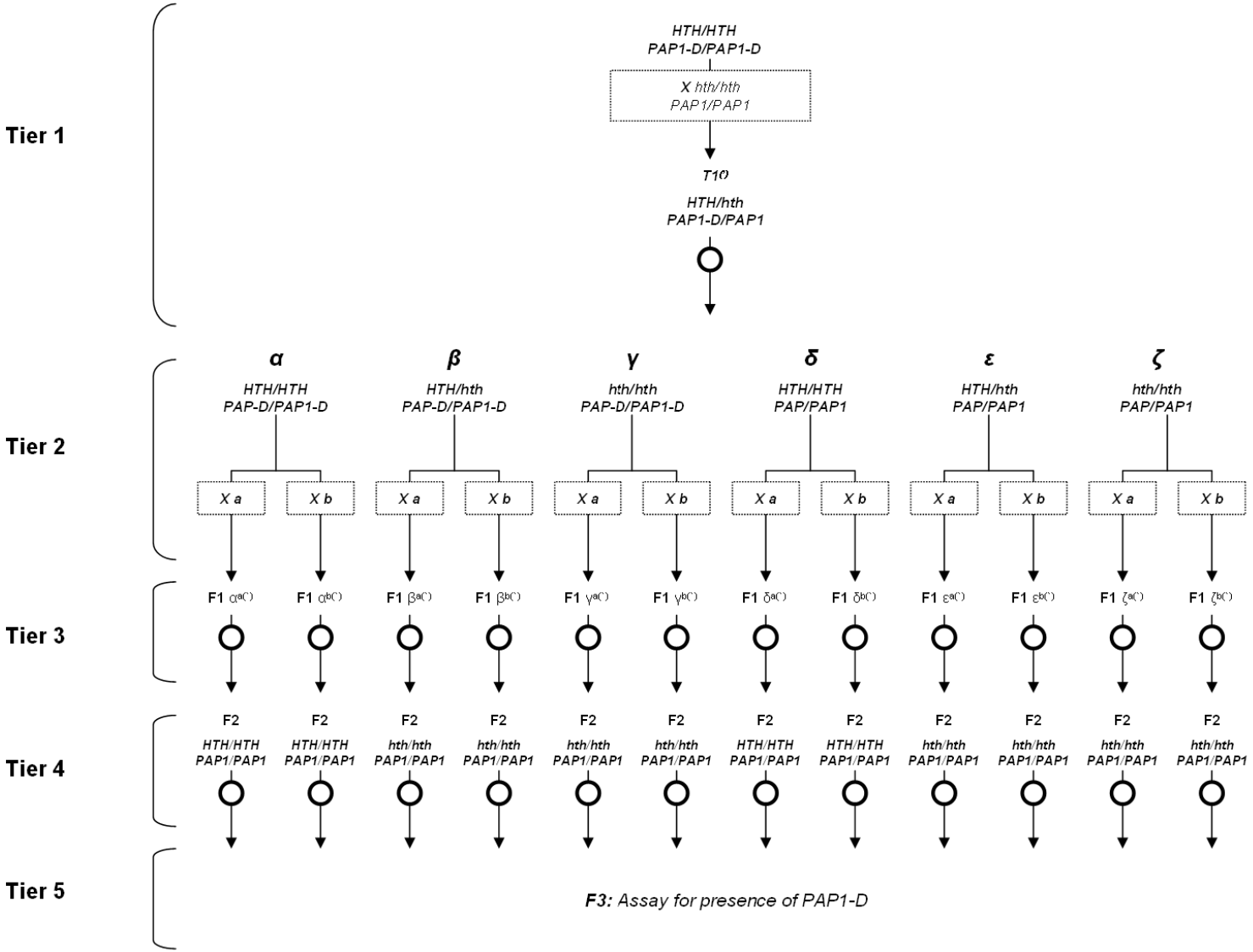
3.2 Results

3.2.1 Pedigree Construction

Constructing PAP1-D pedigrees in the presence of hth

Three replicates of the experimental pedigree (shown in **Figure 3.4** and described in detail in **Appendix A**) were undertaken to monitor the segregation of the *PAP1-D* and *HTH* alleles in the presence of the *hth* mutation. One pedigree was constructed using the *hth-4* allele (designated the *hth-4** pedigree) and two were constructed using the *hth-8* allele (designated the *hth-8** and *hth-8* pedigrees). To facilitate the cataloguing of seed stocks and individual plants, a detailed cataloguing system was devised (described in detail in **Appendix C**). Following the various crossing and self-pollination steps required for the pedigree construction, the required genotypes were determined using a combination of phenotypic analysis and molecular genotyping. Since the original *PAP1-D* line was generated in the Col-0 ecotype (Borevitz *et al.*, 2000) and all of our *hth* alleles were generated in the Landsberg *erecta* (L*er*) background, the Tier 1 *PAP1-D/hth* hybrids (shown in **Appendix D**) were also L*er*/Col hybrids. The characteristic recessive *erecta* (*er*) allele was therefore scored in many Tier 3 and Tier 4 individuals, thus providing an additional marker to monitor for instability in Tier 5.

The Tier 2 (F1) hybridizations and the resulting Tier 3 founder lines of the *hth4**, *hth8**, and *hth8* pedigrees are summarized in **Appendix E i – iii**, respectively. A number of Tier 3 individuals were not generated due to failed hybridizations or the unavailability of a Tier 2 parent with the appropriate genotype. For example, in Tier 2 of all three pedigrees (*hth-4**, *hth-8**, and *hth-8*), the recombinant genotype [*HTH/HTH, PAP1/PAP1*] (δ lineage) could not be identified among individuals that descended from a Tier 1 hybridization in which the *hth/hth* parent was ♀ (*hth8*.T1.1014- and *hth8*.T1.1051-derived lines). The opposite was not true. In other words, if the *hth/hth* parent was ♂, the [*HTH/HTH*, *PAP1/PAP1*] genotype could be recovered.



F3: *Assay for presence of PAP1-D*

Figure 3.4: A flow chart summarizing the 5-Tier pedigree used to test the segregation of *PAP1-D* and *HTH* in 6 lineages (α – ζ) in the presence of a *hth* allele. In the Tier 2 backcrosses, "a" represents the genotype [*HTH/HTH, PAP1/PAP1*] while "b" represents [*hth/hth, PAP1/PAP1*]. All hybridizations in Tiers 1 and 2 were conducted reciprocally, as indicated by the dotted lines. Open circles indicate self-pollination. *Prime* (') indicates lineages for which reciprocals were also generated.

The difficulty associated with identifying some Tier 2 recombinant genotypes at the frequencies which would be expected from independent assortment initiated the experimental determination of the genetic distance between *HTH* and *PAP1* using a simple test cross. Briefly, the dihybrid [*HTH/hth, PAP1-D/PAP1*] was crossed to *hth* mutant plants [*hth/hth*, *PAP1/PAP1*] in a series of replicates and the recombinant genotypes were scored using phenotypic analysis (see **Appendix F**). The results indicate that the *HTH* and *PAP1* loci are 22.0 cM apart.

A detailed summary of the Tier 3 (F1) individuals selected for self-pollination in the *hth4*, hth8*,* and *hth8* pedigrees is provided in **Appendix G i – iii**, respectively. Similarly, the resulting Tier 4 lines which were selected as F2 parents for the *hth4**, *hth8**, and *hth8* pedigrees are shown in **Appendix H i – iii**, respectively.

Constructing PAP1-D pedigrees in the absence of hth

One replicate of the pedigree designed to test the segregation of *PAP1-D* and *HTH* in the absence of mutant *hth* alleles (designated the *HTH* pedigree, shown in **Figure 3.5** and described in further detail in **Appendix A**) was undertaken. As was the case with the pedigrees described above, the genotypes of each Tier were determined using a combination of molecular and phenotypic analysis. The Tier 1 hybrids are shown in **Appendix I**. The Tier 2 hybridizations and resulting Tier 3 (F1) lines are shown in **Appendix J**. The individuals selected for self-pollination in Tier 3 and Tier 4 are shown in **Appendix K** and **Appendix L**, respectively.

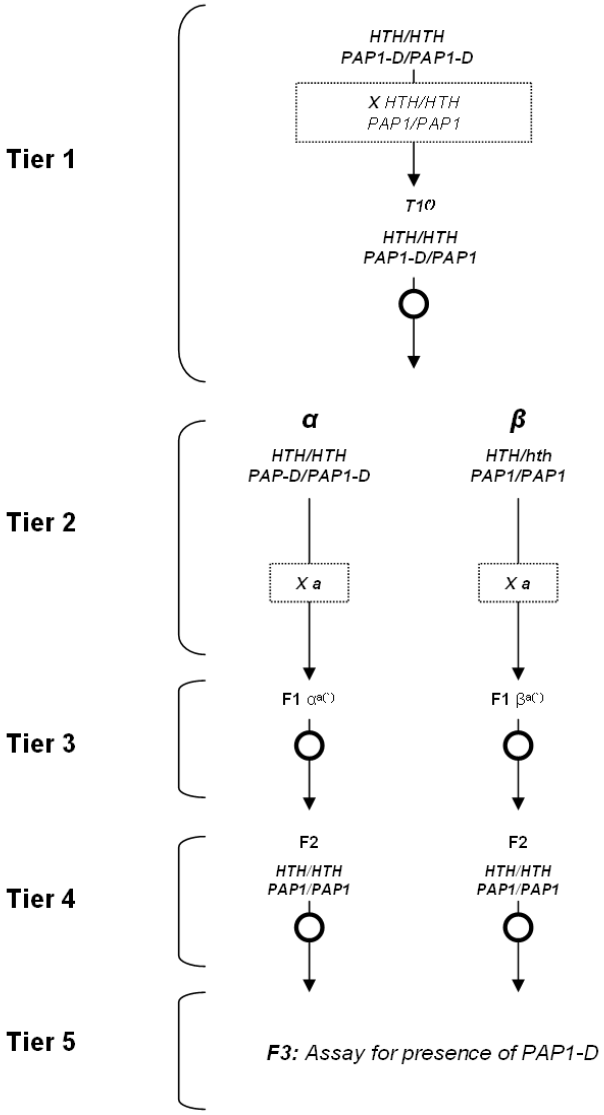


Figure 3.5: A flow chart summarizing the 5-Tier pedigree used to test the segregation of *PAP1-D* and *HTH* in the absence of a *hth* allele (wild-type *HTH* background). The pedigree is analogous to that described in **Figure 3.4** except that a mutant *hth* allele is not present. In the Tier 2 backcrosses, "a" represents the genotype [*HTH/HTH, PAP1/PAP1*]. All hybridizations in Tiers 1 and 2 were conducted reciprocally. Open circles indicate self-pollination. *Prime* (') indicates lineages for which reciprocals were also generated.

3.2.2 Tier 5 results: screening for segregation of non-parental *PAP1-D*, *HTH*, and *ER* alleles

The Tier 5 (F3) progeny were planted *en masse* to test for segregation of non-parental alleles. The number of Tier 5 seeds available for planting was limited by the fecundity of the Tier 4 (F2) parent and, to a greater extent, restricted by the availability of growth chamber space. Wild-type L*er* plants, the original *PAP1-D/PAP1-D* (CS3884) transgenic line, and several F2 progeny of plants segregating *PAP1-D* and *hth* were included in each batch of plantings to provide a phenotype reference for any variation that might occur given the specific growth conditions experienced by each batch of plants.

The detailed Tier 5 results from the *hth-4**, *hth-8**, *hth-8,* and *HTH* pedigrees are presented separately below.

3.2.2.1 hth-4** pedigree Tier 5 results*

Phenotypic analysis

Due to limited space, the *hth4** pedigree Tier 5 planting was divided into 2 smaller plantings. The first planting (01/07/2009) coincided with an error in the programming of the chamber lighting conditions, during which the light intensity increased overnight from the desired intensity of approximately 160 µmol m^{-2} sec^{-1} at pot level to approximately 220 µmol m^{-2} sec^{-1} at pot level. This fluctuation of light intensity stressed the plants and compromised their vegetative growth rate and flowering response. One additional response to this unexpected source of stress was a noticeable increase in production of a purple pigment (resembling anthocyanin accumulation) which became particularly apparent after 2 – 3 weeks of growth. This dramatic increase in stress pigmentation was observed in the majority of Tier 5 plants which experienced the environmental fluctuations due to the programming error, including the wild-type L*er* controls. The appearance of the stress pigmentation differed markedly from the *PAP1-D* phenotype in general color, intensity, localization on the plant, and time of onset. Therefore 1 – 2 week old plants could still be confidently scored for *PAP1-D*. The chamber programming error was corrected before the second

*hth-4** Tier 5 planting (02/24/2009), therefore stress-induced pigment accumulation was no longer a major concern.

In total, the progeny of 42 Tier 4 parents representing 33 distinct ancestries (i.e. distinct with respect to the presence or absence of a *hth* allele, F1 *hth* zygosity, and the direction of the Tier 1 and Tier 2 crosses) were screened for segregation of the non-parental alleles using phenotypic analysis. The results of the phenotypic analysis of the Tier 5 progeny are shown in **Table 3.1**.

Of the 42 Tier 4 parents, 30 were of the genotype [*hth/hth, PAP1/PAP1*] (22 distinct ancestries) and 12 were [*HTH/HTH, PAP1/PAP1*] (11 distinct ancestries). Based on visual screening, restoration of the *PAP1-D* phenotype was not observed in any of the 4065 Tier 5 progeny generated in the *hth-4** pedigree.

The floral phenotype of 4028 Tier 5 individuals was also scored. A total of 37 individuals had not yet bolted at the time therefore their floral phenotype could not be determined. Based on the assessment of floral morphology, the *hth-4* allele was stably inherited in the 1166 progeny that descended from the 30 Tier 4 parents of the genotype [*hth/hth, PAP1/PAP1*]. As expected, all of the 2862 progeny that descended from the twelve Tier 4 parents of the genotype [*HTH/HTH, PAP1/PAP1*] exhibited the wild-type floral morphology. The average number of Tier 5 progeny scored per [*HTH/HTH, PAP1/PAP1*] line was over two-fold higher than in the [*hth/hth, PAP1/PAP1*] lines (68 versus 28, respectively).

The *erecta* phenotype could only be scored for 15/42 Tier 4 parents in the *hth-4** pedigree. Of these 15 individuals, six were [*er/er*] while the remainder were wild-type. The [*er/er*] phenotype was stably inherited in the progeny of the six [*er/er*] Tier 4 parents. The nine Tier 4 parents scored as *ER* all segregated the *ER* allele among their Tier 5 progeny.

Table 3.1: Phenotypic screening of Tier 5 of the *hth-4** pedigree. See body text for details.

Tier 4 Parents				Tier 5 Progeny						
Tier 4 Catalog Number	***HTH* genotype**	***PAP1* genotype**	**Growth Habit**	**Tier 5 Total**	**Fused Flowers (*hth/hth*)**	**WT Flowers (*HTH*)**	**Did Not Bolt**	**Purple, Hirsute (*PAP1-D*)**	***ER* Segregating?**	**Planting Date**
hth4*.T3.α[a].4101	*HTH/HTH*	*PAP1/PAP1*	*ER*	242	0	242	0	0	y	02/24/09
hth4*.T3.α[a'].4001	*HTH/HTH*	*PAP1/PAP1*	*er/er*	68	0	68	0	0	n	02/24/09
hth4*.T3.α[b'].4107	*HTH/HTH*	*PAP1/PAP1*	*er/er*	102	0	102	0	0	n	02/24/09
hth4*.T3.β[b].4117	*hth/hth*	*PAP1/PAP1*	n/a	26	25	0	1	0	n/a	01/07/09
hth4*.T3.β[b].4118	*hth/hth*	*PAP1/PAP1*	n/a	32	28	0	4	0	n/a	01/07/09
hth4*.T3.β[b'].4120	*hth/hth*	*PAP1/PAP1*	n/a	3	3	0	0	0	n/a	01/07/09
hth4*.T3.γ[a].4124	*hth/hth*	*PAP1/PAP1*	n/a	23	23	0	0	0	n/a	01/07/09
hth4*.T3.γ[a'].4128	*hth/hth*	*PAP1/PAP1*	n/a	27	26	0	1	0	n/a	01/07/09
hth4*.T3.γ[b].4005	*hth/hth*	*PAP1/PAP1*	n/a	101	101	0	0	0	n/a	01/07/09
hth4*.T3.γ[b'].4009	*hth/hth*	*PAP1/PAP1*	n/a	44	43	0	1	0	n/a	01/07/09
hth4*.T3.γ[b'].4010	*hth/hth*	*PAP1/PAP1*	n/a	35	35	0	0	0	n/a	01/07/09
hth4*.T3.ε[a].4013	*hth/hth*	*PAP1/PAP1*	n/a	49	47	0	2	0	n/a	01/07/09
hth4*.T3.ε[a'].4019	*hth/hth*	*PAP1/PAP1*	n/a	68	68	0	0	0	n/a	01/07/09
hth4*.T3.ε[b'].4023	*hth/hth*	*PAP1/PAP1*	*er/er*	49	49	0	0	0	n	02/24/09
hth4*.T3.ε[b'].4024	*hth/hth*	*PAP1/PAP1*	*er/er*	36	36	0	0	0	n	02/24/09
hth4*.T3.ζ[a].4160	*hth/hth*	*PAP1/PAP1*	n/a	39	38	0	1	0	n/a	01/07/09
hth4*.T3.ζ[a].4162	*hth/hth*	*PAP1/PAP1*	n/a	49	47	0	2	0	n/a	01/07/09
hth4*.T3.ζ[a'].4051	*hth/hth*	*PAP1/PAP1*	n/a	37	36	0	1	0	n/a	01/07/09
hth4*.T3.ζ[a'].4053	*hth/hth*	*PAP1/PAP1*	n/a	35	34	0	1	0	n/a	01/07/09
hth4*.T3.ζ[b].4094	*hth/hth*	*PAP1/PAP1*	n/a	60	60	0	0	0	n/a	01/07/09
hth4*.T3.ζ[b'].4095	*hth/hth*	*PAP1/PAP1*	n/a	55	55	0	0	0	n/a	01/07/09
hth4*.T3'.α[a].4054	*HTH/HTH*	*PAP1/PAP1*	*ER*	215	0	215	0	0	y	02/24/09
hth4*.T3'.α[a'].4058	*HTH/HTH*	*PAP1/PAP1*	*ER*	454	0	454	0	0	y	02/24/09
hth4*.T3'.α[b].4135	*HTH/HTH*	*PAP1/PAP1*	*ER*	256	0	256	0	0	y	02/24/09
hth4*.T3'.α[b'].4185	*HTH/HTH*	*PAP1/PAP1*	*er/er*	91	0	91	0	0	n	02/24/09
hth4*.T3'.α[b'].4188	*HTH/HTH*	*PAP1/PAP1*	*ER*	90	0	90	0	0	y	02/24/09
hth4*.T3'.β[a'].4143	*hth/hth*	*PAP1/PAP1*	n/a	21	18	0	3	0	n/a	01/07/09
hth4*.T3'.β[b].4151	*hth/hth*	*PAP1/PAP1*	n/a	23	23	0	0	0	n/a	01/07/09
hth4*.T3'.γ[a].4156	*hth/hth*	*PAP1/PAP1*	n/a	35	33	0	2	0	n/a	01/07/09
hth4*.T3'.γ[b].4063	*hth/hth*	*PAP1/PAP1*	n/a	40	40	0	0	0	n/a	01/07/09
hth4*.T3'.γ[b].4065	*hth/hth*	*PAP1/PAP1*	*er/er*	83	83	0	0	0	n	02/24/09
hth4*.T3'.δ[a].4097	*HTH/HTH*	*PAP1/PAP1*	*ER*	364	0	364	0	0	y	02/24/09
hth4*.T3'.δ[a'].4096	*HTH/HTH*	*PAP1/PAP1*	*ER*	283	0	283	0	0	y	02/24/09
hth4*.T3'.δ[b].4070	*HTH/HTH*	*PAP1/PAP1*	*ER*	463	0	463	0	0	y	02/24/09
hth4*.T3'.δ[b'].4069	*HTH/HTH*	*PAP1/PAP1*	*ER*	234	0	234	0	0	y	02/24/09
hth4*.T3'.ε[a].4075	*hth/hth*	*PAP1/PAP1*	n/a	45	44	0	1	0	n/a	01/07/09
hth4*.T3'.ε[a'].4079	*hth/hth*	*PAP1/PAP1*	n/a	30	28	0	2	0	n/a	01/07/09
hth4*.T3'.ε[a'].4081	*hth/hth*	*PAP1/PAP1*	n/a	38	34	0	4	0	n/a	01/07/09
hth4*.T3'.ε[b].4166	*hth/hth*	*PAP1/PAP1*	n/a	31	29	0	2	0	n/a	01/07/09
hth4*.T3'.ε[b].4169	*hth/hth*	*PAP1/PAP1*	n/a	32	26	0	6	0	n/a	01/07/09
hth4*.T3'.ζ[a].4090	*hth/hth*	*PAP1/PAP1*	n/a	33	31	0	2	0	n/a	01/07/09
hth4*.T3'.ζ[b].4098	*hth/hth*	*PAP1/PAP1*	n/a	24	23	0	1	0	n/a	01/07/09
			Total:	4065	1166	2862	37	0		

Screening for partial <u>PAP1-D</u> restoration

The identification of genetic mosaics among other experimental lines screened with our indel markers prompted a more comprehensive screening of some Tier 5 progeny. In particular, we looked for mosaic patterns of *PAP1-D* reversion in the *hth-4** pedigree which we hypothesized would be revealed as localized anthocyanin accumulation on an otherwise green leaf. Among the 4065 Tier 5 progeny of the *hth-4** pedigree, one two week old individual (catalogue number *hth4**.T5'.$\beta^{a'}$.5006^{4143}) was identified with particularly striking localized purple pigment accumulation in one of its rosette leaves (see **Figure 3.6**). The rosette leaf was removed and dissected into three pieces (P1, P2, P3), each of which was immediately subjected to PCR genotyping for the *PAP1-D* allele, as described in section *3.1.4*. The results are shown in **Figure 3.7**. Amplification of DNA samples obtained from P2 tissue gave a positive result with both the BAR61L/375R and PAP1D4839L/5258R primer sets, consistent with the genotype being [*PAP1-D/PAP1*], although the BAR61L/375R PCR product was very faint. P1 could not be definitively genotyped due to unsuccessful amplification with the PAP1D4839L/5258R primer set, yet the BAR61L/375R PCR produced an extremely faint signal from this sample. P3 was determined to be of the parental genotype (*PAP1/PAP1*). Several attempts to reproduce this result with the original P2 DNA sample or to re-amplify the P2 BAR61L/375R PCR product with a second round of PCR were unsuccessful, as were attempts to clone the 314 bp P2 BAR61L/375R PCR product (data not shown).

