A Study on Intraorganismal Genetic Heterogeneity in *Arabidopsis thaliana* in Response to Stress

by

Maye Chin Saechao

A thesis

presented to the University of Waterloo

in fulfilment of the

thesis requirement for the degree of

Master of Science

in

Biology

Waterloo, Ontario, Canada, 2012

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

In sexually reproducing individuals, intraorganismal genetic heterogeneity (IGH) or mosaicism is thought to occur infrequently while genetic homogeneity is presumed the norm. In organisms that undergo modular development, such as long-lived plants, IGH has been substantially documented. In *Arabidopsis thaliana* we have shown that non-parental DNA that is inherited at low but detectable rates can also manifest on single plants as genotypically distinct somatic sectors suggesting that even short-lived annual plants show IGH. The underlying mechanism responsible for generating this type of IGH remains unknown.

In order to better understand this phenomenon I have tested the hypothesis that among genome changes that occur in response to stress, these putative triggers also up-regulate IGH. Metabolic stress, cold stress, mechanical damage and ROS exposure were examined. To test for IGH, transgene markers and polymorphic molecular markers were used. Also, presented in this thesis is work investigating the effect of *in vitro* propagation through tissue culture on IGH frequencies. Regenerated plants as well as undifferentiated callus tissue were genotyped and assayed for sequence reversions.

Molecular genotyping revealed an outcome contrary to that predicted by the initial hypothesis showing instead that a high frequency of restoration occurred in the progeny of un-treated control plants. With the exception of samples passed through tissue culture, molecular marker changes, including single and double reversions of alleles, were detected in every line at some low level Furthermore, many of the revertants were found to be genetic mosaics. DNA sequence analyses revealed that sequences flanking three molecular markers that had undergone reversion were near identical to the great-grandparent of the sequenced individual. These results suggest that stress is perhaps an inhibitor of restoration. Although there may be other explanations for the results described in this thesis, the evidence implicates genome restoration as a mechanism for generating IGH.

Acknowledgements

I would like to thank my present and past labmates over the past couple of years. Pearl Chang, Marianne Hopkins, Meghan Doerr, Daryl Enstone, and Holly Loucas – I appreciate all of the support and advice that have been given, both personal and technical. I would also like to thank Chris Hammill for his contribution to this project.

I would especially like to thank my supervisors Dr. Susan Lolle and Dr. Barb Moffatt for accepting me as a student and believing in me. Without their tremendous efforts and patience I would not be where I am now or accomplished half as much. I would also like to thank my committee members, Dr. Bruce Reed and Dr. Mungo Marsden, for their 'non-botanist' point-of-view, suggestions and advice.

I could not have done this without the funding provided by the University of Waterloo and by the Natural Sciences and Engineering Research Council of Canada.

Table of Contents

List of Figures

List of Tables

Introduction

Many plant species reproduce both sexually and asexually and can leverage the selective advantages offered by each reproductive strategy. According to the laws of Mendelian inheritance, however, only those alleles that are present in the genomes of the parents have the potential to be passed on to the next generation. This tenet holds true irrespective of whether reproduction is sexual or asexual. Typically, in sexually reproducing organisms, alleles exist in pairs that separate during meiosis. Offspring produced by sexual hybridizations receive one gamete from each parent. On the other hand, in asexual reproduction, genetic information is passed on to the offspring mitotically. This method preserves the genetic identity or allelic composition of the individual from which the offspring are derived (Russell, 2006). Novel alleles may arise as a result of mechanisms such as gene conversions, mobile element activity or chromosomal rearrangements and can be inherited in the next generation or the next cell division. However, these alleles may manifest as non-Mendelian ratios deviating from expected segregation ratios (Chen et. al. 2007). The allelic variations, aforementioned, occur at the DNA sequence level.

Hereditary changes that occur in the absence of underlying DNA sequence alteration fall under the umbrella of "epigenetics" (Probst et. al. 2009). Several types of modifications may be passed on, such as DNA methylation, histone or chromatin modifications, nuclear RNA or higher order organization, or even positional information (Probst et. al. 2009). Such modifications serve to regulate gene expression and can be passed on from mother cell to daughter cell or from one generation to the next. The epigenetic expression states can also be perpetuated in the absence of the conditions that created them (Richards, 2006).

A controversial and as yet unresolved example of non-Mendelian inheritance was first described in 2005 by Lolle et al. who documented sequence-level changes that involved the apparent reacquisition of ancestral genetic information; this phenomenon was termed "restoration". This type of inheritance was seen in the progeny of the *Arabidopsis thaliana* organ fusion mutant *HOTHEAD (HTH;* Lolle et. al. 2005) where wild-type function of *HTH* was restored in progeny of plants whose parents were homozygous for the recessive mutant *hth* allele. These plants, termed "revertants", had cryptically reacquired the wild-type allele. Some revertants also carried unexpected sequence changes at other locations within the

genome (Lolle et. al. 2005). Lolle et al (2005) suggested that these changes were directed by templates provided by an extra-genomic source of genetic information stored in the form of an RNA "cache". The genetic information stored within this cache would be of ancestral origin, based on experimental results.

No previously known genetic or epigenetic phenomena could readily explain the genetic instability seen in *hth* mutant plants, however, several alternate explanations have been proposed. There could exist a DNA cache (Ray, 2005), or there could have been ectopic gene conversion using intercalary fragments of embedded genomic DNA sequences (Chaudhury, 2005). Toxic and mutagenic materials may have accumulated and subsequently promoted mutational events at the *hth* locus which, at some frequency, gave rise to normal wild-type plants (Comai, 2005). Out-crossing was also proposed as an explanation due to the known susceptibility of *hth* to out-crossing (Mercier et. al. 2008). Although out-crossing is a possible explanation, it can be argued that it was not consistent with many of the experimental results (Lolle et al., 2006).

RNA has been shown to function as a source of epigenetic information in a variety of organisms. RNA silencing is a gene suppression mechanism that can be heritable over several generations. For example, it was reported that in *Caenorhabditis elegans* heritable phenotypic changes were caused by the introduction of double-stranded RNA (Alcazar et al., 2008). In mice, after induction, gene transcripts and microRNA were shown to persist over several generations (Rassoulzadegan et al., 2006; Wagner et al., 2008). Inherited RNA is not functionally limited to gene silencing mechanisms and has, in fact, been shown to control gene expression, genome rearrangement and chromosome number in the developing nucleus of *Oxytricha trifallax* (Nowacki et al., 2008; Nowacki et al., 2010).

The enigmatic inheritance of non-parental alleles is not exclusive to Arabidopsis but has been described in other plant species. In flax (*Linum usitatissimum*), for example, non-random phenotypic and genomic changes were observed in response to an altered growth environment (Cullis and Charlton, 1981; Chen et al., 2005, Cullis, 2005). Under certain environmental conditions several stable lines, termed genotrophs, were developed from the susceptible inbred flax variety, Stormont Cirrus (Pl). Genotrophs are distinct from the original P1 line in that they breed true when grown in a number of differential environments. Some of the changes observed include differences in capsule septa hair number, plant weight and

height at maturity, and shifts in the mobility of isozymes of peroxidise and acid phosphatase. Genomic changes affected the copy number of ribosomal RNA genes and many repetitive sequence families (Chen et al., 2009).

In 2009, Chen et al. extended this work and demonstrated that an insertion of a novel single copy 5.7 kilobase (kb) DNA fragment, termed LIS-1, could be also be induced by environmental shifts. However, the insertion events themselves were shown to be dependent not only on growth conditions but also on the genetic background used. There were two inducing growth conditions, N-treatment and water treatment. The N-treatment consisted of watering plants with 100 ml per pot of a 1% ammonium sulfate solution, whereas in the water treatment plants were watered with tap water. In these two growth conditions, the LIS-1 insertion appeared during vegetative development and always became homozygous and stably heritable. In the third growth condition, NPK, plants were fed a commercial fertilizer. This growth condition, however, resulted in a variable frequency of LIS-1 insertion without heritable transmission of LIS-1 to the progeny. No insertion of LIS-1 was observed under the control growth condition. The LIS-1 insertions also appeared in another flax line, Hollandia. However, in Hollandia, insertions were not stably integrated unless the plants were grown continuously in the inducing conditions (Chen et al., 2009). Three other flax lines were also tested under the same experimental conditions, but they did not spontaneously acquire the LIS-1 insertions. Based on this research, the Pl line appears to be a highly sensitized line for LIS-1 insertions.