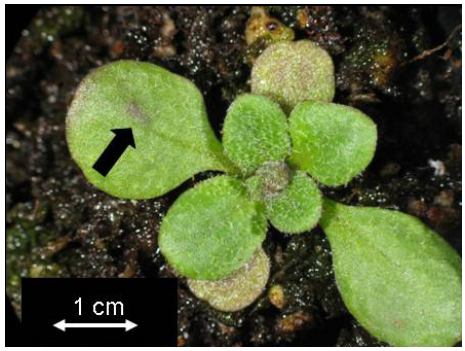


Figure 3.6: A localized patch of purple pigmentation in the cauline leaf of individual *hth4**.T5'.$\beta^{a'}$.5006^{4143} prompted PCR-based molecular screening for the *PAP1-D* allele.

Molecular screening of rosette leaves from an addition 22 Tier 5 individuals with varying levels of anthocyanin-like pigment accumulation revealed another individual (*hth4**.T5'.$\delta^{a'}$.5036^{4097}) that was

determined to be *PAP1-D*/*PAP1* based on *PAP1-D* PCR genotyping (**Figure 3.8**). As was the case with *hth4**.T5'.$\beta^{a'}$.5006^{4143}, the BAR61L/375R PCR product was faint, and could not be reproduced nor re-amplified when subjected to a second round of PCR. Attempts to clone the 314 bp PCR product were also unsuccessful.

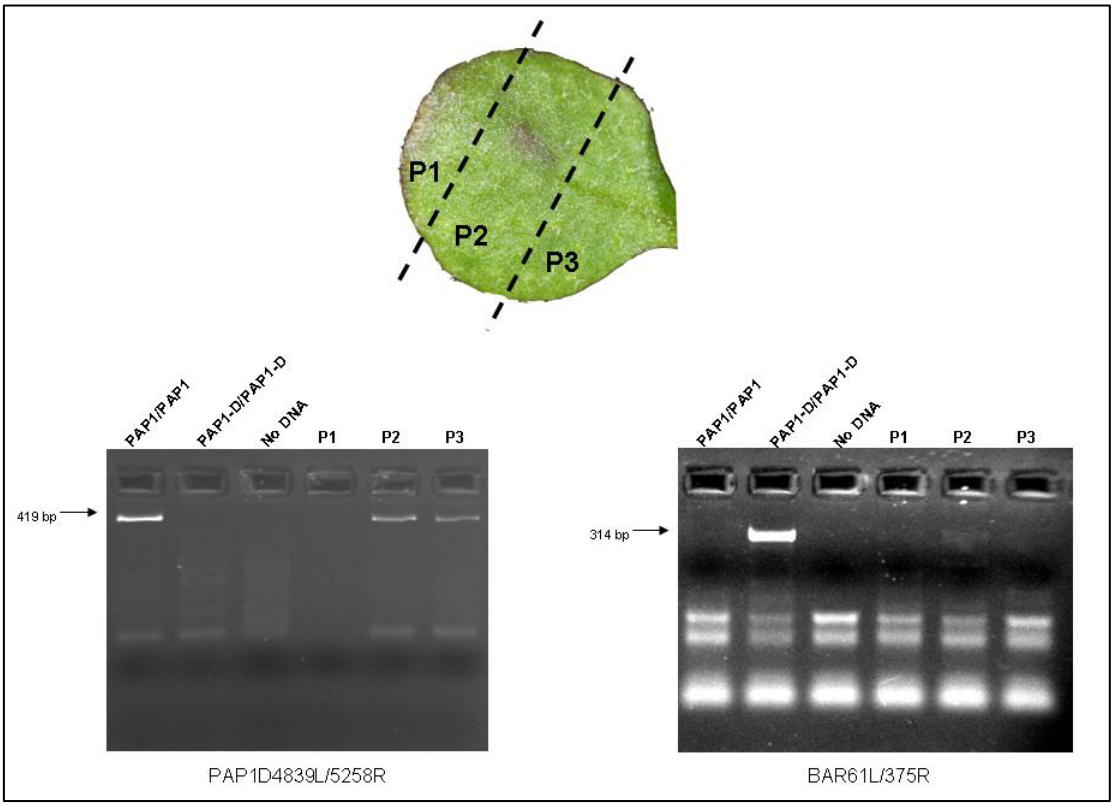


Figure 3.7: *PAP1-D* genotyping of P1, P2, and P3 samples of a cauline leaf of *hth4**.T5'.$\beta^{a'}$.5006^{4143}. P2 produced positive signals that co-migrated with the positive controls for both the PAP1D4839L/5258R and BAR61L/375R assays. The combined results suggest that P2 is heterozygous for the *PAP1-D* transgene, although the BAR61L/375R PCR product was very faint. An extremely faint 314 bp band is also visible in the P1 lane. The contrast of the BAR61L/375R electrophoretogram has been increased to improve visualization of the faint bands.

To test if Tier 5 individuals *hth4**.T5'.$\beta^{a'}$.5006^{4143} and *hth4**.T5'.$\delta^{a'}$.5036^{4097} could segregate the *PAP1-D* allele among their progeny, these two individuals were allowed to self pollinate and seed was collected. Approximately 100 progeny from each of the two lines were scored. Segregation of non-parental alleles was not observed in any of the progeny of either Tier 5 individuals (**see Table 3.2**).

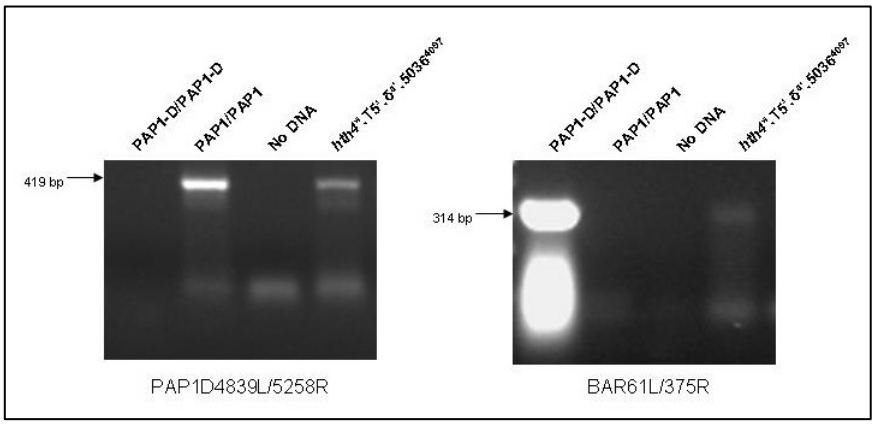


Figure 3.8: *PAP1-D* genotyping a cauline leaf of Tier 5 individual *hth4**.T5'.$\delta^{a'}$.5036^{4097}. Positive signals that co-migrated with the positive controls from both the PAP1D4839L/5258R and BAR61L/375R assays indicate that the sample is heterozygous for the *PAP1-D* allele. The contrast of the BAR61L/375R electrophoretogram was increased to improve visualization of the 314 bp PCR product.

Table 3.2: Results of the phenotypic scoring of progeny from the selfed Tier 5 individuals *hth4**.T5'.$\beta^{a'}$.5006^{4143} and *hth4**.T5'.$\delta^{a'}$.5036^{4097}.

Tier 5 Parents				Progeny						
Tier 5 Catalog Number	*HTH* genotype	*PAP1* genotype	Growth Habit	Total Progeny	Fused Flowers (*hth/hth*)	WT Flowers (*HTH*)	Did Not Bolt	Purple, Hirsute (*PAP1-D*)	*ER* Segregating?	Planting Date
hth4*.T4'.$\beta^{a'}$.4143	*hth/hth*	*PAP1/PAP1*	*n/a*	93	93	0	0	0	n	03/24/09
hth4*.T4'.δ^{a}.4097	*HTH/HTH*	*PAP1/PAP1*	*ER*	76	0	76	0	0	y	03/24/09

As a negative control for the *PAP1-D* genotyping conducted above, 100 wild-type seven-day old seedlings (50 each of the L*er* and Col ecotypes) were assayed using the BAR61L/375R primer set. These lines were obtained from our wild-type seed stocks that presumably never had the *PAP1-D* or *BAR* allele in their ancestry. Nevertheless, a single Col seedling (#39) yielded a faint BAR61L/375R PCR product which was highly reminiscent of the faint BAR61L/375R PCR products observed in *hth4**.T5'.$\beta^{a'}$.5006^{4143} and *hth4**.T5'.$\delta^{a'}$.5036^{4097} (see **Figure 3.9**). The remaining 99 seedlings were negative for BAR61L/375R PCR (data not shown).

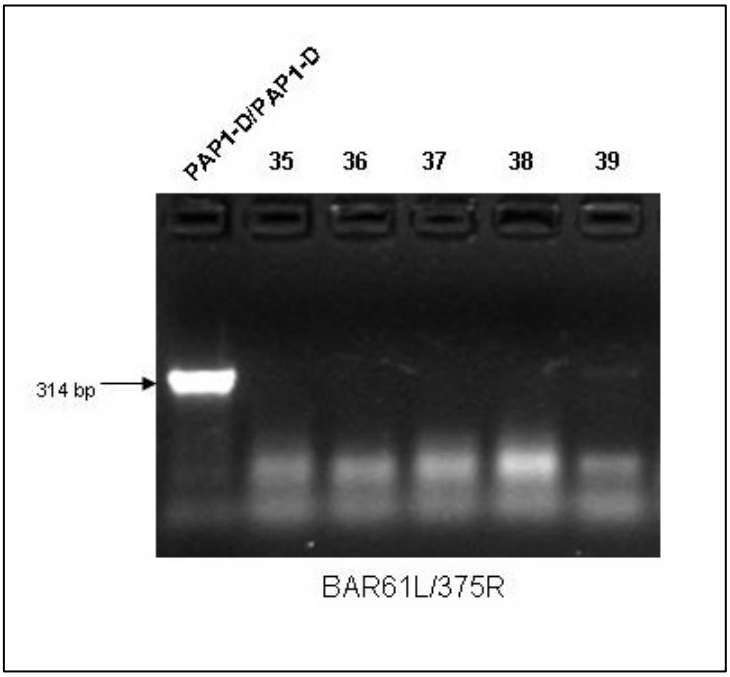


Figure 3.9: Genotyping 50 wild-type Col and L*er* seedlings with the BAR61L/375R primer set revealed one Col individual (#39) that exhibited a faint band that co-migrated with the 314 bp positive control PCR product. The band was faint and not unlike those observed in the Tier 5 individuals *hth4**.T5'.$\beta^{a'}$.5006^{4143} and *hth4**.T5'.$\delta^{a'}$.5036^{4097}. The contrast of this electrophoretogram was adjusted to improve visualization of the 314 bp PCR product in sample #39.

3.2.2.2 hth-8 pedigree Tier 5 results*

In the *hth-8** pedigree, the progeny of 31 Tier 4 parents representing 27 distinct ancestries were planted and screened for the segregation of non-parental alleles using phenotypic analysis. The results of the phenotypic analysis of the Tier 5 progeny of the *hth-8** pedigree are shown in **Table 3.3**.
Of the 31 Tier 4 parents, 27 were of the genotype [*hth/hth, PAP1/PAP1*] (23 distinct ancestries) and four were [*HTH/HTH, PAP1/PAP1*] (four distinct ancestries). Based on visual screening for the *PAP1-D* purple hirsute phenotype, restoration of the *PAP1-D* phenotype was not observed in any of the 7181 Tier 5 progeny of the *hth-8** pedigree.

The floral phenotype of 7066 Tier 5 individuals was determined. A total of 115 individuals had not bolted at the time of scoring therefore their floral phenotype could not be scored. According to the presence of the fusion floral morphology, the *hth-8* allele was stably inherited in the 5706 progeny that descended from the 27 Tier 4 parents that were of the genotype [*hth/hth, PAP1/PAP1*]. All of the 1360 progeny that descended from the four Tier 4 parents of the genotype [*HTH/HTH, PAP1/PAP1*] exhibited the wild-type floral morphology. The average number of Tier 5 progeny scored per [*HTH/HTH, PAP1/PAP1*] line was 340 compared to 211 in the [*hth/hth, PAP1/PAP1*] lines.

The *erecta* phenotype was scored for all Tier 4 parents in the *hth-8** pedigree. Of the 31 parents, 15 were [*er/er*] and 16 were wild-type (*ER*). The [*er/er*] phenotype was stably inherited in the progeny of all 15 [*er/er*] parents. The 16 parents scored as *ER* all segregated the *ER* allele among their Tier 5 progeny.

Table 3.3: Phenotypic screening of Tier 5 of the *hth-8** pedigree. See body text for details.

Tier 4 Parents				Tier 5 Progeny						
Tier 4 Catalog Number	***HTH* genotype**	***PAP1* genotype**	**Growth Habit**	**Tier 5 Total**	**Fused Flowers (*hth/hth*)**	**WT Flowers (*HTH*)**	**Did Not Bolt**	**Purple, Hirsute (*PAP1-D*)**	***ER* Segregating?**	**Planting Date**
hth8*.T4.γ^{a}.4059	*hth/hth*	*PAP1/PAP1*	*ER*	245	245	0	0	0	y	05/20/2009
hth8*.T4.γ$^{a'}$.4061	*hth/hth*	*PAP1/PAP1*	*er/er*	382	371	0	11	0	n	05/20/2009
hth8*.T4.γ^{b}.4055	*hth/hth*	*PAP1/PAP1*	*ER*	416	400	0	16	0	y	05/20/2009
hth8*.T4.γ$^{b'}$.4057	*hth/hth*	*PAP1/PAP1*	*ER*	211	207	0	4	0	y	05/20/2009
hth8*.T4.ε$^{a'}$.4053	*hth/hth*	*PAP1/PAP1*	*ER*	157	157	0	0	0	y	05/20/2009
hth8*.T4.ε^{b}.4047	*hth/hth*	*PAP1/PAP1*	*ER*	316	313	0	3	0	y	05/20/2009
hth8*.T4.ζ^{a}.4043	*hth/hth*	*PAP1/PAP1*	*er/er*	219	218	0	1	0	n	05/20/2009
hth8*.T4.ζ$^{a'}$.4045	*hth/hth*	*PAP1/PAP1*	*ER*	182	182	0	0	0	y	05/20/2009
hth8*.T4.ζ^{b}.4039	*hth/hth*	*PAP1/PAP1*	*ER*	209	209	0	0	0	y	05/20/2009
hth8*.T4.ζ$^{b'}$.4041	*hth/hth*	*PAP1/PAP1*	*ER*	202	202	0	0	0	y	05/20/2009
hth8*.T4'.α^{a}.4009	*HTH/HTH*	*PAP1/PAP1*	*ER*	319	0	315	4	0	y	05/20/2009
hth8*.T4'.α$^{a'}$.4011	*HTH/HTH*	*PAP1/PAP1*	*er/er*	466	0	448	18	0	n	05/20/2009
hth8*.T4'.α^{b}.4013	*hth/hth*	*PAP1/PAP1*	*er/er*	112	112	0	0	0	n	05/20/2009
hth8*.T4'.α^{b}.4014	*hth/hth*	*PAP1/PAP1*	*er/er*	290	287	0	3	0	n	05/20/2009
hth8*.T4'.α^{b}.4015	*hth/hth*	*PAP1/PAP1*	*er/er*	101	101	0	0	0	n	05/20/2009
hth8*.T4'.α$^{b'}$.4016	*hth/hth*	*PAP1/PAP1*	*ER*	398	386	0	12	0	y	05/20/2009
hth8*.T4'.α$^{b'}$.4017	*hth/hth*	*PAP1/PAP1*	*ER*	198	192	0	6	0	y	05/20/2009
hth8*.T4'.β^{b}.4037	*hth/hth*	*PAP1/PAP1*	*er/er*	213	213	0	0	0	n	05/20/2009
hth8*.T4'.β^{b}.4038	*hth/hth*	*PAP1/PAP1*	*er/er*	76	76	0	0	0	n	05/20/2009
hth8*.T4'.γ^{b}.4018	*hth/hth*	*PAP1/PAP1*	*ER*	216	216	0	0	0	y	05/20/2009
hth8*.T4'.γ$^{b'}$.4019	*hth/hth*	*PAP1/PAP1*	*er/er*	246	240	0	6	0	n	05/20/2009
hth8*.T4'.δ^{a}.4001	*HTH/HTH*	*PAP1/PAP1*	*er/er*	243	0	241	2	0	n	05/20/2009
hth8*.T4'.δ$^{a'}$.4003	*HTH/HTH*	*PAP1/PAP1*	*er/er*	363	0	356	7	0	n	05/20/2009
hth8*.T4'.δ^{b}.4005	*hth/hth*	*PAP1/PAP1*	*er/er*	119	118	0	1	0	n	05/20/2009
hth8*.T4'.δ$^{b'}$.4007	*hth/hth*	*PAP1/PAP1*	*ER*	83	82	0	1	0	y	05/20/2009
hth8*.T4'.ε^{b}.4025	*hth/hth*	*PAP1/PAP1*	*er/er*	88	88	0	0	0	n	05/20/2009
hth8*.T4'.ε$^{b'}$.4027	*hth/hth*	*PAP1/PAP1*	*er/er*	201	195	0	6	0	n	05/20/2009
hth8*.T4'.ζ^{a}.4033	*hth/hth*	*PAP1/PAP1*	*ER*	389	384	0	5	0	y	05/20/2009
hth8*.T4'.ζ$^{a'}$.4035	*hth/hth*	*PAP1/PAP1*	*er/er*	104	103	0	1	0	n	05/20/2009
hth8*.T4'.ζ^{b}.4029	*hth/hth*	*PAP1/PAP1*	*ER*	271	263	0	8	0	y	05/20/2009
hth8*.T4'.ζ$^{b'}$.4031	*hth/hth*	*PAP1/PAP1*	*ER*	146	146	0	0	0	y	05/20/2009
			Total:	7181	5706	1360	115	0		

3.2.2.3 hth-8 pedigree Tier 5 results

In the *hth-8* pedigree, the progeny of 38 Tier 4 parents collectively representing 32 distinct ancestries were planted and screened for the segregation of non-parental alleles using phenotypic analysis. The results of this screening are shown in **Table 3.4**.

Of the 38 Tier 4 parents, 31 were of the genotype *[hth/hth, PAP1/PAP1]* (25 distinct ancestries) and seven were *[HTH/HTH, PAP1/PAP1]* (seven distinct ancestries). Complete restoration of the *PAP1-D* phenotype was not observed in any of the 9091 Tier 5 progeny.

The floral morphology of 8647 Tier 5 individuals was scored (444 individuals had not bolted at the time of scoring). Based on floral phenotype, the *hth-8* allele was stably inherited in the 5905 progeny that descended from the 31 Tier 4 parents of the genotype [*hth/hth, PAP1/PAP1*]. All of the 2742 progeny that descended from the seven Tier 4 parents of the genotype [*HTH/HTH, PAP1/PAP1*] exhibited the wild-type floral morphology. The average number of Tier 5 progeny scored per [*HTH/HTH, PAP1/PAP1*] line was 391 compared to 190 in the [*hth/hth, PAP1/PAP1*] lines.

The *erecta* phenotype was scored for all Tier 4 parents in the *hth-8* pedigree. Of the 38 parents, 18 were [*er/er*] and 20 were wild-type (*ER*). The [*er/er*] phenotype was stably inherited in the progeny of all 18 [*er/er*] parents. The 20 parents scored as *ER* all segregated the *ER* allele among their Tier 5 progeny.

Table 3.4: Phenotypic screening of Tier 5 of the *hth-8* pedigree. See body text for details.

Tier 4 Parents				Tier 5 Progeny						
Tier 4 Catalog Number	*HTH* genotype	*PAP1* genotype	Growth Habit	Tier 5 Total	Fused Flowers (*hth/hth*)	WT Flowers (*HTH*)	Did Not Bolt	Purple, Hirsute (*PAP1-D*)	*ER* Segregating?	Planting Date
hth8.T4.αa.4021	*HTH/HTH*	*PAP1/PAP1*	*ER*	511	0	482	29	0	y	06/05/09
hth8.T4.αa'.4020	*HTH/HTH*	*PAP1/PAP1*	*ER*	409	0	388	21	0	y	06/05/09
hth8.T4.αb.4023	*hth/hth*	*PAP1/PAP1*	*ER*	260	252	0	8	0	y	06/05/09
hth8.T4.αb.4024	*hth/hth*	*PAP1/PAP1*	*ER*	212	198	0	14	0	y	06/05/09
hth8.T4.αb'.4022	*hth/hth*	*PAP1/PAP1*	*ER*	362	344	0	18	0	y	06/05/09
hth8.T4.βa.4025	*hth/hth*	*PAP1/PAP1*	*er/er*	114	114	0	0	0	n	06/05/09
hth8.T4.βb.4026	*hth/hth*	*PAP1/PAP1*	*ER*	198	186	0	12	0	y	06/05/09
hth8.T4.βb.4027	*hth/hth*	*PAP1/PAP1*	*er/er*	336	319	0	17	0	n	06/05/09
hth8.T4.γa'.4028	*hth/hth*	*PAP1/PAP1*	*ER*	48	45	0	3	0	y	06/05/09
hth8.T4.γa'.4029	*hth/hth*	*PAP1/PAP1*	*ER*	440	421	0	19	0	y	06/05/09
hth8.T4.γb.4030	*HTH/HTH*	*PAP1/PAP1*	*er/er*	123	0	117	6	0	n	06/05/09
hth8.T4.εa.4032	*hth/hth*	*PAP1/PAP1*	*ER*	200	193	0	7	0	y	06/05/09
hth8.T4.εa.4033	*hth/hth*	*PAP1/PAP1*	*ER*	404	377	0	27	0	y	06/05/09
hth8.T4.εa'.4031	*hth/hth*	*PAP1/PAP1*	*ER*	219	208	0	11	0	y	06/05/09
hth8.T4.εb.4035	*hth/hth*	*PAP1/PAP1*	*ER*	388	367	0	21	0	y	06/05/09
hth8.T4.εb.4036	*hth/hth*	*PAP1/PAP1*	*ER*	304	291	0	13	0	y	06/05/09
hth8.T4.εb'.4034	*hth/hth*	*PAP1/PAP1*	*ER*	146	142	0	4	0	y	06/05/09
hth8.T4.ζa.4038	*hth/hth*	*PAP1/PAP1*	*er/er*	16	4	0	12	0	n	06/05/09
hth8.T4.ζa'.4037	*hth/hth*	*PAP1/PAP1*	*er/er*	86	83	0	3	0	n	06/05/09
hth8.T4.ζb.4040	*hth/hth*	*PAP1/PAP1*	*er/er*	97	93	0	4	0	n	06/05/09
hth8.T4.ζb'.4039	*hth/hth*	*PAP1/PAP1*	*er/er*	190	185	0	5	0	n	06/05/09
hth8.T4'.αa.4002	*HTH/HTH*	*PAP1/PAP1*	*ER*	391	0	374	17	0	y	06/05/09
hth8.T4'.αa'.4001	*HTH/HTH*	*PAP1/PAP1*	*er/er*	455	0	436	19	0	n	06/05/09
hth8.T4'.αb.4003	*hth/hth*	*PAP1/PAP1*	*ER*	299	285	0	14	0	y	06/05/09
hth8.T4'.βa.4006	*hth/hth*	*PAP1/PAP1*	*er/er*	171	165	0	6	0	n	06/05/09
hth8.T4'.βb.4007	*hth/hth*	*PAP1/PAP1*	*er/er*	78	75	0	3	0	n	06/05/09
hth8.T4.γa'.4008	*hth/hth*	*PAP1/PAP1*	*ER*	374	363	0	11	0	y	06/05/09
hth8.T4'.γb'.4009	*hth/hth*	*PAP1/PAP1*	*er/er*	159	135	0	24	0	n	06/05/09
hth8.T4'.δa.4011	*HTH/HTH*	*PAP1/PAP1*	*er/er*	540	0	515	25	0	n	06/05/09
hth8.T4'.δa'.4010	*HTH/HTH*	*PAP1/PAP1*	*er/er*	443	0	430	13	0	n	06/05/09
hth8.T4'.δb.4012	*hth/hth*	*PAP1/PAP1*	*ER*	309	295	0	14	0	y	06/05/09
hth8.T4'.εa'.4013	*hth/hth*	*PAP1/PAP1*	*er/er*	49	42	0	7	0	n	06/05/09
hth8.T4'.εb.4016	*hth/hth*	*PAP1/PAP1*	*er/er*	51	47	0	4	0	n	06/05/09
hth8.T4'.εb'.4014	*hth/hth*	*PAP1/PAP1*	*er/er*	80	80	0	0	0	n	06/05/09
hth8.T4'.εb'.4015	*hth/hth*	*PAP1/PAP1*	*er/er*	137	128	0	9	0	n	06/05/09
hth8.T4'.ζa.4018	*hth/hth*	*PAP1/PAP1*	*ER*	260	248	0	12	0	y	06/05/09
hth8.T4'.ζa'.4017	*hth/hth*	*PAP1/PAP1*	*er/er*	10	5	0	5	0	n	06/05/09
hth8.T4'.ζb.4019	*hth/hth*	*PAP1/PAP1*	*ER*	222	215	0	7	0	y	06/05/09
			Total:	9091	5905	2742	444	0		

3.2.2.4 HTH pedigree Tier 5 results

In the *HTH* pedigree, the progeny of eight Tier 4 parents (representing eight distinct ancestries) were planted and screened for segregation of the non-parental *PAP1-D* allele using phenotypic analysis. The results of the phenotypic screening of Tier 5 of the *HTH* pedigree are shown in **Table 3.5**.

All eight of the Tier 4 parents were *[HTH/HTH, PAP1/PAP1]* and exhibited the *ER* growth habit. A total of 2671 Tier 5 individuals were scored for pigmentation, and 2593 were scored for their floral phenotype and growth habit (78 individuals had not bolted at the time of scoring). Based on the phenotypic screening, the *HTH* and *PAP1* alleles were both stably inherited among all progeny. All of the Tier 4 parents segregated the *ER* allele among their Tier 5 progeny.

Table 3.5: Phenotypic screening of Tier 5 of the *HTH* pedigree. See body text for details

Tier 4 Parents				Tier 5 Progeny						
Tier 4 Catalog Number	*HTH* genotype	*PAP1* genotype	Growth Habit	Tier 5 Total	Fused Flowers (*hth/hth*)	WT Flowers (*HTH*)	Did Not Bolt	Purple, Hirsute (*PAP1-D*)	*ER* Segregating?	Planting Date
HTH.T4.α^{a}.4002	*HTH/HTH*	*PAP1/PAP1*	*ER*	342	0	330	12	0	y	05/22/09
HTH.T4.$\alpha^{a'}$.4001	*HTH/HTH*	*PAP1/PAP1*	*ER*	297	0	295	2	0	y	05/22/09
HTH.T4.β^{a}.4004	*HTH/HTH*	*PAP1/PAP1*	*ER*	333	0	323	10	0	y	05/22/09
HTH.T4.$\beta^{a'}$.4003	*HTH/HTH*	*PAP1/PAP1*	*ER*	329	0	320	9	0	y	05/22/09
HTH.T4'.α^{a}.4007	*HTH/HTH*	*PAP1/PAP1*	*ER*	301	0	290	11	0	y	05/22/09
HTH.T4'.$\alpha^{a'}$.4006	*HTH/HTH*	*PAP1/PAP1*	*ER*	381	0	366	15	0	y	05/22/09
HTH.T4'.β^{a}.4009	*HTH/HTH*	*PAP1/PAP1*	*ER*	323	0	311	12	0	y	05/22/09
HTH.T4'.$\beta^{a'}$.4010	*HTH/HTH*	*PAP1/PAP1*	*ER*	365	0	358	7	0	y	05/22/09
			Total:	2671	0	2593	78	0		

Chapter 4
Discussion

4.1 Characterizing reverted loci

The purpose of the work described in Chapter 2 was to characterize non-parental loci and flanking chromosomal regions at the DNA sequence level thereby providing a more detailed molecular picture of reversion events than had been described previously. Two approaches were used to accomplish this: (1) DNA sequence analysis to determine the fidelity and phase of non-parental sequences, and (2) analysis of flanking polymorphic markers to provide insight into the physical extent of non-parental sequences. This work complemented parallel studies in our laboratory and has improved our understanding of the extent, fidelity, and *cis*/*trans* relationship of reverted loci. The following sections focus on what we have learned about how the reappearance of non-parental alleles affects the genome of revertant individuals, the developmental patterns of restoration events, possible mechanisms governing restoration, and future research goals.

4.1.1 The footprint of restoration

Profiling revertant individuals using a series of PCR markers flanking the unstable indels F15H11_35.66 and MNJ8_44.33 in the progeny of 194E9PCL10 provided some insight into the genomic extent of these reversion events. In all cases, reversion at these indel markers appeared to be accompanied by reversion of the flanking markers, which suggests that markers separated by at least 80 kb can revert *concurrently*. Since all markers tested exhibited reversion, the endpoints of the putative reversion tracts could not be determined. Nevertheless, these data support the notion that reversion tracts can be extensive and may encompass large genomic intervals.

It is possible that concurrent reversion events extend across genomic regions which dwarf the scale of the 80 – 90 kb regions interrogated in this study. Some clues of the absolute extent of reversion may be

found in the results of the initial genome-wide profiling of the F3 progeny of 194E9PCL10 (Hopkins *et al.*, unpublished data). These data reveal that reversion at any given indel marker is frequently accompanied by reversion at other distal markers. For example, the indel marker MNJ8_44.33 on chromosome 5 is flanked by an indel marker (F2P16_80.35) approximately 5.3 Mb upstream and a second indel marker (MGI19_23.65) approximately 10.8 Mb downstream. Of the 53 F3 progeny that exhibited reversion at MNJ8_44.33, 84.9% (45/53) exhibited reversion of the flanking markers. Even more striking, of the F3 individuals that segregated non-parental markers, 45.6% (47/103) exhibited simultaneous reversion at seven independent markers across the genome. The number of F3 individuals in which only one marker was affected was comparatively small – only 4.9% (5/103). Collectively, these results suggest that concurrent events occur with some frequency and individual tracts can extend across large physical distances spanning large segments of chromosomes. This prediction is consistent with the observation that flanking markers within small (200-800 bp) genomic intervals revert concurrently, as revealed by the isolation and sequencing of *cis* chromosomal regions harboring non-parental indel markers.

The aforementioned results strongly support the hypothesis that the reversion events described in these individuals is a *template-directed* process (Lolle *et al.*, 2005a) and cannot be explained by any previously described stochastic mechanisms. DNA sequence data provide additional and compelling support for this hypothesis. Sequence analyses revealed that these novel sequences corresponded precisely to sequences found in the grandparental generation (F1 hybrids), suggesting that the template is of ancestral origin. The reappearance of insertions and flanking polymorphisms in these individuals cannot be attributed to random somatic mutation. However, since these molecular markers are co-dominant, it impossible to deduce if the *cis/trans* distribution (phase) of the markers is identical to that of the F1 without directly sequencing the affected intervals. However, sequencing data that was obtained demonstrated that short intervals reverted concurrently and reinstated sequences with absolute fidelity in *cis*. Therefore it is

possible that the concurrent restoration events observed across larger genomic regions also occurred in *cis*.

In the sample population tested in this study, no individuals were identified in which a double reversion event of one locus had occurred (i.e. an event affecting both homologous loci). However, such events have been observed previously (Lolle et al, 2005a; S.J. Lolle, personal communication: Ryan Lee, unpublished data). As shown in Figure **4.1**, work by Ryan Lee identified seven F3 individuals out of 100 tested that manifested double reversion events nested within a genomic interval that had experienced multiple single reversion events in flanking sequences. Lolle *et al.* (2005a) also described the recovery of *HTH/HTH* embryos from a *hth/hth* plant. From these data alone it is difficult to ascertain precisely how these sequence configurations might have been generated, however, these findings do suggest that *trans* reversion events can occur.

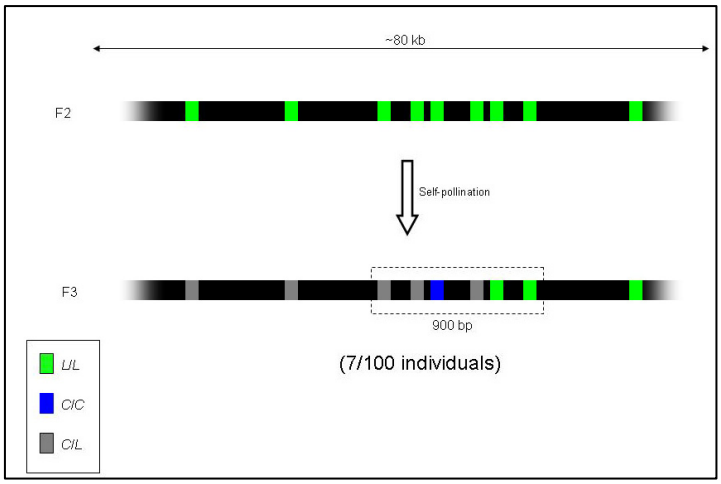


Figure 4.1: A single PCR marker on the F13C5 BAC exhibited a double reversion event from *L/L* to *C/C* in 7/100 F3 progeny, independent of the flanking markers which exhibited restoration to heterozygous (*C/L*). Another significant finding of these data is that the 'border' of a restoration event seems to have been identified. Adapted from Lee *et al.* (unpublished data).

One long-term goal of our ongoing research is to provide a detailed description of the extent, fidelity, and phase of reversion tracts (collectively, the 'footprint' of restoration). The possibility that these events occur concurrently across large genomic distances and, in some instances, affect both homologues, greatly complicates this endeavor. The sequence data obtained in this study suggest that concurrent reversion events occur in continuous uninterrupted tracts at least for intervals up to 800 bp in length, yet it is difficult to extend this observation to larger genomic regions which have only been profiled at very low

resolution with a small number of molecular markers. The 'holy grail' of this project would be to construct high resolution haplotype profiles that describe the entire diploid genome of one or more revertant individuals. Since an important goal is to deduce the phase of reverted markers, this will require determining haplotypes directly as opposed to querying individual loci. Several approaches to determine long-range DNA haplotypes directly have been described (Douglas *et al.,* 2001; Burgtorf *et al.*, 2003; Mitra *et al.*, 2003; Xiao *et al.*, 2009) but these techniques are technically demanding, generally limited to only dozens of kb; and require *a priori* knowledge of the polymorphisms which are to be queried. Long range single-read DNA sequencing would be an ideal solution, but Sanger-based sequencing technology is limited to roughly 1 kb per read and while 'next generation' sequencing technologies can produce up to 100 Mb of sequence in a single run, this sequence is compiled from hundreds of thousands of individual parallel short reads, each typically being well below 400 bp in length (Schuster, 2008; Shendure and Ji, 2008). Assembling individual sequence reads to reproduce the phase of genomic polymorphisms is essentially an impossible task due to the ambiguity of homologous sequences. It would therefore be of extreme utility if, prior to sequencing, somatic tissue from a revertant individual could be converted from diploid to haploid such that the haploid genomes could be sequenced directly. The *in vitro* conversion of somatic cells from diploid to haploid has been demonstrated in animal cells, where stable hybrids created between human and rodent cells resulted in a proportion of hybrids that contained isolated homologues (Yan *et al.*, 2000). Unfortunately, an analogous method has not been reported in plants. It is possible to generate haploid plant cell lines from the tissue culture of anthers (Nitsch and Nitsch, 1969) and also to generate double-haploid plants through modified chromosome doubling techniques (Wan *et al.*, 1989) but both of these techniques require at least one additional generation, which could potentially compromise the somatic phase of the original line following meiotic recombination events.

Sequence analyses of *cis* chromosomal regions described in this thesis have provided the most detailed assessment of the haplotypes of revertant individuals to date, but only at a small scale. It is clear that

determining long-range high resolution haplotype profiles from revertant individuals presents a great challenge, but this information is essential for providing a detailed picture of how restoration affects the genomes of revertant individuals and, in turn, inferring the nature of the template which is guiding restoration. The work presented in this thesis has provided only a snapshot of how the reappearance of non-parental sequences affects isolated genomic regions, therefore one can only speculate on the nature of the template. Given the evidence presented here two possibilities can be entertained. Either an entire ancestral genome is cached or a subset of the ancestral genome is cached as sequestered templates. Whether the template exists as a small number of large templates or, alternatively, many small templates is not clear. Since the flanking markers characterized in this study appear to revert concurrently, it is most parsimonious to hypothesize that the template exchanges information with the 'target genome' in *continuous tracts*, as opposed to individual polymorphisms reverting independently. It is also apparent that the sizes of these 'restoration tracts' seem to vary, although the upper and lower limits cannot be clearly defined. None of the reversion events characterized in this study had clearly delimited endpoints, yet Ryan Lee observed restoration of a homozygous *C* marker flanked by heterozygous markers (one being 96 bp upstream, and the other being 629 bp downstream) which suggests that restoration tracts can be comparatively small. The concurrent heterozygous restoration events identified by Ryan Lee also appear to indicate the presence of a 'border' between a restoration tract and an unaffected region of the genome (**Figure 4.1**). The identification of such borders may be an exceptionally rare event, particularly if the majority of restoration tracts are large (which would produce fewer borders). Nevertheless, it would be of great interest to identify and study more borders such as this, which would be an important step towards characterizing the genomic footprint of restoration.