Research in our laboratory has also elucidated more facets of restoration over the past several years. Extensions to the original work from 2005 have revealed instability with insertion-deletion polymorphisms (indels) at numerous loci that map to both genic and intergenic regions across the genome. In *hth* plants these indels revert at a frequency as high as 22.7%. Revertants were also seen in the *A. thaliana* wild-type hybrid background at a frequency of approximately 5.4% (Hopkins et al., 2011).

One possible explanation for this large number of revertant wild-type offspring is cross-pollination although documented rates of out-crossing for wild-type Arabidopsis plants fall well below these percentages, averaging 0.3 to 2.5% (Abbott and Gomes, 1989; Bergelson et al., 1998; Bakker et al., 2006). To quantify the degree of out-crossing in *hth* plants, *hth* and *eceriferum-10* (Koornneef et al., 1989) floral fusion mutants, wild-type Landsberg and

glufosinate-resistant transgenic lines were grown together. These experiments show that mutants with fused flora phenotypes have enhanced rates of out-crossing but also verify that wild-type lines experience a much reduce rate (0.02-0.89% for fusion mutants and 0.01% for wild-type plants: Hopkins et al., 2011).

In the course of determining out-crossing frequencies and doing segregation analysis, a rare mosaic *hth* mutant plant with a large phenotypically wild-type floral sector was isolated (Figure 1; Hopkins et al., 2011). For this individual, phenotype was found to correspond to genotype with the wild-type *HTH* and mutant *hth-4* alleles both detected in the wild-type sector (Figure 1B). This individual provided the first robust phenotypic evidence showing that single Arabidopsis plants were capable of producing genetically distinct somatic sectors representing a case of intraorganismal genetic heterogeneity (IGH).

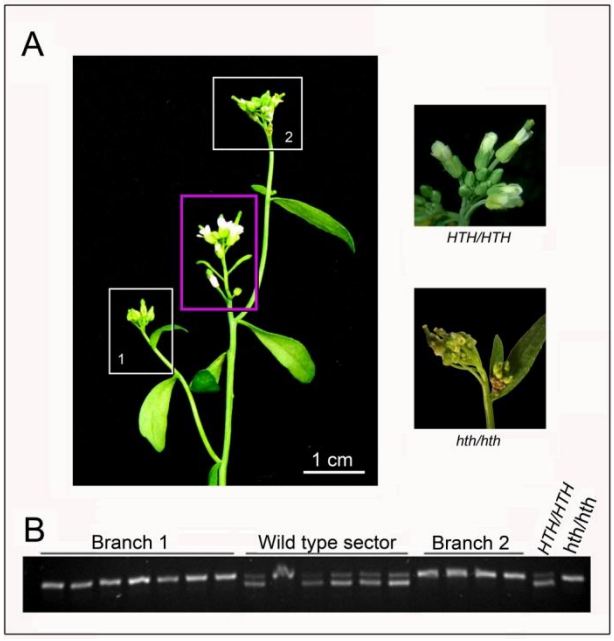


Figure 1: The mosaic mutant hth-4 plant showing phenotypically wild-type sector (Hopkins et al., 2011). A. Wild-type sector (magenta box) among mutant branches (white boxes). Examples of mutant *hth* and wild-type flowers are to the right. B. Molecular analysis was conducted on mutant and wild-type branches. The mutant branches scored homozygous (*hth-4/hth-4*), while the wild-type branch scored heterozygous (*HTH-4/hth-4*).

IGH can take the form of chimerism or mosaicism, distinguishable by their functional origin, relative frequency and degree of genetic change (Santelices, 2004). Mosaic individuals arise from an intrinsic genetic change, such as somatic mutations, mitotic recombination, changes in ploidy levels, or genome duplications (Santelices, 1999). Chimeric organisms differ in that they result from grafting or allogenic fusion (Santelices, 2004). IGH has been found in bacteria, protists, fungi, and plants, as well as invertebrate animals, such as cnidaria and tunicates, and vertebrates like marmosets, cats and humans. In the cases involving animals, IGH generally has either a neutral or detrimental effect (Pineda-Krch and Lehtilä, 2004). Genetic homogeneity has

traditionally been assumed for the majority of individual organisms. However, there is increasing evidence that IGH is more common than previously considered.

The developmental program of unitary organisms, such as vertebrates, is determinate and closed, and would therefore afford no advantage if IGH were common-place. In contrast, the developmental program of organisms, such as plants, is open-ended and can begin with either a single cell (zygote or stem cell) or multicellular stage (vegetative propagules) (Fagerström, 1998). Organisms with repeating basic structural units, called modules, allow the adult organism to have variable number of parts. Such organisms show low levels of differentiation, are developmentally plastic, and tend not to sequester their germ line (Pineda-Krch and Lehtilä, 2004). In fact, it has been reported that long-lived trees benefit from module-level selection driven by somatic mutations by increasing tree fitness and reducing local adaptation in the herbivore (Folse and Roughgarden, 2011). IGH could similarly benefit *Arabidopsis* as it has adopted an inbreeding reproductive strategy and thus has limited its adaptive potential.

Clones derived from differentiated somatic cell or nuclear founders are expected to be phenotypically and genotypically identical. However, contrary to expectation, phenotypic variation or somaclonal variation can be found in organisms regenerated from tissue culture and this variation is stable and can be passed down to the next generation. In animals, the phenotypic variation is thought to be due mostly to epigenetic reprogramming of gene expression (Humpherys et al., 2001). However, genome-wide studies of regenerated *Arabidopsis* lineages revealed a considerable elevation in DNA sequence mutation rates (Jiang et al., 2011) suggesting that DNA sequence mutations may in large part underlie the phenotypic (somaclonal) variation seen in plant tissue culture regenerants.

The molecular events underlying somaclonal variation have also been studied in plant species other than *Arabidopsis.* In *japonica* rice, a purple sheath mutation was recovered as a somaclonal mutant designated Z418. The original plant, C418, had a non-functioning OsC1allele due to a 34 base pair (bp) deletion in the gene, the candidate gene believed to control the purple sheath trait. In Z418 line there was a gain-of-function in the OsC1 gene. Sequence analyses determined that Z418 harbored an allele identical to the OsC1 sequence of the rice-coloured line T65 (Gao et al., 2011). The source of the novel DNA sequence has not yet been determined.

Two factors suggest that Arabidopsis plants might benefit from IGH. First, Arabidopsis plants undergo modular development and secondly, they predominantly reproduce by self-fertilization (inbreeding). However, the mechanism by which IGH occurs has not been experimentally verified. One possibility is that the genetic heterogeneity seen in the *hth* mutants could be a response to stress imposed by the loss of normal *HTH* gene function, as posited by Lolle et al (2005). The metabolic stress in this case could be analogous to the environmental induction used in the flax experiments described above. If genome sequence changes are induced in response to metabolic stress, plants experiencing any number of metabolic stresses might manifest genome changes. By extension, environmental factors that induce stress could also mobilize genome changes.