4.1.2 Restoration and mosaicism

To this point, I have made the assumption that the non-parental alleles characterized in this project resulted from bona fide restoration events conditioned by a novel genetic process analogous to that

described by Lolle *et al.* (2005a). However, it must be emphasized that the F2 parent 194E9PCL10 was not grown in strict isolation from potentially contaminating pollen sources. This is important because it has been demonstrated that *hth* mutants outcross at frequencies as high as 15%, depending on their proximity to the pollen source (Peng *et al*., 2006; Mercier *et al.*, 2008), although work in our laboratory has shown that outcrossing frequencies do not exceed 2% (S.J. Lolle, personal communication). Still, this is in stark contrast to wild-type *Arabidopsis* plants which are essentially exclusive self-pollinators (Abbott and Gomes, 1989). It has been proposed that the tendency of *hth* plants to outcross may be a consequence of the fused floral morphology which typically results in stigmas emerging from the closed flower prior to anther maturation (Peng *et al.*, 2006). Given the available data, it is impossible to definitively rule out pollen contamination as being the primary source of non-parental alleles in the progeny of 194E9PCL10. However, one important observation supports the hypothesis that the majority of reverted genotypes identified in the F3 progeny are the result of genuine restoration events: in the majority of cases, progeny which where collected from the same 'branch' (inflorescence) of 194E9PCL10 exhibited *identical* genotype profiles (**Table 4.1**). For example, most progeny from branches 1, 2, 11 and 16 were revertant and exhibited congruent profiles. Similarly, branches 3, 5, 7, 8, 9, 12 – 14, 15, and 18 – 20 primarily gave rise to non-revertant progeny. These patterns are inconsistent with outcrossing since it seems unlikely that foreign pollen would saturate some branches while leaving other branches unaffected. These data also resemble the unpublished data of Ryan Lee which identified 7/100 individuals with congruent restoration profiles that, at least at one locus, could not have resulted from outcrossing (since one of the reverted loci was homozygous *C*).

The branch-specific profiles observed in the progeny of 194E9PCL10 lead us to hypothesize that 194E9PCL10 could be a *mosaic* wherein a number of the branches are genetically distinct. If one or more restoration events had occurred in the floral meristem or in the developing inflorescence the end result could be a distinct revertant branch that could give rise to a cohort of progeny with congruent restoration profiles. The possibility that the mutant *hothead* plants used in these studies might be genetic mosaics became clear with the discovery of a chimeric F2 fusion mutant (*hth/hth*) that sported a single phenotypically wild-type inflorescence that was genotypically heterozygous (*HTH/hth*) (Hopkins *et al.*, unpublished data; **Figure 4.2**). This individual plant provided the first compelling evidence that *in vivo* restoration events in an F2 parent could lead to the segregation of revertant F3 plants. While a chimeric plant with such a plainly visible revertant sector seems to occur with extreme rarity (S.J. Lolle, personal communication), root/shoot restoration chimeras identified through molecular genotyping (such as those studied in this thesis) seem to be comparatively common.

The wild-type branch shown in **Figure 4.2** is highly reminiscent of a 'bud sport' – a horticultural term used to describe a phenotypically and genetically distinct branch which is thought to originate from somatic mutations in the meristem (Whitham and Slobodchikoff, 1981). While the underlying mechanism of restoration is clearly distinct from the stochastic process of somatic mutation, it is possible that restoration events arise in single cells and propagate clonally to yield chimeric individuals in the same way that somatic mutations do.

Table 4.1: Many of the progeny of 194E9PCL10 that were derived from the same branch exhibited congruent genotype profiles. Bold text indicates a majority of revertant profiles, while non-bold text indicates a majority of parental profiles. ND, not determined. (Hopkins *et al.*, unpublished data).

Branch	Progeny With Congruent Genotype Profiles
1	**13/24 (54.2%)**
2	**29/32 (90.6%)**
3	9/10 (90%)
4	ND
5	21/22 (95.5%)
7	12/12 (100%)
8	26/26 (100%)
9	27/28 (96.4%)
10	16/21 (76.2%)
11	**3/4 (75.0%)**
12	12/15 (80.0%)
13	17/21 (81.0%)
14	2/2 (100%)
15	15/16 (93.8%)
16	**7/7 (100.0%)**
17	19/20 (95.0%)
18	7/9 (77.8%)
19	7/9 (77.8%)
20	9/10 (90.0%)

The propagation of somatic mutations (and perhaps restoration events) can be understood by considering the developmental mechanisms of plant meristems, which have been deduced largely though clonal analysis of plant chimeras. Clonal analysis involves generating chimeras by phenotypically 'marking' one or more somatic plant cells with a cell-autonomous marker (for example, through mutagenesis) thus facilitating monitoring the lineage of the marked cells as they proliferate (Poethig, 1987). Studies such as these have lead to a good understanding of how distinct plant sectors can arise from a single clonal lineage.

Figure 4.2: The rare observation of a visible revertant sector on this chimeric *hth/hth::HTH/hth* F2 individual strongly supports the hypothesis that *hth* plants are genetic chimeras. Molecular genotyping confirmed that the wild-type inflorescence was *HTH/hth* while the remainder of the plant was *hth/hth* (Hopkins *et al.*, unpublished data).

Angiosperm meristem chimeras can be classified as either *sectorial* chimeras or *periclinal* chimeras. In sectorial chimeras, the marked cells typically exist in a circumferential 'slice' of the meristem, whereas in periclinal chimeras, the marked cells exist in distinct layers of the meristem when viewed in longitudinal section (Leyser and Day, 2003). A sectorial chimera can give rise to a plant with a distinct vertical band of marked cells which are clones of the original and may give rise to additional shoots consisting wholly of marked cells (**Figure 4.3a**). Periclinal chimerism (**Figure 4.3b**) is more complex, since the lineage of plant cells is determined by the layer of the meristem in which they are localized. The

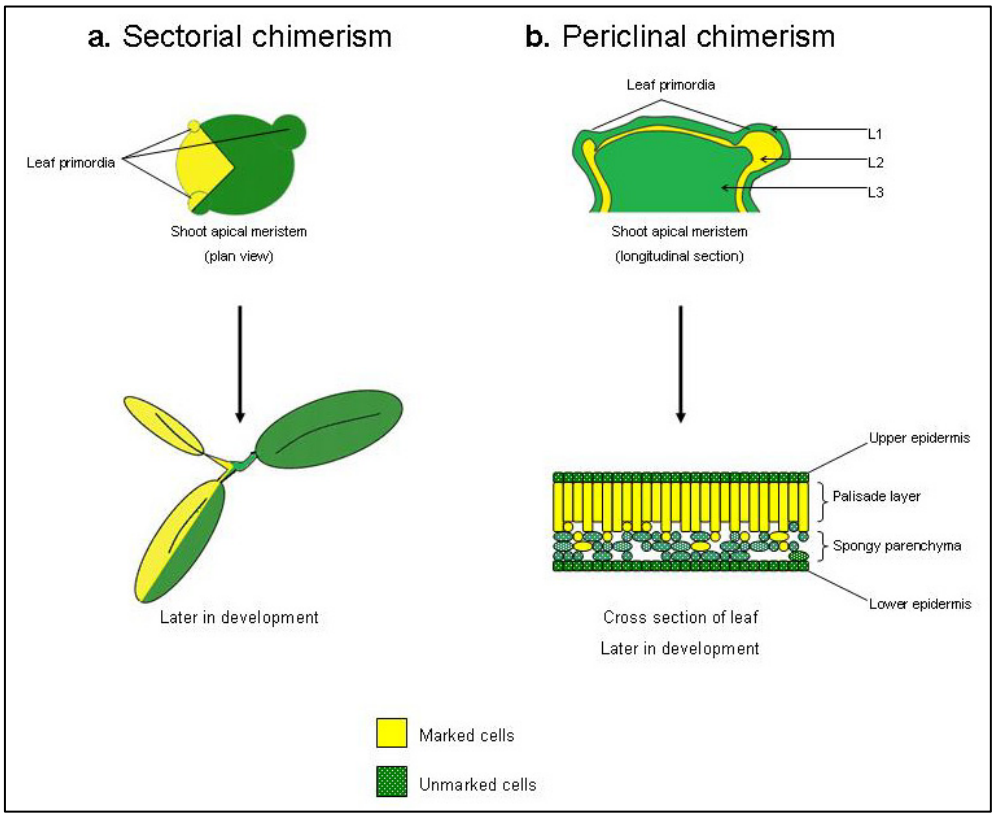


Figure 4.3: Clonal analysis of plant chimeras. **a:** Sectorial chimerism results in a distinct vertical bands of marked cells which propagated clonally form a small number of marked cells at the top of the meristem. **b:** In periclinal chimeras, marked cells can arise in the L1, L2, or L3 layers of the meristem resulting in clonal lineages of marked cells in the tissues derived from that layer.

shoot apical meristems of most dicots consist of three layers when viewed in longitudinal section, each giving rise to distinct tissues. The outer-tunica layer (L1), gives rise to the epidermis, the inner-tunica layer (L2) gives rise to the sub-epidermal cells (including the anthers and ovules), and the cells of the corpus layer (L3) produce more central tissues such as vasculature (Satina *et al.,* 1940; Dermen, 1953; Leyser and Day, 2003). Therefore, if the marked cell of a periclinal chimera lies in the L2 layer, that meristem is likely to yield marked vasculature and gametes. Similarly, if the marked cell lies in the L1 or L3 layer, the fate of the marked cells will be in the epidermis or central tissues, respectively. Periclinal

chimeras have revealed that L2 cells frequently invade the L3 layer, where they adopt the lineage fate of L3 cells (Szymkowiak and Sussex, 1996).

If the assumption is made that the spatial origin and propagation of a *de novo* restoration event is analogous to somatic mutation (i.e. originating in one or a few cells which proliferate clonally), sectorial and periclinal chimerism can help explain how *de novo* restoration events yield the different anatomical patterns of restoration we observe (while saying nothing of the actual restoration *mechanism*). One possible explanation for how the chimeric plant shown in **Figure 4.2** arose would involve a meristematic *hth/hth* → *HTH/hth* restoration event that produced a sectorial chimera. The resulting circumferential sector of *HTH/hth* tissue could give rise to a 'revertant bud sport' consisting entirely of *HTH/hth* cells that are clones of the original revertant cell. If this is the case, it would be expected that the revertant flowers will have the wild-type floral phenotype, produce *HTH* gametes, and segregate *HTH* in a 1:2:1 Mendelian fashion. Unfortunately, the plant in **Figure 4.2** was unintentionally uprooted and therefore progeny could not be collected for segregation analysis.

It is also possible that a large, phenotypically distinct sector could arise from a periclinally chimeric restoration event. Since organ fusion is an epidermis-specific phenomenon (Lolle *et al.*, 1998), there exists the fascinating possibility that the sector in **Figure 4.2** originated from a *hth/hth* → *HTH/hth* restoration event confined to the L1 layer of the meristem, such that the resulting inflorescence would have primarily wild-type epidermal cells. This could correct the floral fusion phenotype despite the fact that the sub-epidermal tissues, including the gametes, would remain *hth/hth*. The transposon-mediated reversion of *fiddlehead* (*fdh*) fusion mutants results in wild-type flowers which rarely segregate wild-type progeny – an observation which has been attributed to L1-specific reversion events (Pruitt *et al.*, 2000). The hypothesis that the plant shown in **Figure 4.2** may be 'a sheep in a wolf's clothing' (i.e. a phenotypic revertant that does not segregate revertant progeny) could be tested by monitoring the segregation of the progeny of the wild-type branch, or by comparing the genotypes of cells of the vasculature and the

gametes to those of the epidermis. The latter approach could be achieved with laser-capture microdissection, a technique which allows individual cells to be isolated from tissue sections (Kerk *et al.*, 2003).

Interestingly, no phenotypically wild-type branches were identified on the 194E9PCL10 parent plant, despite the fact that the majority of its F3 progeny that exhibited non-parental indel loci were also phenotypically wild-type for floral fusion (Hopkins *et al.*, unpublished data). How might phenotypically mutant branches give rise wild-type progeny? One possibility is that some 194E9PCL10 branches were periclinal chimeras in which the reversion events were confined to the L2 layer. The branch, however, might have remained phenotypically mutant because the epidermis retained the *hth* genotype even though the sub-epidermal cells (including the gametes) were revertant (effectively making 194E9PCL10 'a wolf in a sheep's clothing'). However, if the sector was large and relatively stable, an important prediction is that the chimeric branches should manifest Mendelian segregation for any the revertant alleles, whether at an indel marker or the *HTH* locus. Such segregation, however, was *not* observed. For example, progeny derived from branch 2 were heterozygous at all indel loci tested, and uniformly exhibited wild-type flowers. The expected 3:1 phenotypic ratio of wild-type:fused flowers was not observed (Hopkins *et al.*, unpublished data).

The absence of Mendelian segregation is perplexing. These results could be explained by the introduction of exogenous pollen however, as discussed earlier, the branch-specific reversion patterns observed in the progeny of 194E9PCL10 are not easily explained by pollen contamination. If we are to maintain that this parent was a genuine mosaic of restoration events, it would appear that: (1) mechanisms exist which diminish the probability that revertant gametes will combine in a Mendelian fashion and/or (2) restoration events propagate by a mechanism that is fundamentally distinct from that of somatic mutations. For example, it is possible that the persistence of heterozygosity could reflect a persistence of genome flux, resulting in the generation of multiple sectors within one branch, each having distinct

genotypes. If multiple small sectors populate a single branch, the probability of producing two gametes with an identical haplotype is greatly diminished. Alternatively, it may be that the template is sequestered in a *latent* state in *all* cells, but is only activated in some. The presence of an activated template in sectors of the plant could result in branches that are not genetically distinct, but are *distinct in their potential to segregate revertant progeny*. One could then speculate that the activated template is inherited with the gametes and exchanges its information with the genome of the F3's during early embryogenesis. While highly speculative, these mechanisms could possibly explain why an apparent mosaic such as 194E9PCL10 does not segregate restored alleles in a Mendelian fashion.

In considering the results discussed above, an argument can be made that the non-parental alleles in the progeny of 194E9PCL10 could have been introduced through cross-pollination events. Although this seems unlikely, this possibility cannot be dismissed outright. However, the results obtained from the analyses of individual seedlings in which the genotype of the shoot and root was found to differ offer the most compelling evidence that *Arabidopsis* plants harbor a cryptic source of extra-genomic sequence information that can be inherited in a non-mendelian fashion. Nevertheless, sampling error could provide a trivial explanation for the identification of root and shoot samples with distinct genotypes. With this in mind, the experiment which initially identified chimeric progeny in the PL113H12C2_11_12 line was repeated independently by the author using the same batch of seed. A nearly identical frequency of root/shoot chimerics were identified among 100 of the progeny, supporting the hypothesis that the restoration events are genuine.

The results demonstrating that F3 progeny display distinct genetic profiles between organ systems implies that restoration is not limited to the F2 generation but occur can occur *de novo* in each generation. Is seems likely that genetically uniform (non-chimeric) revertant F3 individuals result from restoration events that took place in the parent plant. The idea that F3 chimerism results from *de novo* restoration events implies that the template driving these sequence modifications can be inherited through multiple

generations, which is consistent with the observation that instability at the *hth* locus can persist through multiple generations (Lolle *et al.*, 2005a).

It would be of interest to compare the frequencies of occurrence of chimeric and non-chimeric F3 revertants, although PCR genotyping alone makes distinguishing the two somewhat challenging. Using PCR and traditional sampling methods, it is not possible to distinguish tissue composed of genetically uniform cells that are all heterozygous from tissue composed of two genetically distinct cell populations (where the one population of cells has the parental genotype and the other the non-parental genotype). The presence of a small number of revertant cells would 'mask' the non-revertant genotype of its neighbors. Conversely, if the genotype of a tissue sample is determined to be non-revertant, it *cannot* be concluded that the plant is not a genetic mosaic unless the *entire plant* is genotyped.

Ascribing a genotype to an F2 parent based on extrapolating PCR results obtained from single tissue sample becomes problematic if the individual is actually a genetic mosaic. This raises the point that a missing piece of the work described in this thesis is the genotyping of tissue derived from each of the 20 branches of 194E9PCL10, which would have helped answer the question of whether or not the branches truly were genetically distinct. Consequently, our laboratory has transitioned from 'traditional' genotyping (in which the results of a single tissue sample are extrapolated to the entire individual) towards 'localized' genotyping (i.e. genotyping numerous cauline leaves and flower parts from a single parent). This work has provided further support for the idea that at least some F2 parents are genetic mosaics (Hopkins and Lolle, unpublished data). It will be of great interest to see if this mosaicism translates into the segregation of revertant progeny. Another useful approach to studying chimerism would be to monitor a cytological marker for restoration, such as green fluorescent protein. This would allow the direct visualization of sequence restoration and facilitate distinguishing restoration events among chimeric cell populations. This proposed study relies on the assumption that transgenic loci restore in the same fashion that native alleles do. Initial work described in this thesis may have provided supporting evidence that recently introduced

transgenic constructs can restore at least on a cell-by-cell basis thus producing chimeric tissues (see section *4.2* below).

4.1.3 Is restoration unique to *Arabidopsis*?

A pertinent question is how broadly (or narrowly) distributed restoration is throughout the tree of life (Lolle *et al.*, 2005a). If restoration is a fundamental genetic process, it would be somewhat surprising if this novel process is found to be limited to *Arabidopsis*, but furthermore to *hothead* mutants. This thesis described the characterization of restored insertion sequences in 2/100 F3 progeny of the wild-type hybrid CL11B1. To the best of our knowledge, this line had never encountered a mutant *hth* allele in its ancestry. These findings suggest that the restoration phenomenon can occur in the absence of a mutant *HTH* gene. The frequency of restoration observed in the progeny of CL11B1 was low relative to the frequencies that have been observed in the *hth* mutant background (for example, approximately 50% of the F3 progeny of the *hth-3* mutant PL113H12C2_11_12 exhibited an indel restoration event). The observation of restoration in wild-type plants is consistent with the hypothesis of Lolle *et al.* (2005a) that all *Arabidopsis* plants may have the capacity to exhibit restoration, but that the *hth* mutation somehow conditions restoration to occur at higher and therefore readily measureable frequencies.

A number of atypical inheritance patterns have been reported in plants other than *Arabidopsis*. Examples include the appearance of cultivar-specific RFLP markers in tissue cultures of soybean cultivars that lacked the marker (Roth *et al.*, 1989) and environmentally-induced genome changes in *Linum* sp. which include the high frequency appearance of a novel insertion (Schneeberger and Cullis, 1992; Chen *et al*., 2009). However, it is difficult to confidently draw parallels between restoration in *Arabidopsis* and these studies for several reasons. First, these studies document the appearance of cultivar-specific markers that were not known to have existed in the ancestry of the inbred lines that were studied. Second, the appearance of RFLP markers documented by Roth *et al.* (1989) and changes in nuclear DNA content and repetitive regions of the genome (Schneeberger and Cullis, 1992) were not accompanied by a detailed

molecular analysis, making it difficult to determine the specificity of these changes at the DNA level. Finally, while Chen *et al.* (2009) have documented in detail the highly reproducible *in vivo* mosaic appearance of a novel insertion, it is not clear whether this mechanism is limited specifically to the insertion studied, or is a more generalized genome-wide phenomenon.

Strikingly, the enigmatic genetic phenomena which most closely resemble restoration in *Arabidopsis* have surfaced over the last 15 years from studies of genetic disorders in humans. Patients afflicted with a wide variety of recessive disorders have exhibited what has been described as *reverse mosaicism* – somatic mosaicism resulting from reversion to normal of an inherited mutation (Hirschhorn, 2003). Two representative studies that were introduced in Chapter 1 of this thesis will be described here in greater detail such that similarities and differences to restoration in *Arabidopsis* can be highlighted and possible mechanisms discussed.