In Arabidopsis, adenosine (Ado 5' phospho-transferase) kinase (ADK; ATP: Ado, EC 2.7.1.20) is constitutively expressed in all cells. Loss of ADK interferes with adenylate pools, methylation and cytokinin interconversion (Moffatt et al., 2002). ADK phosphorylates adenosine (Ado) and Ado analogues (Schomberg and Stephan, 1997) and is a key enzyme in the purine salvage pathways for Ado. Recycling of Ado follows two principle routes wherein the direct route, ADK catalyzes the following reaction: ATP + Ado → ADP + AMP (Schomberg and Stephan, 1997). ADK is also involved in the interconversion of cytokinin (CK) ribosides (Burch and Stuchbury, 1987). In the indirect route, Ado is hydrolysed to adenine by Ado nucleosidase and is then converted to AMP by adenine phosphoribosyltransferase (Moffatt et al., 2002). Since CK nucleotides are less active than ribosides, ADK contributes to intracellular CK homeostasis by reducing the abundance of active CKs. *Arabidopsis* plants deficient in ADK activity have increased CK riboside levels as compared to the wild type (Moffatt et al. 2002).

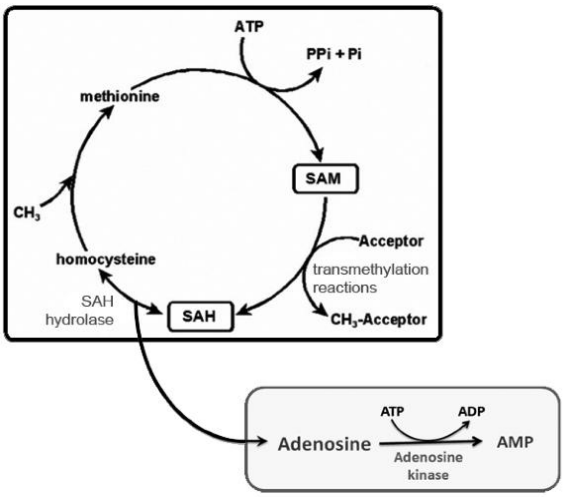


Figure 2: Role of ADK in methyl recycling via S-adenosyl-L-homo-cysteine (SAH). Adenosine kinase catalyzes the production of AMP through phosphorylation of adenosine

Figure 3: Arabidopsis thaliana representative ADK-deficient lines created by gene silencing. Phenotypes varied from wild-type-like (far right) to ADK deficient (far left).

Therefore the loss of ADK activity might impose a substantial metabolic stress on affected plants. Such plants have a wide range of phenotypic variation, spanning the full spectrum from plants suffering severe morphological abnormalities to plants that look relatively normal and are comparable to wild-type. ADK-silenced plants have curled and twisted leaves, floral malformations, and reduced primary shoot height (Figure 3; Moffatt et al. 2002). ADK deficient plants also senesce later and have increased leaf cell numbers (Schoor et al., 2011). However, about 30% of the plants are phenotypically wild-type and may be genetic revertants.

In addition to metabolic stress, plants experience environmental challenges and are exposed to a constantly changing suite of factors such as predators, pathogens and, for many parts of the world, drastically varying temperature ranges. Many biological processes are influenced by growth temperature, including photosynthesis, transpiration, and respiration. Tissue can be damaged by herbivores, pathogens, or by physical means. Plants can respond to mechanical wounding locally in damaged tissue, as well as, systemically at distal sites from the initial wounded area. Plant responses to environmental stimuli are regulated by several phytohormones, such as jasmonic acid (JA), salicylic acid, ethylene and absiscic acid.

JA and its bioactive derivatives regulate many protective responses to abiotic and biotic stress through large-scale changes in gene expression. JAs regulate many biological processes such as systemic wound responses (Koo et al., 2009), secondary metabolism (Gundlach et al., 1992), reproductive development (Browse, 2005) and growth control (Balbi

and Devoto, 2008). Considered a phytohormone, JA accumulates rapidly in tissues both proximal and distal to injury sites and has been shown to accumulate in wounded *Arabidopsis* leaves within 120s of wounding. Also, systemic signal displacement from wounded to unwounded leaves leads to accumulation of jasmonic acid in distal leaves. Responses to JA, such as the expression of some *JASMONATE-ZIM* domain genes, can take place within 15 minutes in unwounded leaves (Glauser et al., 2009).

Wounding, exposure to adverse environmental conditions such as extreme temperatures, excessive light, pollution, drought and salinity, can lead to the increased production and accumulation of damaging concentrations of reactive oxygen species (ROS), a process referred to as oxidative stress. ROS include compounds such as H_2O_2, superoxide anion, and hydroxyl radicals (Gechev et al., 2002). A defense mechanism against oxidative stress is the activation of the cell antioxidant system. Antioxidant enzymes, including glutathione reductase, catalases, peroxidases and superoxide dismutase, are often found to have elevated activities in stress-resistant plants. In non-toxic concentrations, H_2O_2 can act as a signalling molecule. Signalling cascades involve secondary messengers such as ROS, Ca^{2+}, and phosphatidic acid (Apel and Hirt, 2004; Testerink and Munnik, 2005). After mechanical wounding, there is an overlap of biotic and abiotic stress responsive plant genes (Fujita et al., 2006). Although phytohormone pathways are well studied, the mechanisms behind stress perception and initial signaling events are still not as well defined.

In addition to natural inducers, ROS can be mimicked by exposure to a number of different reagents. To study the effect of oxidative stress, 3-amino-1,2,4-triazole (AT) (Gechev et al., 2002), a catalase inhibitor, has been used and is known to elevate H_2O_2 levels. Buthionine sulfoximine (BSO) (Griffith and Meister, 1979), a γ-glutamylcysteinyl synthetase inhibitor, when used in conjunction with AT, suppresses increased glutathione synthesis.

The goal of this project was to test whether IGH could be induced in certain conditions. This project tested the hypotheses that metabolic stress, in the form of ADK deficiency and environmental stress in the forms of cold stress, mechanical damage, and ROS exposure, induce IGH. Transgenic markers and molecular markers polymorphisms between the Columbia and Landsberg *erecta* genetic backgrounds were used to genotype *Arabidopsis* plants and test for IGH. Also presented here is work done by Chris Hammill investigating the effect of tissue culture on restoration frequencies. Regenerated plants as well as

undifferentiated callus derived from the wild-type hybrid lines used in the mechanical damage studies were genotyped and assayed for sequence reversions. Contrary to expectation, the result of the study revealed that the frequency of IGH was reduced when plants experienced mechanical wounding or ROS exposure in the genetic lines tested here.

Materials and Methods

Construction of Plant Lines

amiADK hybrid lines: the amiADK 7-7 line was generated in the Moffatt lab using the Columbia accession. amiADK 7-7 transgenic plants harbour a construct encoding an artificial microADK designed to produce a microRNA targeting the ADK genes (Appendix A, Figure 21). The amiADK 7-7 plants were crossed with WT Ler plants and F1 seeds from the amiADK 7-7 ♀ and Ler ♂ cross were grown to maturity. F1 plants were allowed to self-fertilize. F2 plants harbouring the transgene were identified as glufosinate resistant. Plants were sprayed with glufosinate solution (200µM) once per day over a period of 10 days. Resistant plants were screened for the presence of the amiADK transgene with PCR using the primers, pSAT-F and pXCS-R (Appendix C, Table 15). The progeny from a transgene negative F2 plant were used for assessing temperature effects on genomic stability.

Hybrid ADK lines (Hadk): the ADK1-GFP line was generated as described in Schoor et al. (2011) (Appendix A, Figure 22). True breeding ADK1-GFP plants with weak to strongly silenced ADK phenotypes were crossed with WT Ler plants (Figure 4). Five F1 plants were allowed to self-fertilize and were used to generate the F3 generation of seeds (summarized in Table 1). Transparent mylar sheets were wrapped around individual plants to minimize out-crossing. F1 plants were genotyped using the indel primers (Appendix C, Table 13) and examined for eGFP fluorescence. F2 plants were genotyped and several plants from each lineage negative for the transgene were used to produce the F3 generation. The presence or absence of the transgene was determined by PCR using the following primer pairs (adktestp-RF and EGFPm-R, and adktestp-RF and adktestp-R; Appendix C, Table 15). The F3 progeny from transgene negative F2 plants were screened using PCR and tissue fluorescence.

Table 1: Hybrid ADK1-GFP x Ler F1 Lines

F1 label	Pollen recipient ♀	Pollen donor ♂
Hadk1	WT Ler #1	ADK1-GFP #1
Hadk2	WT Ler #3	ADK1-GFP #3
Hadk3	WT Ler #4	ADK1-GFP #4
Hadk4	ADK1-GFP #3	WT Ler #3
Hadk5	ADK1-GFP #2	WT Ler #1

Figure 4: Hadk parental plants.

Wild-type hybrid lines: wild-type Col and wild-type Ler plants were crossed to generate F1 hybrid lines. Seeds were collected from two F1 plants from Col♀ x Ler♂ and two F1 plants from the recipricol cross, Ler♀ x Col ♂. F2 to F4 generations were self-fertilized. Each plant used in this study was genotyped using marker specific primer sets (Appendix C, Table 13) to verify genotype and lineage.

Naming Convention

Lines were named in the format demonstrated in Figure 5. The initial cross is indicated first, with the pollen recipient (♀) preceding the pollen donor (♂). The descendent lineage is then indicated with the plant number in each generation starting from F1 to the current generation. Dashes are used to separate each generation.

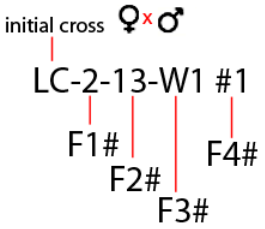


Figure 5: Naming convention for hybrid lines

Lami = Landsberg *erecta* x amiADK 7-7

Hadk = hybrid ADK1-GFP (see Table 1)

LC = Landsberg *erecta* x Columbia

W (for F3 generation) = wounding treatment experimental group

C (for F3 generation) = control, no treatment group

Treatments

Cold stress: amiADK hybrid F3 seedlings from the line amiL-5-1 were grown at 21^{o}C (no treatment), 15^{o}C and 4^{o}C. The plants grown at 4^{o}C were grown in a refrigerator with a glass door, unlike the plants grown at the other two temperatures, which were grown in growth chambers.

Mechanical wounding: wild-type hybrid F3 plants from each F2 progenitor were divided into two groups: wounding (experimental treatment) and no treatment (control). Each group consisted of 15 plants grown in individual 5.5 cm pots. Upon inflorescence emergence, approximately 2 weeks following germination, plants were mechanically damaged by pinching the leaves with ribbed forceps. Approximately 50% of the leaf surface was damaged and inflorescence buds excised.

ROS exposure: the F4 seedling populations derived from wild-type hybrid control groups used in the wounding experiment were used to test the effect of ROS inducing media: ½ MS agar + 40 µM BSO (Sigma) + 2 µM AT (Sigma).

Growth Conditions

Plants and callus tissue were grown 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), unless otherwise stated. Growth chambers were illuminated by both incandescent and fluorescent lights (approximately 140 - 170 µmol m^{-2} sec^{-1} at sample level) with a 16 hour light cycle, 8 hour dark. The ROS plates were placed in a growth chamber with approximately 80-100 µmol m^{-2} sec^{-1} illumination. For callus growth requiring 24 hours of dark, light was blocked by wrapping plates in aluminum foil. All seeds were cold stratified by placing seeds (on soil or media) in a 4^{o}C environment for 2 to 3 days prior to transfer to the growth chamber. Transparent mylar sheets were used for the following plants: amiADK hybrid F2s, Hadk parents, F1s and F2s, and wild-type hybrid parents, F1s, and F2s. All plants were grown at 21^{o}C with the following exceptions: Hadk plants were grown at 19^{o}C and two groups of amiADK hybrid F3 plants were grown at 15^{o}C or 4^{o}C (see section on treatments).

Parent, F1 and F2, and wt hybrid F3 plants were sown directly onto soil, with the exception of the amiADK hybrid F1 plants, which were first germinated on ½ MS agar then transplanted onto soil. In general, F3 and F4 seedlings used for molecular genotyping were grown on ½ MS agar with plates oriented vertically to promote root growth along the surface of the agar. However, the ROS inducing plates were grown on a 45^{o} angle while plants grown for qPCR genotyping were grown horizontally. Approximately 20-30 seeds were distributed on each plate and genotyped using the indel primers (Appendix C, Table 13). Figure 6 is an example of a plate with seedling growth. Plants used for qPCR genotyping were grown at a density of approximately 9-12 seeds per plate.

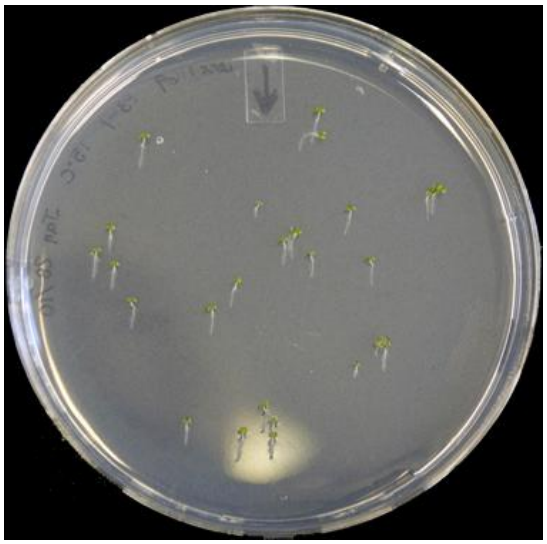


Figure 6: amiADK x Ler F3 14 day old seedlings grown on half MS media plates. Seedlings grew along the surface of the plates in one orientation due to the vertical placement of the plates.

All seeds sown on media were surface sterilized using chlorine gas prior to plating. Seeds were sterilized in 1.7ml microcentrifuge tubesfilled up to the 0.1 ml mark. Uncapped tubes are placed in a glass container. To create chlorine gas, a beaker, placed inside the glass container, was filled with 100 ml household bleach and 4 ml of concentrated HCl. The container was sealed and seeds are exposed to gas for 1 to 2 hours. After sterilization the tube was sealed and the seeds stored at room temperature.

Tissue Sampling

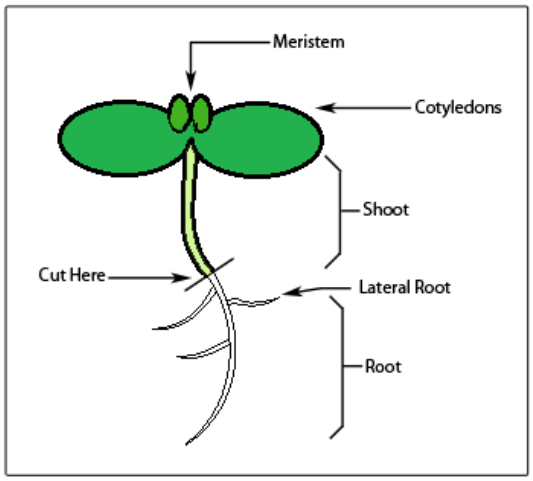


Figure 7: Seedling dissection diagram. Seedlings were bisected at the junction between the root and shoot portions.

Seedling dissections: amiL-5-1 F3 seedlings grown at the three different temperatures were grown to until cotyledons had fully opened. This stage was achieved approximately 7-10 days after transfer to the growth chambers for the 21^{o}C environment, 10-14 days for the 15^{o}C environment, and 18-21 days for the 4^{o}C environment. Once the correct growth stage was reached, root and shoot were separated by bisecting seedlings (Figure 7). The root and shoot were collected into separate 1.7 ml microcentrifuge tubes and placed on ice.

Sampling tissue from adult plants: samples from rosette leaves (approximately 0.5 cm^2), cauline leaves, or flowers were taken from the plants during the mid-flowering growth stage (days post germination?). For the majority of the collections, at least two samples were taken from each plant. One sample was used for DNA extraction and the other was stored at -20^{o}C, with a few exceptions. For nine F2 plants from the amiADK 7-7 hybrid line Lami-1, one rosette sample and one sample from each floral branch was collected. Also, two samples were taken from the rosette of the wild-type hybrid F3 plants (control and experimental) prior to wounding treatment. After bolting, tissue was collected from each inflorescence branch and pooled for each individual plant.