Mosaic patterns of genetic reversion from mutant to wild-type in liver samples of Tyrosinemia Type I patients

The autosomal recessive disease Tyrosinemia Type I is characterized by the loss of function of fumarylacetoacetase (FAH) which is a key enzyme in tyrosine degradation (Kvittingen *et al.*, 1994). Individuals with this disease typically exhibit severe liver damage due to the accumulation of toxic metabolites and pronounced cell regeneration at sites of damage (Lindblad *et al.*, 1977). A study by the group of Kvittengen *et al*. (1993) identified a group of five Norwegian Tyrosinemia patients who exhibited a mosaic pattern of FAH activity in liver samples subjected to immunological staining, suggesting that FAH activity had been restored in some regions of liver tissue. The Kvittengen group expanded on this work in a later study which identified similar patterns of mosaic FAH activity in liver tissue samples in 15 of the 19 patients investigated (Kvittingen *et al.*, 1994). In the 1994 study, however, the immunochemical analysis was supplemented with a molecular genetic analysis of mosaic liver tissue samples from four patients (two homozygotes and two compound heterozygotes), who collectively

exhibited 5 different Tyrosinemia-causing mutant alleles. The molecular analysis revealed that, for all four patients, the restoration of FAH activity in the immunopositive regions of the liver samples was due to heterozygous genetic reversion to wild-type. Among the four patients, reversion was observed at three different alleles, each of which exhibited reversion of the mutant AT nucleotide pair to the wild-type GC.

The Kvittingen group explored several possible explanations in an attempt to understand their perplexing results. For the compound heterozygous patients, it seemed reasonable to propose that a gene conversion event or mitotic recombination could explain the reversion, wherein a section of one mutant allele provided the wild-type template for the other. However, this explanation fails to account for the reversion which was observed in the homozygous individuals, where no wild-type template was available. Furthermore, there are no known FAH pseudogenes which could have supported the gene conversion hypothesis. Another proposed explanation was that the accumulation of alkylating metabolites due to the absence of FAH activity could have induced the reversion, although the alkylating metabolites in question have not been implicated in mutagenesis (Kvittingen *et al.*, 1994). Nevertheless, these metabolites are known to cause significant damage in the liver of Tyrosinemia patients, leading to increased cell division and often the development of cancer (Lindblad *et al.*, 1997). The authors therefore proposed that since rapidly replicating cells are more prone to mutations, wild-type revertant cells could arise at low frequencies. Since these revertant cells would have a selective advantage, they could proliferate to the point that they are detectable by clinical methods (Kvittingen *et al*., 1994). An alternative explanation (which was not explicitly discussed by the authors) is the possibility that only one parent transmitted a mutant allele to the patient and the wild-type allele was subjected to a *de novo* somatic mutation early in development, thus giving rise to a mosaic individual. This possibility cannot be wholly discounted since the transmission of the mutant alleles from parent to patient was not specifically demonstrated. Nevertheless, the likelihood that *de novo* mutations arose in all four patients resulting in mutant alleles identical to those present in the parents is extremely unlikely, thus supporting the argument for genuine

reversion. It is known from studies in humans that cells of maternal origin can persist in progeny well into adult life (Maloney *et al.*, 1999) resulting in what has been termed 'microchimerism'. The possibility that reversion may have been an artifact of maternal cell colonization was not specifically addressed by Kvittengen *et al.* (1994). Nevertheless, another study focusing on reverse mosaicism in Tyrosinemia patients specifically ruled out maternal cell colonization as the source of wild-type alleles in revertant liver samples (Bergeron *et al.*, 2004).

***In-vivo* mosaic reversion from mutant to wild-type in an adenosine deaminase deficient patient**

Adenosine deaminase (ADA) deficiency results from the complete loss of activity of the purine salvage enzyme ADA due to homozygosity for certain mutant alleles at the *ADA* locus. ADA deficiency leads to the neonatal onset of severe combined immunodeficiency (SCID), a disorder which is lethal prior to two years of age in the absence of therapy (Hirschhorn *et al.*, 1996). In a study by Hirschhorn *et al.* (1996), a compound heterozygous patient was initially identified as having complete ADA deficiency, yet he gradually exhibited a progressively milder phenotype and, without therapy, was healthy at the age 12 (at the time of publication). A detailed molecular analysis revealed that this individual was a somatic mosaic with residual ADA activity, and that the mosaicism resulted from a spontaneous *in vivo* reversion to wild-type of the maternally transmitted mutant allele (Hirschhorn *et al.,* 1996).

In this elegant study, contrasting levels of ADA activity were initially identified in patient samples. Lymphoid cell lines displayed residual (15%) ADA activity and lower levels of metabolite accumulation, both of which were comparable to the levels observed in the heterozygous parental cell lines. Erythrocytes (RBCs) however, had no detectable activity in the patient, whereas both of the heterozygous parents exhibited approximately 50% of normal RBC ADA levels. A comprehensive molecular analysis of cDNA clones from the lymphoid cell line revealed that the maternally transmitted mutant allele was virtually absent, apparently having reverted to the wild-type. The maternal mutant allele was, however, detectable in the patients' RBCs, and the paternally transmitted mutant allele was present in both cell lines.

An important aspect of this study was the use of RFLP's to authenticate the lymphoid cell lines and to eliminate the possibility of contamination by paternal cells. Most importantly, a 'private' maternal intragenic *ADA* RFLP unique to the family was used to confirm that the mutant allele was indeed inherited from the mother since this RFLP was identified in the revertant lymphoid cell line, as well as the non-revertant RBCs.

To test the possibility that the reversion could have occurred *in vitro* during cell culture manipulations, clones derived from patient peripheral blood samples were analyzed for the presence of both the paternal and maternal mutations. Of the 89 clones investigated, 17% carried neither mutation, demonstrating that reversion had occurred *in vivo*.

As was the case with the study by Kvittingen *et al.* (1994), the Hirschhorn group struggled to provide an explanation for their novel results. Since the patient studied was a compound heterozygote, the possibility of gene conversion exists, but this was dismissed as extremely unlikely since only 12 bp separated the maternal and paternal mutations. Hirschhorn *et al.* noted that, like Tyrosinemia type I, ADA deficiency leads to rapidly dividing cells at sites of damage and regeneration and the accumulation of harmful metabolites, one of which (deoxyadenosine) is known to induce DNA strand breaks *in vitro* (Hirschhorn *et al.*, 1996, and references within). As was proposed for FAH revertants, the presence of these conditions could provide selective pressure that might favor the proliferation of cells harboring *de novo* mutations that restore wild-type ADA function. Despite the absence of a demonstrated increased rate of mutation in damaged cells of ADA deficient individuals, this hypothesis was again favored as the most likely underlying mechanism for this fascinating example of 'autonomous gene therapy'.

These studies of reverse mosaicism exhibit both similarities and differences to restoration in *Arabidopsis*. First of all, the pedigrees in which these reversion events are observed in humans and *Arabidopsis* are the same (i.e. the homozygous mutant progeny of two heterozygotes exhibits mosaic functional and genetic reversion to the wild-type allele). There is also a clear similarity between the

reverse mosaicism of homozygous alleles in human patients and restoration of the *HTH* allele, particularly the reports of homozygous mutant tyrosinemia patients reverting to the exact wild-type allele (Kvittingen *et al.*, 1994). In both the Kvittengen *et al.* (1994) and Hirschhorn *et al.* (1996) studies, a toxicity induced mutation-selection model was proposed to explain reversion to a wild-type allele. A nearly identical model was proposed to explain reversion of the *hth* allele (Comai and Cartwright, 2005) soon after the results of Lolle *et al.* (2005a) emerged. These models, however, cannot account for the more recent work in our laboratory (restoration of L*er*/Col polymorphisms) for two important reasons: (1) These models rely heavily on revertant cells having a selective advantage yet there is no evidence that any of the L*er*/Col polymorphisms characterized in this thesis have any functional importance which would facilitate their selection following reversion; (2) These models essentially describe a stochastic process based on random mutation which could account for single base pair changes, but *cannot* account for the reappearance of extensive sequence tracts.

A recent genetic study on patients suffering abdominal aortic aneurysms has revealed that even non-diseased patients harbor allele variants that were present in some tissues but absent in others (Gottlieb et al 2009). This provides a striking example that allelic variants of unknown origin can lead to mosaicism in humans. This is reminiscent of the cryptic mosaicism we have observed in *Arabidopsis* plants that is only detectable by molecular analysis.

An important distinction between reverse mosaicism seen in humans and restoration in *Arabidopsis* is that reversion in humans has only been shown to restore function to loss of function mutations in a highly site-specific manner. The most striking example of the specific nature of reversion in humans is demonstrated in the study of *ADA* reversion by Hirschhorn *et al.* (1996) where a flanking maternal RFLP marker and several other genomic RFLPs appeared to be unaffected by the nearby reversion of the maternally inherited mutation. This finding is reminiscent of the unpublished work of Ryan Lee in which a *trans* restoration event did not affect closely flanking loci, but is in contrast to the restoration events

described in this thesis. Although the initial observation leading to the discovery of restoration in *Arabidopsis* was based on the recovery of the wild-type function of the *hth* mutation (therefore also appearing to be highly specific and directional) the work described in this thesis suggests that restoration occurs in tracts and that the targets of restoration are not limited to loss of function mutations. Until more work is conducted in larger populations of humans to determine if other more distal flanking loci or unlinked loci (including indels) revert concurrently with disease loci, it is difficult to draw further analogies to the events observed in *Arabidopsis*. Nevertheless, there exists the fascinating possibility that a fundamentally similar yet elusive mechanism is governing these processes across kingdoms.

4.2 Segregation and stability of the *PAP1-D* transgene and *HTH* allele

The primary goal of the work described in Chapter 3 of this thesis was to determine how a transgenic marker (*PAP1-D*) and native alleles (*HTH* and in some cases *ER*) segregate in a variety of distinct ancestries. These experiments attempted to shed light on the types of sequences that could become unstable, the role of *hth* in conditioning restoration, and the ancestral requirements of restoration. To achieve this goal, pedigrees compromising a number of distinct ancestries were constructed in both the presence and absence of two mutant *hth* alleles, and the segregation of *PAP1-D*, *HTH,* and *ER* monitored. The results of this project can be summarized as follows: (1) Based on phenotypic analysis, the segregation of non-parental alleles was not observed in any of the 119 F2 (Tier 4) lineages that were studied, irrespective of ancestry or the presence of a mutant *hth* allele; (2) PCR-based screening suggested that restoration of the *PAP1-D* allele may have occurred in a subpopulation of cells in two F3 (Tier 5) individuals, but these results were inconclusive.

In the paper published by Lolle *et al*., the frequency at which *HTH* revertants occurred fell within the range of 10^{-1} to 10^{-2} F3 revertants per chromosome per generation (Lolle *et al.*, 2005a). The stable inheritance observed at the *hth*, *PAP1*, and *er* loci as reported in Chapter 3 of this thesis seems at first inconsistent with the results of Lolle *et al.* (2005a). Furthermore, attempts by other investigators to

reproduce the high frequency of genetic instability reported at *hth* have been unsuccessful (Peng *et al.*, 2006; Mercier *et al.*, 2008). What proportion of *hth/hth* F2's segregate non-parental alleles? This is a critical yet universally overlooked question in all published discussions of restoration to date. Based on work conducted in our lab, genetic instability at *hth* was only observed in approximately 1/300 (0.3%) of F2 lines (S.J. Lolle, personal communication). Therefore, *HTH* restoration appears to occur at low frequencies within populations and is not manifested by every individual within that population. However, if multiple loci are monitored using molecular screens, most *hth/hth* individuals exhibit genetic instability of at least one indel marker in the F3 generation, even in the absence of *HTH* restoration (S.J. Lolle, personal communication). Knowing that the vast majority of F2 individuals do not exhibit genetic instability at *hth* has profound implications when studying restoration and brings to attention limitations in the design of the experiments described here.

When the pedigrees described in Chapter 3 of this thesis were designed, the studies aimed at determining the relative susceptibility of *hth* mutants to outcrossing and, by extension, revealing the actual frequency of instability among *hth/hth* F2's had not yet been completed. As a result, the number of F2 (Tier 4) lines used was suboptimal. Instead, this experiment was specifically designed to generate the maximum number of distinct ancestries and explore their affect on restoration frequencies. Therefore, the fact that *HTH* restoration was not observed among the experimental population could simply reflect an insufficient sample size. Of the 119 F2 (Tier 4) parents that were generated, 88 were *hth/hth* (the other 31 being *HTH/HTH*). If only 0.3% of *hth/hth* F2's exhibit restoration of the *HTH* allele, this sample population was probably too small to observe *HTH* restoration in even one lineage with high confidence. Furthermore, many of these 88 F2's were derived from ancestries in which *HTH* restoration has not been documented. This may have further decreased the likelihood of observing restoration. The low frequency of genetic instability within *hth/hth* populations may also serve to explain why independent studies have not been able to reproduce the results of Lolle *et al.* (2005a). Peng *et al.* (2006) did not specifically

discuss the number of F2 parents while Mercier *et al.* (2008) investigated the progeny of only 41 isolated *hth/hth* F2 parents.

The relative ease with which indel restoration events can be identified in the progeny of F2 *hth/hth* mutants is in stark contrast to the low frequency of F2's that segregate wild-type *hth* plants. This could simply be a result of the small sample size of alleles interrogated. This study focused on three specific alleles (*HTH*, *PAP1*, and *ER*), whereas the molecular screens for indel reversion interrogate 16 different alleles. It is reasonable to assume that studying a larger number of loci will increase the probability of observing a restoration event. It is also possible that certain loci or genomic regions are more recalcitrant to reversion that others.

Alternatively, there could be an inherent bias when screening for phenotypic reversion. As was discussed in section *4.1.2*, restoration appears to occur in mosaic patterns, which could shed some light on the apparent disparity between the frequency of indel restoration events (which are screened with PCR) and restoration events that are only screened with phenotypic analysis. Any localized restoration event could be detected by PCR, but complete phenotypic restoration of *HTH*, *ER*, or *PAP1-D* would depend on the restoration event propagating to yield a phenotype. It is therefore possible that subpopulations of cells did experience restoration of *HTH, ER,* and *PAP1-D* in these experiments, but that these events never gave rise to a detectable phenotype. Therefore, restoration would only have been detectable through a comprehensive molecular analysis of multiple tissue samples.

Although molecular screening for *HTH* and *ER* was not conducted, 23 Tier 5 individuals from the *hth-4** pedigree were screened for the presence of the *PAP1-D* allele using PCR. Two individuals (*hth4**.T5'.$\beta^{a'}$.5006^{4143} and *hth4**.T5'.$\delta^{a'}$.5036^{4097}) gave a positive result with the BAR61L/374R primer set, indicating the presence of the *PAP1-D* allele. In one case, the molecular data were further corroborated by the presence of a distinct purple sector on the rosette leaf sampled. However, these results remain inconclusive for two reasons. First, the PCR products amplified from these samples were very

faint and re-amplification was unsuccessful. Second, wild-type L*er* and Col seedlings screened using the same *BAR* gene primer set gave one false positive result. Therefore PCR contamination seems one plausible explanation for these results. The source of this contamination, however, is not clear. PCR contamination could of course also offer a trivial explanation for the appearance of non-parental indel markers in the studies described earlier, but given that these results are consistently reproducible, and always include negative controls in which no DNA template is added, this seems unlikely.

Determining if the transgenic marker *PAP1-D* would exhibit genetic instability was a primary goal of the experiments described in this thesis. Complete phenotypic *PAP1-D* restoration was not observed, but it cannot be ascertained if this was due to an insufficient sample size, or because this genomic locus or this specific construct is recalcitrant to restoration, or simply because recently introduced transgenic markers do not restore. Therefore, the question of whether or not transgenic markers can restore remains open.

Future work to test for restoration of specific alleles such as *HTH* and *PAP1-D* must take into account the low frequency at which restoration appears within populations. Therefore future experiments should focus on a single pedigree that is known to manifest genetic instability (such as that described by Lolle *et al.*, 2005a), with an F1 that is also heterozygous for a transgenic marker. If hundreds of replicate F2 lines are produced, this should increase the probability that restoration of *HTH* will be observed in the F3. Consistently observing phenotypic restoration at one or more native mutant alleles (such as *HTH* or *ER*) is an essential experimental control in an experiment which seeks to determine if phenotypic restoration can also occur for a recently introduced transgenic marker. If restoration of the transgenic marker is not observed, the restoration of *HTH* or *ER* would provide the necessary positive control to draw a meaningful conclusion from a negative result.

Testing a transgenic marker for tissue-specific restoration events (i.e. mosaicism) could be conducted in the same experiment proposed above, possibly with PCR based genotyping. However, to eliminate the

possibility of sampling errors and contamination (which would become increasingly relevant at such a large scale experiment) it would be desirable to test a transgenic marker with a cytological phenotype (such as green fluorescent protein) such that restoration events could be monitored directly at the cellular level, without need for manipulation of sample extracts.

4.3 Possible mechanisms of restoration

The work presented here and parallel studies of restoration conducted in our laboratory have yielded the following key observations: (1) Restoration results in the reappearance of DNA sequences which are identical to those of a recent ancestor; (2) Restoration appears to affect both nearby and distant genomic regions concurrently; (3) Restoration can occur in *cis* and in *trans* (4) Restoration is frequently localized within single plants, giving rise to sectored or chimeric individuals; (5) Restoration events are more readily revealed by PCR-based genotyping than phenotypic screening.

These observations contribute primarily to our descriptive understanding of restoration without providing any *direct* insight into the underlying mechanism. Nevertheless, these observations provide clues of the underlying mechanism which allow us to rule out a number of possible explanations. It has already been discussed that the trivial explanations of pollen contamination, sampling error, and PCR-contamination are unlikely to account for all cases of restoration. Stochastic mechanisms such as random somatic mutation and the mutation-selection hypothesis suggested by Comai and Cartwright (2005) have also been ruled out in favor of a highly specific template-directed process. The template appears to be ancestral in origin, contains tracts of the ancestral genome that may be upward of 80 kb in length, is heritable through multiple generations, and is capable of introducing non-parental genetic information into the genome.

Several hypotheses have been proposed in attempts to describe the nature of the template. Chaudhury (2005) suggested that the template for reversion could be present *within the genome* as short (13-18 bp) sequences homologous to the wild-type *HTH* allele which were detected in BLAST searches of the L*er*

database. Short RNA transcripts of these homologous sequences could be used to convert the mutant alleles to wild-type by a mismatch-repair mechanism. However, the existence of such short homologous sequences could simply be a chance occurrence (Lolle *et al.* 2005b). More importantly, similar homologous sequences which could provide the templates for the extensive indel restoration events described in this thesis are not present in either the L*er* or Col genomes, as revealed by BLAST searches.

Other hypotheses have suggested the existence of an extra-genomic template. Ray (2005) proposed a DNA-based template model in which the template is an 'archival form' of supernumary chromosome fragments ("genome trash") which were preserved in the germline of the F1 heterozygote. It was speculated that the way in which these fragments are archived (i.e. heterochromatinized and covalently modified) could render them silent and possibly even inaccessible to PCR and DNA hybridization techniques, yet they could nevertheless guide restoration by a gene-conversion mechanism in subsequent generations. Furthermore, these archived fragments may only be present in a small number of meristematic cells which would further complicate their detection in a typical genotyping experiment. The model proposed by Ray is consistent with our observations of extensive and precise tracts of restored F1 sequences as well as chimeric restoration events (which could be explained as the detection of supernumary chromosomal fragments or *de novo* conversion events). This model is based on the demonstrated existence of supernumery chromosomal fragments in plants (Ray 2005 and references within) and the well established mechanism of gene conversion. The heritability of such fragments, however, implies a novel and currently undemonstrated mechanism. Furthermore, the restriction of these templates to a small number of meristematic cells is inconsistent with the clonal nature of plant development. Lolle *et al.* (2005a) proposed a mechanism in which the template exists as a stable and heritable 'cache' of RNA which could possibly contain the entire genomic sequence of the F1. This RNA cache would likely be double-stranded, can self-replicate, and could guide restoration in subsequent generations. While the heritability of RNA has been demonstrated (Fire *et al.*, 1998; Rassoulzadegan *et*

al., 2006) as has the capacity of RNA to guide genome rearrangements (Nowacki *et al*, 2008), preliminary work in our laboratory has not yet been able to directly detect cached RNA sequences in *hth* mutants (Chang and Lolle, unpublished results). Therefore, the gene conversion-based supernumery chromosome fragment hypothesis proposed by Ray (2005) and the RNA cache hypothesis proposed by Lolle *et al.* (2005a) remain as highly novel, yet currently undemonstrated mechanisms to explain restoration. It will be interesting to see if future studies of restoration can provide direct support for either of these mechanisms.