Whole seedling collection: seedlings were grown for 10-14 days on ½ MS agar and collected individually into microcentrifuge tubes for DNA extraction for the following populations: Hadk F3s, wild-type hybrid F4s (all treatment groups) and progeny from tissue culture regenerated plants.

Sample collection for qPCR : F4 plants were grown for approximately 3 weeks, harvested individually and place into 2 ml screw cap microcentrifuge tubes for DNA extraction. Plants belong to the line LC-2-13 and were taken from 10 no treatment

populations, and 10 wounding populations. Twenty seedlings from each population in the experimental and control groups were analyzed.

Tissue culture: cotyledons were collected as duplicate samples for each seedling that was used for callus induction. One cotyledon was used for DNA preparation and the other set aside as a reserve sample. At the time of transfer to shoot induction media (SIM), samples were taken in duplicate from callus material.

DNA Extraction

Crude DNA extraction method: the following method was adapted from Edwards *et al.* (1991). Tissue was ground in 50 µl TE buffer (10mM Tris-HCl pH 8.0, 1 mM EDTA) in microcentrifuge tubes using a disposable plastic pestles. 350 µl of extraction buffer (200 mM Tris-HCl pH 7.5, 250 mM NaCl, 25mM EDTA, 0.5% SDS from Edwards *et al.* 1991) was added to each sample. After mixing the sample using a vortex set at maximum speed, the sample was centrifuged for 1 minute at maximum speed. 300 µl of supernatant was transferred to new tubes then 300 µl of isopropanol was added. The samples were mixed by vortexing and incubated at room temperature for 2 minutes. Tubes were centrifuged for 5 minutes at maximum speed. Pellets were air dried for 10-15 minutes before being re-suspended in 100 µl TE buffer and stored at -20^{o}C.

DNA extraction from plants used in qPCR: whole plants were collected individually into 2 ml screw cap microcentrifuge tubes and 6 stainless steel ball bearings added (1/8" diameter; Abbott Ball Company, West Hartford, Ct, USA). The tube and tissue were flash frozen in liquid nitrogen. Cells were disrupted using a vortex set at max speed and kept in a frozen state using liquid nitrogen. Following tissue homogenization, 600 µl of extraction buffer (0.1M Tris-HCl pH 8.2, 0.05M EDTA, 0.1M NaCl, 2% SDS, 0.5 mg/ml Proteinase K) was added to each tube. Samples were mixed on a rocker for 30 minutes at room temperature. Each sample was then transferred to a fresh 1.7 ml microcentrifuge tube and centrifuged for 5 minutes at maximum speed. The supernatant was transferred into new tubes, 2 ul of 10 mg/ml RNase A added to each sample and samples incubated at 37^{o}C for 15 minutes. 500 µl of phenol:chloroform:isoamyl alcohol (25:24:1) mixture was added to each tube. Samples were then rocked at room temperature for 15 minutes. Tubes were centrifuged at max speed for 5 minutes. The upper layer was transferred to a new tube and 1/10th the

volume of 3M sodium acetate pH 5.2 and 1 volume isopropanol was added before mixing by inversion. The samples were again centrifuged for 5 minutes at max speed to pellet the DNA. The pellet was resuspended in 500 µl TE buffer and 50 µl 3M sodium acetate pH 5.2. 500 µl of ice cold ethanol was added and the tubes were mixed by inversion. Samples were centrifuged for 10 minutes at maximum speed to pellet the DNA. The supernatant was removed by pipetting and the pellet allowed to air dry for 15 minutes before additional drying in a 50^{o}C heat block for approximately 3 minutes. DNA was resuspended in 50µl TE buffer at 4^{o}C for 16-24h before storage at -20^{o}C.

DAB Staining

Hydrogen peroxide production in plant tissue was visualized using a 3,3'-diaminobenzidine (DAB; Sigma-Aldrich D8001) stain. Whole seedlings and excised leaves were submerged in a solution of 5 mM DAB in 50 mM Tris-acetate pH 3.8. Tissues were stained for 20h in dark at room temperature. After staining, tissue was decolorized for 10 minutes with 95% ethanol at 70^{o}C. Tissue was rinsed and stored in 95% ethanol.

Tissue Culture

This work was undertaken by Chris Hammill as part of his Biol499 project. F4 seeds from several wild-type hybrid lines used in the mechanical wounding and ROS experiments were also used in this project (LC-2-13-[C2, C4, C5, C6 and W1]). Seedlings were grown for 1-2 weeks before dissection and transfer to callus induction media (see Growth Conditions section).

Initial callus induction: roots were cut into 2-3mm segments and transferred to callus inducing media (CIM; base media: 3.2 g/L Gamborg's B5 vitamins with minimal organics Sigma-Alderich Canada ltd., Oakville, ON, 20 g/L d-glucose, 0.5 g/L MES, and 3 g/L phytagel). The base media was supplemented with 500 µg/L 2,4-D (auxin) and 50 µg/L kinetin (cytokinin). Roots were allowed to develop calli for one month before first transfer.

Callus maintenance: To maintain callus size and freshness, callus samples from each line were transferred to fresh CIM every 3 weeks. Transferred callus tissue ranged in size

from 5mm to 20mm in diameter. The transfers were repeated 5 times over the course of these experiments.

Plant regeneration: small masses of callus tissue were transferred to shoot induction media (SIM) to induce shoot formation. The base SIM was identical to the CIM except for the addition of 93 ug/L naphthalene acetic acid (auxin) and 894 ug/L N6-Δ2-isopentenyladenine (cytokinin).

Rooting and transfer to soil: Callus tissue exhibiting shoot formation was transferred to hormone free media to induce root formation. Plantlets were maintained in culture for 1 to 2 weeks prior to transfer to soil. Successfully rooted plantlets were transfered to 1:1 mixture of LC1:LG3 Sungro Sunshine potting mixes, (Sungro Horticulture, Seba Beach, AB). Freshly transferred seedlings were covered with a plastic dome or bag as protection from humidity shock. Seedlings underwent gradual dehumidification by removing the plastic covering over the course of a week. The soil was kept moist and plantlets were watered as needed.

Molecular Genotyping

Insertion-deletion polymorphisms (indels) of the Col and Ler accession were used for molecular genotyping. Sixteen markers with indel sizes ranging from 45-94 bp were chosen where the alleles in Col are insertions and the alleles in Ler are deletions (Appendix C, Figure 23). For each marker PCR primers were designed to amplify genomic regions flanking the indel (Appendix C, Table 13, Figure 23). The PCR program used for amplification was as follows: 94^{o}C for 2 min, 55^{o}C for 15 sec, 72^{o}C for 30 sec, and 39 cycles of 94^{o}C for 15 sec, 55^{o}C for 15 sec, 72^{o}C for 30s. Taq DNA polymerase purified from recombinant *E. coli* stocks was used in NH Buffer with 300 µM of each primer and 400 µM dNTPs for amplification (10x NH Buffer: 200 mM Tris-HCl pH 8.8, 50 mM $MgCl_2$, 100 mM KCl, 100 mM $(NH_4)_2SO_4$). The genotype was determined by size separation of the PCR products using agarose gel electrophoresis (Appedix C, Figure 23). Samples were subsequently genotyped using only those markers that were homozygous for the deletion (harbouring the Ler allele) in the respective parent.