Conclusions

The reappearance of non-parental loci in F3 *Arabidopsis* plants results in the reinstatement of tracts of sequence at least 80 kb in length, appear to occur predominantly in *cis*, and are identical to the corresponding sequences that were present in a recent ancestor (the F1 hybrid). The restoration tracts analyzed in this thesis are indistinguishable from outcrossing events at the genetic level. This, combined with the increased rate of outcrossing which has been demonstrated in *hth* fusion mutants (Peng *et al*., 2006; Mercier *et al.*, 2008), presents some unique challenges when studying non-parental sequences. In many of the experiments detailed here, pollen contamination cannot be absolutely ruled out as the source of the non-parental loci. Nevertheless, the apparent branch-specific nature in which revertant progeny arose from the 194E9PCL10 F2 parent leads us to hypothesize that this individual was a mosaic, wherein many of the branches were either distinct genetically, or distinct in their capacity to produce revertant progeny. The appearance of non-parental loci localized to either the root or the shoot of F3 progeny of PL113H12C2_11_12 and CL11B1 provided clear examples where outcrossing could not account for the appearance of all non-parental sequences. The observation of mosaics provides the strongest evidence to date that restoration is a genuine and novel biological phenomenon. Since the exact reappearance of large insertions cannot be explained by a stochastic mechanism, we have concluded that restoration is a template-directed process. The nature of the template remains elusive, yet it is clear that it contains large tracts of an ancestral genome, is stably inherited through multiple generations, and is not detectable in typical experiments.

Monitoring the segregation of the *HTH, ER,* and *PAP1-D* alleles in both the presence and absence of the *hth-4* and *hth-8* mutations revealed that all loci were inherited stably. This leads us to conclude that, at a population level, the vast majority of individuals do not exhibit phenotypic restoration at any given allele. Future studies testing for the restoration of a transgenic marker such as *PAP1-D* should aim to

produce a population of at least 300 F2 parents that are all derived from the same pedigree in which Lolle *et al.* (2005a) reported restoration at *hth*. The apparent disparity between restoration that yields a phenotype and restoration observed at indel markers may indicate that the majority of restoration events are tissue-specific cryptic events that are best detected through molecular screening.

References

Abbott, R.J., and Gomes, M.F. 1989. Population Genetic-Structure and Outcrossing Rate of *Arabidopsis-thaliana* (L) Heynh. Heredity **62**: pp. 411-418.

Ariga, T., Kondoh, T., Yamaguchi, K., Yamada, M., Sasaki, S., Nelson, D.L., Ikeda, H., Kobayashi, K., Moriuchi, H., and Sakiyama, Y. 2001. Spontaneous *In Vivo* Reversion of an Inherited Mutation in the Wiskott-Aldrich Syndrome. The Jounral of Immunology **166:** pp. 5245-5249.

Ariga, T., Yamada, M., Sakiyama, Y., and Tatsuzawa, O. 1998. A Case of Wiskott-Aldrich Syndrome With Dual Mutations in Exon 10 of the WASP Gene: An Additional *De Novo* One-Base Insertion, Which Restores Frame Shift Due to an Inherent One-Base Deletion, Detected in the Major Population of the Patient's Peripheral Blood Lymphocytes. Blood **92:** pp. 699-701.

Bergeron, A., Lettre, F., Russo, P., Morisette, J., and Tanguay, R.M. 2004. No Evidence of Maternal Cell Colonization in Reverted Liver Nodules of Tyrosinemia Type I Patients. Gastroenterology **127:** pp. 1381-1385.

Borevitz, J.O., Xia, Y., Blount, J., Dixon, R.A., and Lamb, C. 2000. Activation Tagging Identifies a Conserved MYB Regulator of Phenylpropanoid Biosynthesis. The Plant Cell **12:** pp. 2383-2394.

Bruce, G., Lorraine, E.C., Benjamin, M., Nathan, S., Daniel, O., Cherrie, A., Melissa, M., Tomoko, S., Ghassan, B., and Morris, S. 2009. *BAK1* gene variation and abdominal aortic aneurysms. Human Mutation **30:** pp. 1043-1047.

Burgtorf, C., Kepper, P., Hoehe, M., Schmitt, C., Reinhardt, R., Lehrach, H., and Sauer, S. 2003. Clone-Based Systematic Haplotyping (CSH): A Procedure for Physical Haplotyping of Whole Genomes. Genome Research **13:** pp. 2717-2724.

Buss, L.W. 1983. Evolution, development, and the units of selection. Proceedings of the National Academy of Sciences **80:** pp. 1387-1391.

Chaudhury, A. 2005. Plant genetics: *Hothead* healer and extragenomic information. Nature **437** (7055): pp. E1-E1.

Chen Y., L.R., Cullis C.A. 2009. An environmentally induced adaptive (?) insertion event in flax. International Journal of Genetics and Molecular Biology **1**(3): pp. 038-047.

Comai, L., and Cartwright, R.A. 2005. A Toxic Mutator and Selection Alternative to the Non-Mendelian RNA Cache Hypothesis for *hothead* Reversion. The Plant Cell **17:** pp. 2856-2858.

Cullis, C.A. 1973. DNA Differences between Flax Genotrophs. Nature **243**(5409): pp. 515-516.

Cullis, C.A. 1981. Environmental induction of heritable changes in flax: defined environments inducing changes in rDNA and peroxidase isozyme band pattern. Heredity **47**(1): pp. 87-94.

Cullis, C.A., Swami, S., and Song, Y. 1999. RAPD polymorphisms detected among the flax genotrophs. Plant Molecular Biology **41**(6): pp. 795-800.

Dermen, H. 1953. Periclinal cytochimeras and origin of tissues in stem and leaf of peach. American Journal of Botany **40**(3): pp. 154-168.

Douglas, J.A., Boehnke, M., Gillanders, E., Trent, J.M., and Gruber, S.B. 2001. Experimentally-derived haplotypes substantially increase the efficiency of linkage disequilibrium studies. Nature Genetics **28**(4): pp. 361-364.

Dreury, C., and Bateson, W. 1901. Experiments in plant hybridization. Journal of the Royal Horticultural Society **26**: pp. 1-32.

Durrant, A. 1962. The environmental induction of heritable change in *Linum*. Heredity **17**(1): pp. 27-61.

Durrant, A., and Jones, T.W.A. 1971. Reversion of induced changes in amount of nuclear DNA in *Linum*. Heredity **27**(3): pp. 431-439.

Edwards, K., Johnstone, C., and Thompson, C. 1991. A simple and rapid method for the preparation of plant genomic DNA for PCR analysis. Nucleic Acids Research **19**(6): pp. 1349

Evans, G.M., Durrant, A., and Rees, H. 1966. Associated Nuclear Changes in the Induction of Flax Genotrophs. Nature **212**(5063): pp. 697-699.

Hartmann, H.T. and Kester, D.E. 1975. Plant Propagation. Prentice-Hall Inc Englewood Cliffs, New

Jersey.

Hirschhorn, R., Yang, D.R., Puck, J.M., Huie, M.L., Jiang, C.-K., and Kurlandsky, L.E. 1996. Spontaneous in vivo reversion to normal of an inherited mutation in a patient with adenosine deaminase deficiency. Nature Genetics **13**(3): pp. 290-295.

Hirschhorn, R. 2003. *In vivo* reversion to normal of inherited mutations in humans. Journal of Medical Genetics **40:** pp. 721-728.

Jander, G., Norris, S.R., Rounsley, S.D., Bush, D.F., Levin, I.M. and Last, R.L. 2002. *Arabidopsis* map-based cloning in the post-genome era. Plant Physiology **129:** pp. 440-450.

Jonkman, M.F., Scheffer, H., Stulp, R., Pas, H.H., Nijenhuis, M., Heeres, K., Owaribe, K., Pulkkinen, L., and Uitto, J. 1997. Revertant Mosaicism in Epidermolysis Bullosa Caused by Mitotic Gene Conversion. Cell **88**(4): pp. 543-551.

Krolikowski, K.A., Victor, J.A., Wagler, T.A., Lolle, S.J., and Pruitt, R.E. 2003. Isolation and characterization of the *Arabidopsis* organ fusion gene *HOTHEAD*. Plant Journal **35:** pp. 501-511.

Kerk, N.M., Ceserani, T., Tausta, S.L., Sussex, I.M., and Nelson, T.M. 2003. Laser Capture Microdissection of Cells from Plant Tissues. Plant Physiology **132:** pp. 27-35.

Krishnaswamy, L., and Peterson, T. 2007. An Alternate Hypothesis to Explain the High Frequency of "Revertants" in *Hothead* Mutants in *Arabidopsis*. Plant Biology **9:** pp. 30-31.

Kvittengen, E.A., Rootwelt, H., Brandtzaeg, P., Bergan, A., and Berger, R. 1993. Hereditary Tyrosinemia Type I. Self-induced correction of the fumarylacetoacetase defect. Journal of Clinical Investigation **91**(4): pp. 1816.

Kvittengen, E.A., Rootwelt, H., Berger, R., and Brandtzaeg, P. 1994. Self-induced correction of the genetic defect in tyrosinemia type I. Journal of Clinical Investigation **94**(4): pp. 1657-61.

Laibach, F. Uber sommer und winterannuelle rasse von *Arabidopsis thaliana* (L.) heynh ein beitrag zur atiologie der blutenbildung. *Beitr. Biol. Pflantzen* **28** (1951).

Leyser, O., and Day, S. 2003. Mechanisms in Plant Development. Blackwell Science Ltd., Oxford, UK.

Lindblad, B., Lindstedt, S., and Steen, G. 1977. On the enzymic defects in hereditary tyrosinemia. Proceedings of the National Academy of Sciences **74:** pp. 4641-4645.

Linhart, Y.B., and Grant, M.C. 1996. Evolutionary significance of local genetic differentiation in plants. Annual Review of Ecology and Systematics **27:** pp. 237-277

Lolle, S.J., Hsu, W., and Pruitt, R.E. 1998. Genetic Analysis of Organ Fusion in *Arabidopsis* thaliana. Genetics **149:** pp. 607-619.

Lolle, S.J., Victor, J.L., Young, J.M., and Pruitt, R.E. 2005a. Genome-wide non-mendelian inheritance of extra-genomic information in *Arabidopsis*. Nature **434:** pp. 505-509.

Lolle, S.J., Victor, J.L., Young, J.M., and Pruitt, R.E. 2005b. Plant genetics: *Hothead* healer and extragenomic information (reply). Nature **437:** pp. E2.

Lolle, S.J., Victor, J.L., Young, J.M., and Pruitt, R.E. 2006. Increased outcrossing in *hothead* mutants (Reply). Nature **443:** pp. E8-E9.

Maloney, S., Smith, A., Furst, D.E., Myerson, D., Rupert, K., Evans, P.C., and Lee Neelson, J. 1999. Microchimerism of Maternal Origin Persists into Adult Life. Journal of Clinical Investigation **104:** pp. 41-47.

McClintock, B. 1978. Mechanisms that rapidly reorganize the genome. Stadler Symposium **10**: pp. 25-48.

Mendel, G. 1866. Versuche über pflanzenhybriden. (English translation in Dreury and Bateson, 1901). Verh. Naturf. Ver. Brünn. **4:** pp. 3-44.

Mercier, R., Jolivet, S., Vignard, J., Durand, S., Drouaud, J., Pelletier, G., and Nogue, F. 2008. Outcrossing as an Explanation of the Apparent Unconventional Genetic Behaviour of *Arabidopsis thaliana hth* Mutants. Genetics **180:** pp. 2295-2297.

Mitra, R.D., Butty, V.L., Shendure, J., Williams, B.R., Housman, D.E., and Church, G.M. 2003. Digital genotyping and haplotyping with polymerase colonies. Proceedings of the National Academy of Sciences **100:** pp. 5926-5931.

Morrison, R.A., and Evans, D.A. 1988. Haploid Plants from Tissue Culture: New Plant Varieties in a

Shortened Time Frame. Nature Biotechnology **6**(6): pp. 684-690.

Neff, M.M., Neff, J.D., Chory, J. & Pepper, A.E. 1998. dCAPS, a simple technique for the genetic analysis of single nucleotide polymorphisms: experimental applications in *Arabidopsis thaliana* genetics. The Plant Journal **14**: pp. 387-392.

Nitsch, J.P., and Nitsch, C. 1969. Haploid Plants from Pollen Grains. Science **163:** pp. 85-87.

Peng, P., Chan, S.W.L., Shah, G.A., and Jacobsen, S.E. 2006. Plant genetics: Increased outcrossing in *hothead* mutants. Nature **443**(7110): pp. E8-E8.

Pruitt, R.E., Vielle-Calzada, J.-P., Ploense, S.E., Grossniklaus, U., and Lolle, S.J. 2000. *FIDDLEHEAD*, a gene required to suppress epidermal cell interactions in *Arabidopsis*, encodes a putative lipid biosynthetic enzyme. Proceedings of the National Academy of Sciences pp. **97:** pp. 1311-1316.

Ray, A. 2005. Plant genetics: RNA cache or genome trash? Nature **437**(7055): pp. E1-E2.

Roth, E.J., Frazier, B.L., Apuya, N.R., and Lark, K.G. 1989. Genetic Variation in an Inbred Plant: Variation in Tissue Cultures of Soybean [*Glycine max* (L.) Merrill]. Genetics **121:** pp. 359-368.

Satina, S., Blakeslee, A.F., and Avery, A.G. 1940. Demonstration of the Three Germ Layers in the Shoot Apex of *Datura* by Means of Induced Polyploidy in Periclinal Chimeras. American Journal of Botany **27**(10): pp. 895-905.

Schneeberger, R.G., and Cullis, C.A. 1991. Specific DNA Alterations Associated With the Environmental Induction of Heritable Changes in Flax. Genetics **128:** pp. 619-630.

Schuster, S.C. 2008. Next-generation sequencing transforms today's biology. Nature Methods **5**(1): pp. 16-18.

Shendure, J., and Ji, H. 2008. Next-generation DNA sequencing. Nature Biotechnology **26**(10): pp. 1135-1145.

Stewart, D., Candotti, F., and Nelson, D. 2007. The Phenomenon of Spontaneous Genetic Reversions in the Wiskott-Aldrich Syndrome: A Report of the Workshop of the ESID Genetics Working Party at the XIIth Meeting of the European Society for Immunodeficiencies (ESID). Budapest, Hungary October

4–7, 2006. Journal of Clinical Immunology **27**(6): pp. 634-639.

Szymkowiak, E.J., and Sussex, I.M. 1996. What chimeras can tell us about plant development. Annual Review of Plant Physiology and Plant Molecular Biology **47:** pp. 351-376.

Taniguchi T, D.A. 2002. Molecular pathogenesis of fanconi anemia. International Journal of Hematology **75**(2): pp. 123-128.

Thompson, C.J., Movva, N.R., Tizard, R., Crameri, R., Davies, J.E., Lauwereys, M. & Botterman, J. 1987. Characterization of the herbicide-resistance gene *bar* from *Streptomyces hygroscopicus*. The EMBO Journal **6**(9): pp. 2519-2523.

Tost, J., Brandt, O., Boussicault, F., Derbala, D., Caloustian, C., Lechner, D., and Gut, I.G. 2002. Molecular haplotyping at high throughput. Nucleic Acids Research **30**: pp. 96-96.

Wada, T., Schurman, S.H., Otsu, M., Garabedian, E.K., Ochs, H.D., Nelson, D.L., and Candotti, F. 2001. Somatic mosaicism in Wiskott-Aldrich syndrome suggests in vivo reversion by a DNA slippage mechanism. Proceedings of the National Academy of the Sciences **98:** pp. 8697-8702.

Wan, Y., Petolino, J.F., and Widholm, J.M. 1989. Efficient production of doubled haploid plants through colchicine treatment of anther-derived maize callus. TAG Theoretical and Applied Genetics **77**(6): pp. 889-892.

Weigel, D., Ahn, J.H., Blazquez, M.A., Borevitz, J.O., Christensen, S.K., Fankhauser, C., Ferrandiz, C., Kardailsky, I., Malancharuvil, E.J., Neff, M.M., Nguyen, J.T., Sato, S., Wang, Z.-Y., Xia, Y., Dixon, R.A., Harrison, M.J., Lamb, C.J., Yanofsky, M.F., and Chory, J. 2000. Activation Tagging in *Arabidopsis*. Plant Physiology **122:** pp. 1003-1014.

Xiao, M., Wan, E., Chu, C., Hsueh, W.-C., Cao, Y., and Kwok, P.-Y. 2009. Direct determination of haplotypes from single DNA molecules. Nature Methods **6**(3): pp. 199-201.

Yan, H., Papadopoulos, N., Marra, G., Perrera, C., Jiricny, J., Boland, C.R., Lynch, H.T., Chadwick, R.B., de la Chapelle, A., Berg, K., Eshleman, J.R., Yuan, W., Markowitz, S., Laken, S.J., Lengauer, C., Kinzler, K.W., and Vogelstein, B. 2000. Conversion of diploidy to haploidy. Nature **403**(6771): pp. 723-724.

Appendices

Appendix A
Detailed pedigree design rationale

Background – the most basic pedigree to test the segregation of <u>PAP1-D</u> and <u>HTH</u> in the presence of <u>hth</u>

To monitor the stability of *PAP1-D* and *HTH* segregating in the *hth* genetic background would involve at least three steps, mimicking the pedigree in which Lolle *et al.* (2005a) identified restoration of *HTH*. The steps would be:

1. Backcross *PAP1-D* with a *hth* mutant and allow the resulting F1 to self pollinate (F1 = [*HTH/hth, PAP1-D/PAP1*]).
2. Select individual F2 progeny derived from the self-pollinated F1 that are homozygous for *hth* and the *PAP1* allele (F2 = [*hth/hth, PAP1/PAP1]*) and allow these individuals to self-pollinate.
3. Screen the F3 progeny for the reappearance of *PAP1-D* and/or *HTH*.

Based on the results reported by Lolle *et al.* (2005a), approximately 10^{-1} F3 individuals should show restoration to the grandparental (F1) *HTH* allele. If the restoration phenomenon is not limited to sequences native to the *Arabidopsis* genome, restoration of the *PAP1-D* allele may also be observed in the F3.

While the pedigree described above is of primary interest, it provides a limited view of the ancestral requirements of restoration since only one ancestry is investigated. To ensure that all ancestries are thoroughly investigated when testing for the restoration of *PAP1-D* and to obtain a more detailed understanding of the ancestral requirements of *HTH* restoration, a much more comprehensive pedigree was designed and tested.

A comprehensive 'five-tier' pedigree to monitor the segregation of PAP1-D and HTH in a variety of distinct ancestries in the presence of the hth mutation

The pedigree used to monitor the segregation of *PAP1-D* and *HTH* in a variety of distinct ancestries is presented in **Figure 3.4**. Each replicate of the pedigree could potentially produce 48 distinct ancestries, assuming that all lines are generated successfully. The ancestries are *distinct* in the sense that they differ from one another with respect at least one of the following: (1) the presence or absence of *hth* and/or *PAP1-D*, (2) the zygosity of *hth* in the F1, and (3) the direction of hybridization with respect to *hth* and *PAP1-D*. It is not known how these factors may affect gene instability in our *hth* lines therefore determining this is a primary goal of this project, in addition to monitoring the inheritance of *PAP1-D*.

The pedigree is presented as 'Tiers', numbered 1-5. The rationale of the pedigree design is described below for each of the five Tiers.

Tier 1 – Introducing PAP1-D into the hth mutant background

The purpose of Tier 1 is to introduce the transgenic marker *PAP1-D* into the *hth* mutant background and generate genotypically diverse individuals which can be used as parents in the F1 backcrosses (Tier 2). This is accomplished by hybridizing the original *PAP1-D* line with a *hth* fusion mutant, producing the Tier 1 dihybrid:

[HTH/hth, PAP1-D/PAP1]

The Tier 1 dihybrid is allowed to self-pollinate, and the genotypically diverse seed is collected for the generation of Tier 2 founder lines.