Quantitative PCR

Three indel sites were examined: T14G11, F23M2 and T6H20. Primers sets were chosen that would amplify a region flanking the indel, either upstream or downstream of indel sites, termed the external reference sequence (Appendix C, Table 14). These sequences are common in both Ler and Col accessions and were used to create a baseline for relative genomic copies of the region next to the indel locus in a given sample. Additional primer sets were designed with one primer internal to the Col insertion sequence of the indel. In this scenario, only those sequences with the Col insertion sequence are amplified in qPCR. Each set of primers amplified a region approximately 100-300 bp in length.

qPCR was done using Bio-Rad Real-Time thermal cycler CFX96 using the CFX Manager software. DNA samples (6 ng/sample, 3 ng/sample for plants grown on ROS inducing media) from 10 individuals were pooled prior to amplification. Each pooled DNA sample was run in three technical replicates. Samples were also assayed individually. The reaction set-up for samples was as follows: SsoFast EvaGreen Supermix (Bio Rad), 0.5 µM reverse primer, 0.5 µM forward primer, 60 ng of template (30 ng for those plants grown on ROS inducing media), and water to a total volume of 10 µL. The following qPCR program was used: 98^{o}C for 2 min, 39x[98^{o}C for 2 sec, 60^{o}C for 5 sec + plate read], Melt Curve analysis 60 to 95^{o}C, increment 0.1^{o}C for 10 sec + plate read. Data were analyzed using Bio Rad CFX Manager Software version 1.5.

Each qPCR run included standards consisting of serial dilutions of known DNA concentration and composition. This DNA was made of linear pieces of 700 to 900 bp purified amplicon that contained the indel region for the marker of interest and would therefore amplify in qPCR. The primers used for synthesizing the standards for the respective indels can be found in Appendix C, Table 17. Ten-fold dilutions of standard DNA in TE buffer ranging from 5 x 10^{-11} ng to 5 x 10^{-17} ng were used with the external reference sequence primer sets. Dilutions ranging from 5 x 10^{-12} ng to 5 x 10^{-18} ng were used with the internal primer sets. Each standard dilution was run in technical duplicate. A ‘no template’ control was included in technical duplicate to test for DNA contamination. In some of the runs a Ler control (10 ng) was included. Following amplification, the qPCR products were size separated and visualized using agarose gel electrophoresis.

DNA quantification and dilutions: DNA was quantified using a spectrophotometer (NanoDrop 1000 3.7.1 Thermo Fisher Scientific Inc). Dilutions of 20 ng/µl of DNA were made for each sample. Aliquots of 20 ng/µl for 10 pooled DNA samples (from the same population) were made for the plants grown on ½ MS agar, so that each aliquot needed to be thawed only once. Due to the lower amount of recovered DNA for plants grown on ROS inducing media, aliquots for these plants were diluted to 10 ng/µl.

Copy number calculations: The copy numbers of DNA sequences for the serial dilutions of standard DNA were calculated using the following equations with the Microsoft Office Excel 2007 software.

$$grams\ per\ mole\ of\ product = product\ length \times 330\ grams\ per\ mol\ per\ base \times 2$$

$$mass\ of\ one\ molecule\ of\ product = \frac{\frac{g}{mol} of\ product}{Avagadro's\ number}$$

$$copy\ \#\ in\ diluted\ standards = \frac{g\ of\ product}{mass\ of\ one\ molecule\ of\ product}$$

The copy number in the standards were plotted against the $C_{(t)}$ values generated by the CFX Manager software using an automatically generated single baseline threshold set using a baseline subtractive curve fit. An exponential trendline was obtained using the Excel software. This equation was used to calculate the copy numbers for the experimental samples using the average $C_{(t)}$ value of the technical replicates. Each qPCR run included a set of serial dilution standards against which the experimental samples are compared.

DNA Sequencing

Sequencing was performed at The Centre for Applied Genomics, The Hospital for Sick Children, Toronto, Canada. The indel regions T14G11, F23M2, and T6H20, were sequenced using the primers in Appendix C, Table 16. Sequencing was done directly on purified PCR products. A small portion of the PCR products were run on an agarose gel to verify amplification and product size prior to sequencing. PCR products were purified using the EZ-10 SPIN Column PCR Products Purification Kit (Bio Basic Inc.). The PCR program used to amplify products is as follows: 94^{o}C for 2 minutes, 35x [94^{o}C for 30 sec, 52^{o}C for 30 sec, 72^{o}C for 50 sec], 72^{o}C for 7 min., hold at 4 ^{o}C. The extension time at 72^{o}C in the cycle repeated 35x was increased to 80 sec for amplification of LC-2-13 #W1 DNA. 1 unit of Tsg DNA polymerase (Bio Basic Inc.) was used per reaction in the following reaction mixture: 10x Tsg Buffer (Bio Basic Inc.), 3 mM $MgSO_4$, 200 µM dNTPs, and 200 µM of each primer.

Results

The goal of this project was to investigate factors that could promote intraorganismal genetic heterogeneity (IGH) in Arabidopsis plants. This was done by screening for and identifying mosaic *Arabidopsis* plants with the specific goal of identifying marker changes that might have arisen as a result of the reacquisition of non-parental alleles or genomic restoration events. ADK silencing, temperature stress, mechanical wounding and ROS induction were examined as possible environmental triggers. Size-based PCR genotyping was used to assay genetic changes at selected genomic loci in combination with quantitative PCR and screening of fluorescent markers. Two transgenic lines that induced ADK silencing through use of an artificial microRNA (amiADK) or with eGFP expression (ADK::eGFP) were used. The temperature stress was applied to amiADK hybrid F3 seedlings. Wounding and ROS stress was applied to the wild-type hybrid plants and both seedlings and adult plants were genotyped. Wild-type hybrid lineages were derived from crosses made between the Columbia and Landsberg accessions.

Assessing temperature effects on genome stability

Preliminary experiments using the amiADK x Ler line revealed one F3 seedling with a mosaic profile (S. Lolle unpublished results). The seedling (#34) had a homozygous Ler root profile and a heterozygous shoot profile for the indel F12K11 (Figure 8). Its parent was homozygous Ler at this marker.

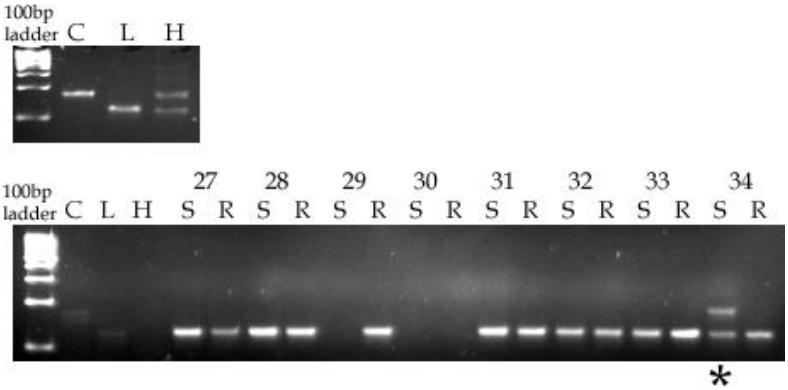


Figure 8: Digital photograph of an agarose gel showing DNA PCR-amplified from Ler x amiADK F3 seedlings (F2#11), using primers for marker F12K11. Information: Ladder – GeneRuler 100bp Plus DNA Ladder; Col (C); Ler (L); Col x Ler hybrid (H); root sample (R); shoot sample (S); 4% TAE agarose gel. The top gel is a photograph of homozygous Col, Ler and heterozygous alleles.

To follow up on this initial finding additional seedlings descended from Columbia-Landsberg hybrids of amiADK line, amiL-5-1, were grown at three temperatures (21˚C, 15˚C and 4˚C), bisected at the root-shoot junction and DNA preparations made for each. The quality of the recovered DNA sample from each sample was variable, with shoot samples being especially problematic. Although in total there were 105 seedlings in each group (103 in the 4^{o}C group), only a portion (shown in Table 2) had good PCR amplification for both the root and shoot preparations. As molecular profiles were being compared between organ systems for individual seedlings, only those samples producing scorable bands following the agarose gel electrophoresis for both the root and shoot were included in the final analyses or tally (Table 2).

Table 2: Summary of total number of seedlings assayed at the temperatures shown using the molecular markers indicated.

	Number of root and shoot samples		
	Growth Temperature		
Marker	**21^{o}C**	**15^{o}C**	**4^{o}C**
F12K11	99	98	64
F23M2	89	92	97
T6A23	105	88	15
T11I18	29	30	N/A
F8D20	82	85	92
MSA6	77	27	15

For the majority of the samples, no deviation from the expected genotype of was found. However, there were two individuals that deviated from the parental genotype in the group grown at 21^{o}C. The seedling, designated amiL-5-1 #7 was found to have a profile different from the F2 parent (designated amiL-5 #1, see Table 4), although the root and shoot had the same profile. The second seedling, (amiL-5-5 #55) had discordant shoot and root profiles for marker MSA6.

Table 3: Summary of molecular profiles for amiL-5-1 #7 and #55 and their F2 parent amiL-5 #1

Marker	**F12K11**		**F23M2**		**T6A23**		**T11I18**		**MSA6**		**F8D20**	
Plant	**R**	**S**	**R**	**S**	**R**	**S**	**R**	**S**	**R**	**S**	**R**	**S**
(F2) amiL-5 #1	L		L		L		L		H		L	
amiL-5-1 #7 (F3)	C	C	C	C	H	H	L	L	C	C	L	L
amiL-5-1 #55 (F3)	L	L	L	L	L	L	NA	NA	C	H	L	L

The marker MSA6 which was heterozygous in the F2 parent, was also used for PCR profiling. For MSA6, the root section of seedling #55 in the 21°C group was found to be homozygous Col but the shoot was heterozygous (Table 3, Figure 9).

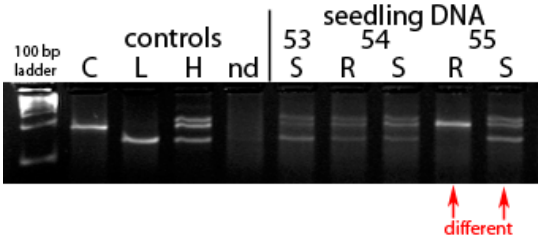


Figure 9: DNA PCR-amplified shoot and root samples obtained from individual F3 seedlings from the line amiL-5-1 grown at 21°C for marker MSA6 on agarose gel. Homozygous Columbia (C), Landsberg *erecta* (L), heterozygous (H), and no template (nd) controls are to the left. The F2 parental plant (amiL-5#1) is heterozygous at MSA6 (see table 4). Shown are root (R) and shoot (S) PCR products for F3 progeny seedlings #53, 54 and 55. The root and shoot profile for seedling #55 differs (red arrows).

Surveying ADK transgenic adult plants for mosaicism

Figure 10: Range of phenotypes found in Lami-1 F2 plants. Phenotypes range from severely ADK-deficient (#18) to wild-type-like (#23).

DNA samples from nine flowering F2 plants (Lami-1 lineage) were profiled using the 16 indel markers. There was a range of phenotypes from severely ADK deficient to wild-type-like (Figure 10). No marker deviations were found for any of the nine plants tested. The results are summarized in Table 4.

Table 4: Summary of PCR genotyping results for the 9 Lami-1 F2 plants sampled.

Lami-1 F2 plant #	# of samples per plant	# of markers in profile (Appendix C, Table #)	# of confirmed sequence changes
16	6	all 16	none
17	6	all 16	
18	15	all 16	
19	5	all 16	
20	13	all 16	
21	12	all 16	
22	10	14 (excluding F4C21 and F16J13)	
23	10	14 (excluding F4C21 and F16J13)	
24	12	14 (excluding F4C21 and F16J13)	

Phenotypic assessment of genome stability using an ADK1::GFP fluorescent line

Fluorescence in the seedlings of the control ADK1-GFP transgenic lines was found to be strong in epidermal and vascular tissue. The most noticeable fluorescence was seen in leaf epidermal cells (Figure 11A). The pattern of eGFP fluorescence is different from auto-fluorescence of the underlying chloroplasts. Cells near damaged tissue tended to auto-fluoresce at the same wavelength of eGFP. The fluorescence in epidermal cells was used as a guide for screening for eGFP fluorescence in non-transgenic segregant F3 seedlings (derived from F2 parents that had not inherited the transgene). One seedling (Hadk4-1-1, Table 6) was positive for eGFP fluorescence. The shoot tissue as well as root tissue expressed eGFP fluorescence in patterns reminiscent of the fluorescent controls (Figure 12).

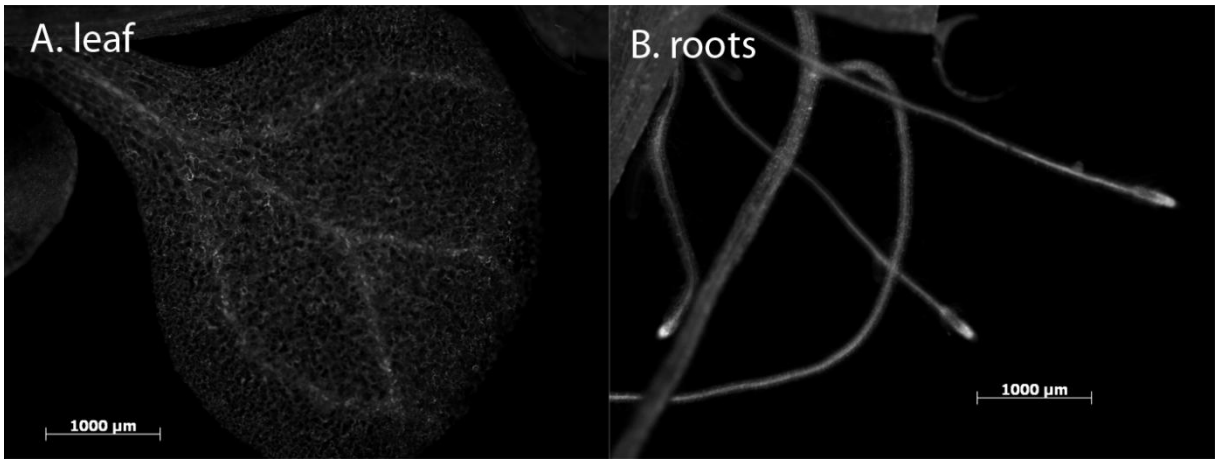


Figure 11: Transgenic A. thaliana (ADK1-GFP) seedling expressing eGFP fluorescence in leaf tissue (A) and root tissue (B).

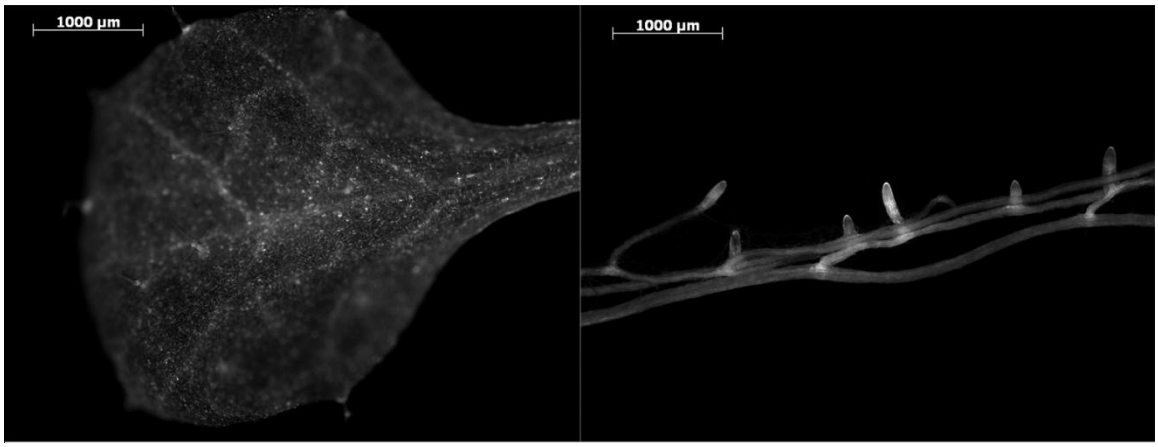


Table 12: Hadk4-1-1 fluorescent seedling eGFP expression in leaf (left) and roots (right).