Tier 2: F1 backcrosses

The progeny of the selfed Tier 1 dihybrid are screened for six distinct genotypes which will serve as parent lines in the F1 backcrosses. The genotypes of the parent lines (designated α – ζ) are:

α: *[HTH/HTH, PAP1-D/PAP1-D]*

β: *[HTH/hth, PAP1-D/PAP1-D]*

γ: *[hth/hth, PAP1-D/PAP1-D]*

δ: *[HTH/HTH, PAP1/PAP1]*

ε: *[HTH/hth, PAP1/PAP1]*

ζ: *[hth/hth, PAP1/PAP1*]

Using these six distinct parents in the F1 backcrosses allows us to explore what effect, if any, the zygosity of the F1 parent (with respect to *hth*) may have on restoration frequencies in subsequent generations (expanding on the basic pedigree described above, which would test only one parental genotype). Furthermore, using non-transgenic (*PAP1/PAP1*) and wild-type *HTH* (*HTH/HTH*) parents in the F1 backcross (δ – ζ) allows us to test if *PAP1-D* and *HTH* alleles introduced in an earlier generation (i.e. the Tier 1 dihybrid) can be transmitted to the F3 progeny.

F1 seed resulting from the Tier 2 backcrosses is collected for the generation of Tier 3 founder lines.

Tier 3: Selecting F1 parents and generating F2 seed

F1 individuals resulting from the Tier 2 backcrosses are screened for the genotypes shown in Tier 3 of **Figure 3.4**. These F1 individuals are allowed to self-pollinate to segregate the *HTH* and *PAP1-D* alleles (when present) and F2 seed is collected for the generation of Tier 4 founder lines.

Tier 4: Selecting F2 parents and generating F3 seed

The F2 progeny are screened for [*hth/hth*, *PAP1/PAP1*] individuals. These individuals are selected as F2 parents and are allowed to self-pollinate to generate F3 seed for restoration analysis in Tier 5.

Tier 5: Screening F3 progeny for the stable inheritance of parental alleles

In Tier 5, the F3 progeny are planted *en masse* and screened for the stable segregation of the parental alleles. Since all F2 (Tier 4) parents are *PAP1/PAP1*, any purple hirsute plants in the F3 (Tier 5) could be considered putative *PAP1-D* restoration events and thus appropriate candidates for more detailed molecular analyses. Similarly, any F3 individuals exhibiting phenotypically wild-type flowers that descended from *hth/hth* F2's could be considered putative *HTH* restoration events.

Additional notes

It was reported by Lolle *et al.* (2005a) that restoration events tend to be biased to the male reproductive system, although recent work in our laboratory has suggested that restoration events are not necessarily limited to the male gametes and in many cases are likely to be somatic events (Hopkins *et al.*, unpublished data). Nevertheless, it may be informative to test if transmission through the male or female parent affects inheritance in subsequent generations, therefore all Tier 1 and Tier 2 hybridizations in the pedigree described above are conducted reciprocally.

Furthermore, to test if the particular *hth* allele has any affect on the frequency of restoration, two of the three *hth* alleles studied by Lolle *et al*. (2005a) (*hth-4* and *hth-8*) were used to generate replicate pedigrees.

When taken as a whole, this work aims to provide a detailed picture of how both native (*hth*) and transgenic (*PAP1-D*) alleles segregate in a wide variety of ancestries, at a level of detail and scale which has not been attempted previously.

A 'five-tier' pedigree to monitor the segregation of PAP1-D and HTH in the absence of the hth mutation

The results of Lolle *et al.* (2005a) suggest that the *hth* mutation somehow conditions *Arabidopsis* plants for restoration. It is not known for how many generations the instability conferred by a *hth* ancestry can persist, but a fundamental assumption of the experiments described here is that allelic information is stably inherited unless conditioned by the presence of a mutant *hth* allele.

To test this assumption, a pedigree was designed and tested to determine the stability of *PAP1-D* segregating in the absence of the *hth* mutation. This pedigree is summarized in **Figure 3.5** and is analogous to the pedigree described above, except that the *hth* allele is not present at any stage.

Appendix B

Results of profiling markers flanking the unstable loci F15H11_35.66 and MNJ8_44.33 in 100 F3 progeny of 194E9PCL10

Detailed results of PCR genotyping molecular markers flanking the indel markers F15H11_35.66 and MNJ8_44.33 (highlighted) in 100 F3 progeny of 194E9PCL10. The location along the BAC in kb corresponds to the molecular markers described in section **2.1.6**. L = homozygous L*er*; H = heterozygous Col/L*er*. The F2 parent (194E9PCL10) was a fusion mutant that was L at all loci. The 100 F3 progeny are grouped according to the 20 branches of 194E9PCL10 from which they were collected. Floral phenotypes of the F3 progeny are also shown (F = fusion, W = wild-type floral morphology). This data is summarized in **Figure 2.6**.

			F15H11 (Chromosome 1)					MNJ8 BAC (Chromosome 5)				
Branch (F2)	Progeny (F3)	Floral Phenotype(F3)	11.76kb	28.03kb	35.66kb	47.48kb	101.57kb	5.67kb	41.49kb	44.33kb	55.83kb	88.03kb
1	#1-5	F	L	L	L	L	L	H	H	H	H	H
2	#1-12	W	H	H	H	H	H	H	H	H	H	H
3	#1-3	F	L	L	L	L	L	L	L	L	L	L
4	#1-5	W	H	H	H	H	H	H	H	H	H	H
5	#1-5	F	L	L	L	L	L	L	L	L	L	L
7	#1-5	F	L	L	L	L	L	L	L	L	L	L
8	#1-5	F	L	L	L	L	L	L	L	L	L	L
9	#1-5	F	L	L	L	L	L	L	L	L	L	L
10	#1, #3-5	F	L	L	L	L	L	L	L	L	L	L
10	#2	W	H	H	H	H	H	H	H	H	H	H
11	#1, #3-4	W	H	H	H	H	H	H	H	H	H	H
11	#2	F	L	L	L	L	L	L	L	L	L	L
12	#1	W	H	H	H	H	H	L	L	L	L	L
12	#2, #4-6	F	L	L	L	L	L	L	L	L	L	L
13	#1-5	F	L	L	L	L	L	L	L	L	L	L
14	#1-2	F	L	L	L	L	L	L	L	L	L	L
15	#1-5	F	L	L	L	L	L	L	L	L	L	L
16	#1-7	W	H	H	H	H	H	H	H	H	H	H
17	#1-5	F	L	L	L	L	L	L	L	L	L	L
18	#1-7	F	L	L	L	L	L	L	L	L	L	L
19	#1-5	F	L	L	L	L	L	L	L	L	L	L
20	#1-5	F	L	L	L	L	L	L	L	L	L	L

Appendix C

Using the seed stock cataloguing system

To keep track of the various individuals and their experimental ancestries, a system of "catalogue numbers" was devised which provides all of the information needed to deduce the ancestry of any individual to whom a catalogue number has been assigned. To clarify how to interpret these catalogue numbers, a hypothetical catalogue number from each Tier is discussed here.

▪ *Tier 1*

Catalogue numbers for Tier 1 hybrids consist of three informative parts separated by decimals.
Tier 1 catalogue number example:

i. "*hth4*"

Two *hth* alleles were studied in independent 5-Tier pedigrees: *hth-4* and *hth-8*. All catalogue numbers begin with either *hth4* or *hth8* to denote which allele was used in the Tier 1 hybridization and Tier 2 backcrosses.

Therefore, the first part of this catalogue number tells us that this individual belongs to a pedigree in which the *hth-4* allele is being investigated.

ii. "T1'"

T1' indicates that this catalogue number refers to a Tier 1 individual. *Prime* (') indicates the direction of the Tier 1 hybridization which generated this individual. The presence of *prime* indicates that *the transgenic parent was ♀ in the parental cross*. Therefore, since this catalogue number is T1', we know that the parental Tier 1 backcross was:

♂[*HTH/HTH, PAP1/PAP1*] x ♀[*hth-4/hth-4*, *PAP1-D/PAP1-D*]

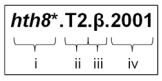
Alternatively, the *absence* of *prime* would indicate that the transgenic individual was ♂ in the parental cross.

iii. "1001"

This 4 digit number identifies the individual plant. For a Tier 1 individual, the first digit is always "1" therefore the 4 digit number is limited to the range 1001 – 1999.

▪ *Tier 2*

Catalogue numbers for the remaining Tiers consist of four informative parts separated by decimals.
Tier 2 catalogue number example:



i. "*hth8"**

As in Tier 1, *hth8* indicates the allele used in this particular hybridization regime. Two pedigrees (one each for *hth-4* and *hth-8*) were constructed in which the homozygosity of *PAP1-D* was not confirmed in the transgenic Tier 2 lineages α – γ. Individuals descending from these Tier 2 lineages are distinguished with an asterisk following the allele in the catalogue number. Therefore for this catalogue number, *hth8** indicates that this individual is part of the *hth-8* pedigree in which the homozygosity of *PAP1-D* was not confirmed in the α – γ Tier 2 lineages.

ii. "T2"

This indicates a Tier 2 individual. The absence of *prime* tells us that this individual is a descendant of a Tier 1 cross in which the transgenic parent was ♂.

iii. "β"

Beta indicates that this individual belongs to the β Tier 2 parental lineage. The six Tier 2 lineages are described in **Appendix A**. Individuals of the β lineage are [*HTH/hth, PAP1-D/PAP1-D*], although since this particular individual is from the *hth8** pedigree, we know that the homozygosity of *PAP1-D* was not confirmed.

iv: "2001"

This is the 4 digit number identifying this individual plant. The number always begins with "2" for a Tier 2 individual.

▪ ***Tier 3***

Tier 3 catalogue number example:

i. "*hth8*"

This individual is from the *hth8* pedigree. The absence of an asterisk indicates that the homozygosity of *PAP1-D* was confirmed in the α – γ Tier 2 lineages.

ii. "T3'"

This indicates a Tier 3 individual *that* descended from a Tier 1 cross in which the transgenic parent was ♀.

iii. "$\varepsilon^{a'}$"

This part of the catalogue number provides information regarding the Tier 2 (F1) backcross from which this Tier 3 individual descended. More specifically, $\varepsilon^{a'}$ tells us the direction of the backcross and whether the backcross was to a L*er* individual (**a** = [*HTH/HTH, PAP1/PAP1*]) or a *hth* mutant (**b** = [*hth/hth, PAP1/PAP1*]). The presence of *prime* tells us that the experimental parent (ε) was ♀ in the Tier 2 backcross. Therefore, $\varepsilon^{a'}$ tells us that this individual descended from the F1 (Tier 2) backcross:

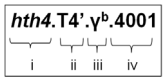
♂[*HTH/HTH, PAP1/PAP1*] x ♀[*HTH/hth, PAP1/PAP1*]

iv. "3001"

This 4 digit number identifies this individual plant. The number always begins with "3" for a Tier 3 individual.

▪ ***Tier 4***

Tier 4 catalogue number example:



i – iv.

This catalogue number is for a Tier 4 individual from the 140C10 experiment. In the Tier 1 cross, the transgenic parent was ♀. The homozygosity of *PAP1-D* was confirmed for the α – γ Tier 2 lineages. This individual descended from the γ lineage and the following Tier 2 (F1) backcross:

♂[*hth/hth, PAP1-D/PAP1-D*] x ♀[*hth/hth, PAP1/PAP1*]

▪ ***Tier 5***

Tier 5 catalogue number example:

i-iii.

This catalogue number is for a Tier 5 individual from the *hth4* experiment. In the Tier 1 cross, the transgenic parent was ♂. The homozygosity of *PAP1-D* was confirmed in the α – γ Tier 2 lineages. This individual descended from the α lineage and the following Tier 2 (F1) backcross:

♂[*HTH/HTH, PAP1-D/PAP1-D*] x ♀[*hth/hth, PAP1/PAP1*]

iv. "5001^{4001}"

In addition to the 4 digit number for this Tier 5 individual, the superscript-4001 indicates the Tier 4 parent. This is important to note because in some cases, Tier 5 seed was collected from multiple Tier 4 sibs.

▪ ***HTH pedigree***

The catalogue system used for the *HTH* pedigree is analogous to that described above. The primary difference is that each catalogue number for the control samples begins with "*HTH*" instead of a particular *hth* mutant allele.

Appendix D

Tier 1 reciprocal hybrids (*hth-4* and *hth-8* mutant backgrounds)

Tier 1 reciprocally generated hybrid lines. Hybrids were generated using two *hth* alleles in the L*er* background: *hth-4* and *hth-8*. The genotypes of the hybrids were confirmed and the individuals shown were allowed to self pollinate to generate seed for Tier 2.

		♀	
	X	***PAP1-D/PAP1-D HTH/HTH ER/ER***	***PAP1/PAP1 hth-4/hth-4 er/er***	***PAP1/PAP1 hth-8/hth-8 er/er***
♂	***PAP1-D/PAP1-D HTH/HTH ER/ER***	-	140C10.T1.1014	208F6.T1.1051
	PAP1/PAP1 hth-4/hth-4 Er/er	140C10.T1'.1001	-	-
	PAP1/PAP1 hth-8/hth-8 Er/er	208F6.T1'.1048	-	-

Appendix E

Tier 2 hybridizations (*hth4**, *hth8**, and *hth8* pedigrees)

i: *hth4** pedigree Tier 2 backcrosses and the resulting Tier 3 founder lines. In the α – ζ Tier 2 lines, "n/a" refers to Tier 2 genotypes which could not be identified. In the Tier 3 founder lines "n/a" refers to unsuccessful hybridizations.

hth4 Tier 2 Backcrosses (hth4.T1.1014 Line)*

X	**α** **(hth4*.T2.α.2029)** ***HTH/HTH PAP1-D/PAP1-D***	**β** **(hth4*.T2.β.2004)** ***HTH/hth PAP1-D/PAP1-D***	**γ** **(hth4*.T2.γ.2001)** ***hth/hth PAP1-D/PAP1-D***	**δ** **n/a** ***HTH/HTH PAP1/PAP1***	**ε** **(hth4*.T2.ε.2058)** ***HTH/hth PAP1/PAP1***	**ζ** **(hth4*.T2.ζ.2046)** ***hth/hth PAP1/PAP1***
a ***HTH/HTH PAP1/PAP1***	hth4*.T3.α^{a} (n/a) hth4*.T3.$\alpha^{a'}$.3033	hth4*.T3.β^{a} (n/a) hth4*.T3.$\beta^{a'}$ (n/a)	hth4*.T3.γ^{a}.3084 hth4*.T3.$\gamma^{a'}$.3092	n/a	hth4*.T3.ε^{a}.3062 hth4*.T3.$\varepsilon^{a'}$.3076	hth4*.T3.ζ^{a}.3045 hth4*.T3.$\zeta^{a'}$.3055
b ***hth4/hth4 PAP1/PAP1***	hth4*.T3.αa.3028 hth4*.T3.$\alpha^{b'}$.3041	hth4*.T3.β^{b}.3108 hth4*.T3.$\beta^{b'}$.3121	hth4*.T3.γ^{b}.3088 hth4*.T3.$\gamma^{b'}$.3095	n/a	hth4*.T3.ε^{b} (n/a) hth4*.T3.$\varepsilon^{b'}$.3081	hth4*.T3.ζ^{b}.3048 hth4*.T3.$\zeta^{b'}$.3060

hth4 Tier 2 Backcrosses (hth4.T1'.1001 Line)*

X	**α** **(hth4*.T2'.α.2035)** ***HTH/HTH PAP1-D/PAP1-D***	**β** **(hth4*.T2'.β.2001)** ***HTH/hth PAP1-D/PAP1-D***	**γ** **(hth4*.T2'.γ.2069)** ***hth/hth PAP1-D/PAP1-D***	**δ** **(hth4*.T2'.δ.2010)** ***HTH/HTH PAP1/PAP1***	**ε** **(hth4*.T2'.ε.2067)** ***HTH/hth PAP1/PAP1***	**ζ** **(hth4*.T2'.ζ.2063)** ***hth/hth PAP1/PAP1***
a ***HTH/HTH PAP1/PAP1***	hth4*.T3'.α^{a}.3156 hth4*.T3'.$\alpha^{a'}$.3162	hth4*.T3'.β^{a} (n/a) hth4*.T3'.$\beta^{a'}$.3184	hth4*.T3'.γ^{a}.3151 hth4*.T3'.$\gamma^{a'}$ (n/a)	hth4*.T3'.δ^{a}.3218 hth4*.T3'.$\delta^{a'}$.3196	hth4*.T3'.ε^{a}.3124 hth4*.T3'.$\varepsilon^{a'}$.3139	hth4*.T3'.ζ^{a}.3175 hth4*.T3'.$\zeta^{a'}$ (n/a)
b ***hth4/hth4 PAP1/PAP1***	hth4*.T3'.α^{b}.3159 hth4*.T3'.$\alpha^{b'}$.3164	hth4*.T3'.β^{b}.3193 hth4*.T3'.$\beta^{b'}$ (n/a)	hth4*.T3'.γ^{b}.3152 hth4*.T3'.$\gamma^{b'}$ (n/a)	hth4*.T3'.δ^{b}.3219 hth4*.T3'.$\delta^{b'}$.3215	hth4*.T3'.ε^{b}.3134 hth4*.T3'.$\varepsilon^{b'}$ (n/a)	hth4*.T3'.ζ^{b}.3176 hth4*.T3'.$\zeta^{b'}$ (n/a)

ii: *hth8** pedigree Tier 2 backcrosses and the resulting Tier 3 founder lines. In the α – ζ Tier 2 lines, "n/a" refers to Tier 2 genotypes which could not be identified. In the Tier 3 founder lines "n/a" refers to unsuccessful hybridizations.

hth8 Tier 2 Backcrosses (hth8.T1.1051 Line)*

X	**α** **n/a** ***HTH/HTH*** ***PAP1-D/PAP1-D***	**β** **n/a** ***HTH/hth*** ***PAP1-D/PAP1-D***	**γ** **(hth8*.T2.γ.2041)** ***hth/hth*** ***PAP1-D/PAP1-D***	**δ** **n/a** ***HTH/HTH*** ***PAP1/PAP1***	**ε** **(hth8*.T2.ε.2006)** ***HTH/hth*** ***PAP1/PAP1***	**ζ** **(hth8*.T2.ζ.2018)** ***hth/hth*** ***PAP1/PAP1***
a ***HTH/HTH*** ***PAP1/PAP1***	n/a	n/a	hth8*.T3.γ^{a}.3149 hth8*.T3.$\gamma^{a'}$.3152	n/a	hth8*.T3.ε^{a} (n/a) hth8*.T3.$\varepsilon^{a'}$.3118	hth8*.T3.ζ^{a}.3099 hth8*.T3.$\zeta^{a'}$.3103
b ***hth8/hth8*** ***PAP1/PAP1***	n/a	n/a	hth8*.T3.γ^{b}.3142 hth8*.T3.$\gamma^{b'}$.3145	n/a	hth8*.T3.ε^{b}.3105 hth8*.T3.$\varepsilon^{b'}$ (n/a)	hth8*.T3.ζ^{b}.3096 hth8*.T3.$\zeta^{b'}$.3098

hth8 Tier 2 Backcrosses (hth8.T1'.1048 Line)*

X	**α** **(hth8*.T2'.α.2015)** ***HTH/HTH*** ***PAP1-D/PAP1-D***	**β** **(hth8*.T2'.β.2041)** ***HTH/hth*** ***PAP1-D/PAP1-D***	**γ** **(hth8*.T2'.γ.2011)** ***hth/hth*** ***PAP1-D/PAP1-D***	**δ** **(hth8*.T2'.δ.2014)** ***HTH/HTH*** ***PAP1/PAP1***	**ε** **(hth8*.T2'.ε.2005)** ***HTH/hth*** ***PAP1/PAP1***	**ζ** **(hth8*.T2'.ζ.2066)** ***hth/hth*** ***PAP1/PAP1***
a ***HTH/HTH*** ***PAP1/PAP1***	hth8*.T3'.α^{a}.3039 hth8*.T3'.$\alpha^{a'}$.3041	hth8*.T3'.β^{a} (n/a) hth8*.T3'.$\beta^{a'}$ (n/a)	hth8*.T3'.γ^{a} (n/a) hth8*.T3'.$\gamma^{a'}$ (n/a)	hth8*.T3'.δ^{a}.3029 hth8*.T3'.$\delta^{a'}$.3031	hth8*.T3'.ε^{a} (n/a) hth8*.T3'.$\varepsilon^{a'}$ (n/a)	hth8*.T3'.ζ^{a}.3072 hth8*.T3'.$\zeta^{a'}$.3074
b ***hth8/hth8*** ***PAP1/PAP1***	hth8*.T3'.α^{b}.3044 hth8*.T3'.$\alpha^{b'}$.3047	hth8*.T3'.β^{b}.3078 hth8*.T3'.$\beta^{b'}$ (n/a)	hth8*.T3'.γ^{b}.3049 hth8*.T3'.$\gamma^{b'}$.3051	hth8*.T3'.δ^{b}.3034 hth8*.T3'.$\delta^{b'}$.3037	hth8*.T3'.ε^{b}.3057 hth8*.T3'.$\varepsilon^{b'}$.3060	hth8*.T3'.ζ^{b}.3068 hth8*.T3'.$\zeta^{b'}$.3071

iii: *hth8* pedigree Tier 2 backcrosses and the resulting Tier 3 founder lines. In the α – ζ Tier 2 lines, "n/a" refers to Tier 2 genotypes which could not be identified. In the Tier 3 founder lines "n/a" refers to unsuccessful hybridizations

hth8 Tier 2 Backcrosses (hth8.T1.1051 Line)

X	**α** **(hth8.T2.α.2009)** ***HTH/HTH PAP1-D/PAP1-D***	**β** **(hth8.T2.β.2006)** ***HTH/hth PAP1-D/PAP1-D***	**γ** **(hth8.T2.γ.2008)** ***hth/hth PAP1-D/PAP1-D***	**δ** **n/a** ***HTH/HTH PAP1/PAP1***	**ε** **(hth8.T2.ε.2038)** ***HTH/hth PAP1/PAP1***	**ζ** **(hth8.T2.ζ.2003)** ***hth/hth PAP1/PAP1***
a ***HTH/HTH PAP1/PAP1***	hth8.T3.α^{a}.3177 hth8.T3.α$^{a'}$.3171	hth8.T3.β^{a}.3211 hth8.T3.β$^{a'}$ (n/a)	hth8.T3.γ^{a} (n/a) hth8.T3.γ$^{a'}$.3227	n/a	hth8.T3.ε^{a}.3254 hth8.T3.ε$^{a'}$.3244	hth8.T3.ζ^{a}.3285 hth8.T3.ζ$^{a'}$.3278
b ***hth8/hth8 PAP1/PAP1***	hth8.T3.α^{b}.3190 hth8.T3.α$^{b'}$.3184	hth8.T3.β^{b}.3221 hth8.T3.β$^{b'}$ (n/a)	hth8.T3.γ^{b}.3236 hth8.T3.γ$^{b'}$ (n/a)	n/a	hth8.T3.ε^{b}.3268 hth8.T3.ε$^{b'}$.3260	hth8.T3.ζ^{b}.3296 hth8.T3.ζ$^{b'}$.3290

hth8 Tier 2 Backcrosses (hth8.T1'.1048 Line)

X	**α** **(hth8.T2'.α.2007)** ***HTH/HTH PAP1-D/PAP1-D***	**β** **(hth8.T2'.β.2039)** ***HTH/hth PAP1-D/PAP1-D***	**γ** **(hth8.T2'.γ.2014)** ***hth/hth PAP1-D/PAP1-D***	**δ** **(hth8.T2'.δ.2034)** ***HTH/HTH PAP1/PAP1***	**ε** **(hth8.T2'.ε.2016)** ***HTH/hth PAP1/PAP1***	**ζ** **(hth8.T2'.ζ.2038)** ***hth/hth PAP1/PAP1***
a ***HTH/HTH PAP1/PAP1***	hth8.T3'.α^{a}.3006 hth8.T3'.α$^{a'}$.3001	hth8.T3'.β^{a}.3031 hth8.T3'.β$^{a'}$ (n/a)	hth8.T3'.γ^{a} (n/a) hth8.T3.γ$^{a'}$.3061	hth8.T3'.δ^{a}.3078 hth8.T3'.δ$^{a'}$.3072	hth8.T3'.ε^{a} (n/a) hth8.T3'.ε$^{a'}$.3102	hth8.T3'.ζ^{a}.3146 hth8.T3'.ζ$^{a'}$.3134
b ***hth8/hth8 PAP1/PAP1***	hth8.T3'.α^{b}.3011 hth8.T3'.α$^{b'}$ (n/a)	hth8.T3'.β^{b}.3050 hth8.T3'.β$^{b'}$ (n/a)	hth8.T3'.γ^{b} (n/a) hth8.T3'.γ$^{b'}$.3067	hth8.T3'.δ^{b}.3089 hth8.T3'.δ$^{b'}$ (n/a)	hth8.T3'.ε^{b}.3123 hth8.T3'.ε$^{b'}$.3114	hth8.T3'.ζ^{b}.3166 hth8.T3'.ζ$^{b'}$ (n/a)

Appendix F

Test crosses to determine the genetic distance between *PAP1* and *HTH*

To determine the genetic distance between *PAP1* and *HTH*, twelve replicate test crosses (*[HTH/hth, PAP1-D/PAP1] x [hth/hth, PAP1/PAP1]*) were conducted. The phenotypes of the progeny were scored to identify the total number of recombinant genotypes 'B' (*[HTH/HTH, PAP1/PAP1]*) and 'C' (*[hth/hth, PAP1-D/PAP1-D]*). Pooling the results yielded an average ($\dot{x}$) recombination frequency (%RF) of 22.0 between *PAP1* and *HTH*, which represents a genetic distance of 22.0 cM.

Test Cross (Replicate)	(A) WT Flowers, Purple	(B) WT Flowers, Green	(C) Fused Flowers, Purple	(D) Fused Flowers, Green	Total	Recombinant (B+C)	%RF
1	27	6	13	29	75	19	25.3
2	22	5	5	25	57	10	17.5
3	24	12	6	41	83	18	21.7
4	38	8	9	24	79	17	21.5
5	23	11	5	27	66	16	24.2
6	22	4	7	29	62	11	17.7
7	10	4	4	19	37	8	21.6
8	26	9	5	26	66	14	21.2
9	21	10	8	30	69	18	26.1
10	8	3	3	7	21	6	28.6
11	28	8	3	18	57	11	19.3
12	28	7	6	28	69	13	18.8
						$\dot{x}$	22.0
						σ	3.5

Appendix G

Tier 3 (F1) founder lines (*hth4**, *hth8**, and *hth8* pedigrees)

i: *hth4** pedigree Tier 3 founder lines. Phenotypic scoring and molecular genotyping was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

hth4* **pedigree Tier 3 founder lines**					
Tier 3 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
hth4*.T3.α$^{a'}$.3033	P	WT	*er/er*	*HTH/HTH*	*PAP1-D/PAP1*
hth4*.T3.α^{b}.3028	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3.α$^{b'}$.3041	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3.β^{b}.3108	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3.β$^{b'}$.3121	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3.γ^{a}.3084	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3.γ$^{a'}$.3092	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3.γ^{b}.3088	P	F	*ER*	*hth/hth*	*PAP1-D/PAP1*
hth4*.T3.γ$^{b'}$.3095	P	F	*ER*	*hth/hth*	*PAP1-D/PAP1*
hth4*.T3.ε^{a}.3062	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3.ε$^{a'}$.3076	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3.ε$^{b'}$.3081	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3.ζ^{a}.3045	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3.ζ$^{a'}$.3055	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3.ζ^{b}.3048	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth4*.T3.ζ$^{b'}$.3060	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth4*.T3'.α^{a}.3156	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
hth4*.T3'.α$^{a'}$.3162	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
hth4*.T3'.α^{b}.3159	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3'.α$^{b'}$.3164	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3'.β$^{a'}$.3184	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3'.β^{b}.3193	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3'.γ^{a}.3151	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth4*.T3'.γ^{b}.3152	P	F	*er/er*	*hth/hth*	*PAP1-D/PAP1*
hth4*.T3'.δ^{a}.3218	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T3'.δ$^{a'}$.3196	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T3'.δ^{b}.3219	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3'.δ$^{b'}$.3215	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3'.ε^{a}.3124	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3'.ε$^{a'}$.3139	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3'.ε^{b}.3134	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3'.ζ^{a}.3175	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth4*.T3'.ζ^{b}.3176	G	F	*ER*	*hth/hth*	*PAP1/PAP1*

ii: *hth8** pedigree Tier 3 founder lines. Phenotypic scoring and molecular genotyping was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

hth8* **pedigree Tier 3 founder lines**					
Tier 3 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
hth8*.T3.γ^{a}.3149	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8*.T3.γ$^{a'}$.3152	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth8*.T3.γ^{b}.3142	P	F	*ER*	*hth/hth*	*PAP1-D/PAP1*
hth8*.T3.γ$^{b'}$.3145	P	F	*ER*	*hth/hth*	*PAP1-D/PAP1*
hth8*.T3.ε$^{a'}$.3118	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3.ε^{b}.3105	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3.ζ^{a}.3099	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3.ζ$^{a'}$.3103	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3.ζ^{b}.3096	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T3.ζ$^{b'}$.3098	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T3'.α^{a}.3039	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
hth8*.T3'.α$^{a'}$.3041	P	WT	*er/er*	*HTH/HTH*	*PAP1-D/PAP1*
hth8*.T3'.α^{b}.3044	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8*.T3'.α$^{b'}$.3047	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth8*.T3'.β^{b}.3078	P	WT	*er/er*	*HTH/hth*	*PAP1-D/PAP1*
hth8*.T3'.γ^{b}.3049	P	F	*n/a*	*hth/hth*	*PAP1-D/PAP1*
hth8*.T3'.γ$^{b'}$.3051	P	F	*n/a*	*hth/hth*	*PAP1-D/PAP1*
hth8*.T3'.δ^{a}.3029	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8*.T3'.δ$^{a'}$.3031	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8*.T3'.δ^{b}.3034	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3'.δ$^{b'}$.3037	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3'.ε^{b}.3057	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3'.ε$^{b'}$.3060	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3'.ζ^{a}.3072	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3'.ζ$^{a'}$.3074	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8*.T3'.ζ^{b}.3068	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T3'.ζ$^{b'}$.3071	G	F	*ER*	*hth/hth*	*PAP1/PAP1*

iii: *hth8* pedigree Tier 3 founder lines. Phenotypic scoring and molecular genotyping was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

***hth8* pedigree Tier 3 founder lines**					
Tier 3 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
hth8.T3.α^{a}.3177	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
hth8.T3.α$^{a'}$.3171	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
hth8.T3.α^{b}.3190	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3.α$^{b'}$.3184	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3.β^{a}.3211	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3.β^{b}.3221	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3.γ$^{a'}$.3227	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3.γ^{b}.3236	P	F	*er/er*	*hth/hth*	*PAP1-D/PAP1*
hth8.T3.ε^{a}.3254	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8.T3.ε$^{a'}$.3244	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8.T3.ε^{b}.3268	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8.T3.ε$^{b'}$.3260	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8.T3.ζ^{a}.3285	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8.T3.ζ$^{a'}$.3278	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8.T3.ζ^{b}.3296	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T3.ζ$^{b'}$.3290	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T3'.α^{a}.3006	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
hth8.T3'.α$^{a'}$.3001	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
hth8.T3'.α^{b}.3011	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3'.β^{a}.3031	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3'.β^{b}.3050	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3.γ$^{a'}$.3061	P	WT	*ER*	*HTH/hth*	*PAP1-D/PAP1*
hth8.T3'.γ$^{b'}$.3067	P	F	*er/er*	*hth/hth*	*PAP1-D/PAP1*
hth8.T3'.δ^{a}.3078	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth8.T3'.δ$^{a'}$.3072	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth8.T3'.δ^{b}.3089	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8.T3'.ε$^{a'}$.3102	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8.T3'.ε^{b}.3123	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8.T3'.ε$^{b'}$.3114	G	WT	*er/er*	*HTH/hth*	*PAP1/PAP1*
hth8.T3'.ζ^{a}.3146	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8.T3'.ζ$^{a'}$.3134	G	WT	*ER*	*HTH/hth*	*PAP1/PAP1*
hth8.T3'.ζ^{b}.3166	G	F	*ER*	*hth/hth*	*PAP1/PAP1*

Appendix H

Tier 4 (F2) founder lines (*hth4**, *hth8**, and *hth8* pedigrees)

i: *hth4** pedigree Tier 4 founder lines. Phenotypic scoring and molecular genotyping was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

hth4* **pedigree Tier 4 founder lines**					
Tier 4 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
hth4*.T4.α^{a}.4101	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4.α^{a}.4001	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4.α^{b}.4107	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4.β^{b}.4117	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.β^{b}.4118	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.β$^{b'}$.4120	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.γ^{a}.4124	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.γ^{a}.4128	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.γ^{b}.4005	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.γ$^{b'}$.4009	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.γ$^{b'}$.4010	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ε^{a}.4013	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ε^{a}.4019	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ε^{b}.4023	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ε$^{b'}$.4024	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ζ^{a}.4160	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ζ^{a}.4162	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ζ^{a}.4051	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ζ^{a}.4053	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ζ^{b}.4094	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4.ζ$^{b'}$.4095	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.α^{a}.4054	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.α^{a}.4058	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.α^{b}.4135	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.α$^{b'}$.4185	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.α$^{b'}$.4188	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.β^{a}.4143	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.β^{b}.4151	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.γ^{a}.4156	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.γ^{b}.4063	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.γ^{b}.4065	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.δ^{a}.4097	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.δ^{a}.4096	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.δ^{b}.4070	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.δ$^{b'}$.4069	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth4*.T4'.ε^{a}.4075	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.ε^{a}.4079	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.ε^{a}.4081	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.ε^{b}.4166	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.ε^{b}.4169	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.ζ^{a}.4090	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*
hth4*.T4'.ζb.4098	G	F	*n/a*	*hth/hth*	*PAP1/PAP1*

ii*: hth8** pedigree Tier 4 founder lines. Phenotypic scoring and molecular genotyping was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

hth8* **pedigree Tier 4 founder lines**					
Tier 4 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
hth8*.T4.γ^{a}.4059	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.γ$^{a'}$.4061	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.γ^{b}.4055	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.γ$^{b'}$.4057	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.ε$^{a'}$.4053	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.ε^{b}.4047	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.ζ^{a}.4043	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.ζ$^{a'}$.4045	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.ζ^{b}.4039	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4.ζ$^{b'}$.4041	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.α^{a}.4009	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8*.T4'.α$^{a'}$.4011	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8*.T4'.α^{b}.4013	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.α^{b}.4014	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.α^{b}.4015	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.α$^{b'}$.4016	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.α$^{b'}$.4017	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.β^{b}.4037	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.β^{b}.4038	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.γ^{b}.4018	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.γ$^{b'}$.4019	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.δ^{a}.4001	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8*.T4'.δ$^{a'}$.4003	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8*.T4'.δ^{b}.4005	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.δ$^{b'}$.4007	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.ε^{b}.4025	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.ε$^{b'}$.4027	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.ζ^{a}.4033	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.ζ$^{a'}$.4035	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.ζ^{b}.4029	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8*.T4'.ζ$^{b'}$.4031	G	F	*ER*	*hth/hth*	*PAP1/PAP1*

iii: *hth8* pedigree Tier 4 founder lines. Phenotypic scoring and molecular genotyping was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

***hth8* pedigree Tier 4 founder lines**					
Tier 4 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
hth8.T4.α^{a}.4021	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth8.T4.α^{a}.4020	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth8.T4.α^{b}.4023	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.α^{b}.4024	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.α^{b}.4022	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.β^{a}.4025	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4.β^{b}.4026	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.β^{b}.4027	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4.γ^{a}.4028	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.γ^{a}.4029	G	F	*ER?*	*hth/hth*	*PAP1/PAP1*
hth8.T4.γ^{b}.4030	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8.T4.ε^{a}.4032	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ε^{a}.4033	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ε^{a}.4031	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ε^{b}.4035	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ε^{b}.4036	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ε^{b}.4034	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ζ^{a}.4038	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ζ^{a}.4037	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ζ^{b}.4039	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4.ζ^{b}.4040	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.α^{a}.4002	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
hth8.T4'.α^{a}.4001	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8.T4'.α^{b}.4003	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.β^{a}.4006	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.β^{b}.4007	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4.γ^{a}.4008	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.γ^{b}.4009	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.δ^{a}.4011	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8.T4'.δ^{a}.4010	G	WT	*er/er*	*HTH/HTH*	*PAP1/PAP1*
hth8.T4'.δ^{b}.4012	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ε^{a}.4013	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ε^{b}.4016	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ε^{b}.4014	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ε^{b}.4015	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ζ^{a}.4018	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ζ^{a}.4017	G	F	*er/er*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ζ^{b}.4019	G	F	*ER*	*hth/hth*	*PAP1/PAP1*
hth8.T4'.ζ^{b}.4019	G	F	*ER*	*hth/hth*	*PAP1/PAP1*

Appendix I

Tier 1 reciprocal hybrids (absence of *hth*)

HTH pedigree Tier 1 hybridizations. The resulting progeny shown were allowed to self pollinate to generate seed for Tier 2.

X	***PAP1-D/PAP1-D*** ***HTH/HTH***
HTH/HTH ***PAP1/PAP1***	HTH.T1.1070 HTH.T1'.1060

Appendix J

Tier 2 hybridizations (*HTH* pedigree)

HTH pedigree Tier 2 backcrosses and resulting Tier 3 founder lines.

HTH Pedigree Tier 2 Backcrosses (HTH.T1.1070 Line)

X	**α** **(HTH.T2.α.2001)** ***HTH/HTH*** ***PAP1-D/PAP1-D***	**β** **(HTH.T2.β.2002)** ***HTH/HTH*** ***PAP1/PAP1***
a ***HTH/HTH*** ***PAP1/PAP1***	HTH.T3.α^{a}.3014 HTH.T3.$\alpha^{a'}$.3012	HTH.T3.β^{a}.3019 HTH.T3.$\beta^{a'}$.3015

HTH Pedigree Tier 2 Backcrosses (HTH.T1'.1060 Line)

X	**α** **(hth8.T2'.α.2009)** ***HTH/HTH*** ***PAP1-D/PAP1-D***	**β** **(HTH.T2'.β.2006)** ***HTH/HTH*** ***PAP1/PAP1***
a ***HTH/HTH*** ***PAP1/PAP1***	HTH.T3'.α^{a}.3005 HTH.T3'.$\alpha^{a'}$.3001	HTH.T3'.β^{a}.3008 HTH.T3'.$\beta^{a'}$.3010

Appendix K

Tier 3 (F1) founder lines (*HTH* pedigree)

HTH pedigree Tier 3 founder lines. Phenotypic scoring and molecular genotyping was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

***HTH* pedigree Tier 3 founder lines**					
Tier 3 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
HTH.T3.α^{a}.3014	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
HTH.T3.$\alpha^{a'}$.3012	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
HTH.T3.β^{a}.3019	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T3.$\beta^{a'}$.3015	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T3'.α^{a}.3005	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
HTH.T3'.$\alpha^{a'}$.3001	P	WT	*ER*	*HTH/HTH*	*PAP1-D/PAP1*
HTH.T3'.β^{a}.3008	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T3'.$\beta^{a'}$.3010	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*

Appendix L

Tier 4 (F2) founder lines (*HTH* pedigree)

HTH pedigree Tier 4 founder lines. Phenotypic scoring was used to determine the genotypes. P = *PAP1-D* (purple hirsute) pigmentation; G = *PAP1/PAP1* (green wild-type) pigmentation; WT = wild-type floral morphology; F = fusion floral morphology.

***HTH* pedigree Tier 4 founder lines**					
Tier 4 Cat. No.	**Pigmentation**	**Floral Morphology**	**Growth Habit**	***HTH* genotype**	***PAP1* genotype**
HTH.T4.α^{a}.4002	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T4.$\alpha^{a'}$.4001	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T4.β^{a}.4004	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T4.$\beta^{a'}$.4003	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T4'.α^{a}.4007	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T4'.$\alpha^{a'}$.4006	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T4'.β^{a}.4009	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*
HTH.T4'.$\beta^{a'}$.4010	G	WT	*ER*	*HTH/HTH*	*PAP1/PAP1*