DNA was isolated from the non-transgenic F3 seedlings and used in size-based PCR genotyping. The summary of total number of F3 seedling samples profiled for each of the progeny populations can be found in Table 5. Also included in the table are the total numbers of markers assessed in the profile. One out of 94 of the non-fluorescent seedlings, designated Hadk1-1-3 #12, had four heterozygous markers whereas the parent was homozygous for the deletion or insertion at these same loci. Four other homozygous markers, however, remained unchanged. The seedling's profile as well as the F2 parent of that particular F3 population is found in Table 6.

Table 5: Summary of PCR genotyping for Hadk F3 plants

Hadk F3 population	**# of F3 seedlings genotyped**	**Markers assessed** (Appendix C, Table #)	**# F3 seedlings with differing profiles**
1-1-3	94	F12K11, F15H11, MSA6, T6H20, MNJ8, MGI19	1
1-1-5	96	F23M2, T11I18, MSA6, MNJ8, MGI19	0
2-3-2	92	F12K11, T6A23, MSA6, F2P16, MNJ8, MGI19	0
2-3-5	96	F23M2, T14G11, T6A23, T11I18, MSA6, T6H20, F4C21, F16J13, MGI19	0
3-1-1	96	F23M2, T6A23, MGI19	0
3-1-5	96	F12K11, F8D20, F2P16, MNJ8, MGI19	0
4-1-1	96	F15H11, T11I18, MSA6, F4C21, F16J13, F8D20, MNJ8, MGI19	1
4-1-3	96	F12K11, F2P16	0
5-1-1	96	F5J5, F6D8, F15H11, F23M2, F2P16, MNJ8, MGI19	0
5-1-2	96	F23M2, T14G11, T6A23, MNJ8	0

One seedling out of 96 in the Hadk4-1-1 population was found to be fluorescent. DNA was isolated from one leaf of that fluorescent plant and used for molecular genotyping. The results are summarized in Table 6. The profile of the fluorescent seedling does not match that of the F2 parent. Five homozygous parental markers were heterozygous for the fluorescent F3 seedling whereas 7 other homozygous markers remained unchanged. This fluorescent seedling was grown to maturity and its seeds were collected.

Table 6: Summary of discordant F3 marker profiles (highlighted) and their corresponding F2 parent.

Plant	**Marker Profile**															
	F12K11	**F5J5**	**F6D8**	**F15H11**	**F23M2**	**T14G11**	**T6A23**	**T11I18**	**MSA6**	**T6H20**	**F4C21**	**F16J13**	**F8D20**	**F2P16**	**MNJ8**	**MGI19**
Hadk1-1 #3 (F2)	L	H	H	L	H	H	C	H	L	L	C	H	H	H	L	L
Hadk1-1-3 #12 (F3)	H	H	H	L	H	L	H	H	H	L	H	L	H	H	L	L
Hadk4-1 #1 (F2)	H	C	C	L	H	C	C	L	L	H	L	L	L	H	L	L
Hadk4-1-1 (F3) fluorescent seedling	L	H	C	L	L	C	C	L	L	C	H	H	H	H	H	L

Assessing the effect of mechanical wounding and ROS exposure on genome stability

Five to eleven tissue samples were harvested from 10 F2 wild-type adult hybrid plants. The rosette leaf sample was profiled using for all 16 indel markers while the remaining samples were profiled using only those markers that scored as homozygous for the deletion (the Ler allele) based on the profile obtained from the rosette leaf. A summary of the results can be found in Table 7. No marker differences were detected for any of the plants.

Table 7: Molecular genotyping of wild-type hybrid F2 plants for detection of mosaicism

F2 Line	Plant #	# of samples	Homozygous Ler markers assessed (Appendix C, Table #)	# of plants with sequence differences
CL-2	11	8	F12K11, MSA6, T6H20, F16J13	0
	12	7	F5J5, F6D8, MSA6, F4C21, F8D20, F2P16, MNJ8	
LC-2	12	8	T6H20, F8D20	
	13	10	F23M2, T14G11, T6A23, T11I18, MSA6, T6H20	
	14	9	F2P16	
CL-3	12	11	F12K11, F23M2, T14G11, T6A23, F8D20	
	13	11	F16J13, F8D20	
	14	7	F6D8, T11I18, MSA6, T6H20	
LC-3	11	7	F12K11	
	12	5	F12K11, F8D20	

Five populations of F3 plants derived from the above F2 plants (75 plants in total) were mechanically damaged as described in the Materials and Methods section. Despite being wounded, plants recovered well and, based on visual inspection, did not seem to be adversely affected by the treatment. Tissue samples taken before wounding and samples taken from each new shoot were genotyped using markers that were homozygous for the deletion in the previous generation. No marker differences were identified between the samples obtained before and after wounding. The summary of the genotyping data can be found in Table 8.

Table 8: Summary of genotyping results for wild-type hybrid F3 plants

Line	Markers assessed	Treatment	# of plants with sequence differences
CL-2-11	F12K11, MSA6, T6H20, F16J13	No treatment	0
		Wounding	
CL-2-12	F5J5, F6D8, MSA6, F4C21, F8D20, F2P16, MNJ8	No treatment	
		Wounding	
LC-2-13	F23M2, T14G11, T6A23, T11I18, MSA6, T6H20	No treatment	
		Wounding	
CL-3-12	F12K11, F23M2, T14G11, T6A23, F8D20	No treatment	
		Wounding	
CL-3-14	F6D8, T11I18, MSA6, T6H20	No treatment	
		Wounding	

Progeny from three of the lines in Table 8 were genotyped using markers that were homozygous Ler in their respective F3 parent. F4 progeny from both the wounding and no-treatment F3 groups were evaluated. No marker changes were found in either group (Table 9).

To evaluate the effect of ROS inducting media, F4 seedlings from F3 plants in the no treatment groups were grown in Petri plates with ROS inducing substrate. DNA was prepared from whole seedlings and genotyped using indel markers (Appendix C, Table 13). A summary of the results for the F4 seedlings is found in Table 9. With one exception (LC-2-13-C13 #48), no marker differences were found in any of the populations. LC-2-13-C13 #48 was grown on ROS inducing media. This seedling was found to be homozygous Col at marker F8D20 and was the only marker out of 16 that did not match that of #48's parent, LC-2-13 #C13.

Table 9: Summary of genotyping results for wild-type hybrid F4 seedlings grown on ½ MS or ROS inducing media.

Treatment	Line	# of seedlings genotyped	Markers assessed	# seedlings with sequence changes
No Treatment	CL-2-12-C5	100	F5J5, F6D8, F15H11, MSA6, F4C21, F16J13, F8D20, F2P16	0
	CL-2-12-C15	89	F5J5, F6D8, T14G11, T6A23, MSA6, T6H20, F4C21, F8D20, F2P16, MNJ8	0
	LC-2-13-C1	100	F23M2, T14G11, T6A23, T11I18, MSA6, T6H20, MGI19	0
	LC-2-13-C13	100	F23M2, T14G11, T6A23, T11I18, MSA6, T6H20, F8D20, F2P16	0
	CL-3-12-C5	100	F12K11, F6D8, F15H11, F23M2, T14G11, T6A23, F8D20	0
	CL-3-12-C15	100	F12K11, F23M2, T14G11, T6A23, F8D20, F2P16	0
	LC-3-2-C1	95	F6D8, F15H11, F23M2, T14G11, T6A23, F4C21, MGI19	0
Wounded	CL-2-12-W5	100	F5J5, F6D8, F15H11, MSA6, F4C21, F8D20, F2P16, MNJ8	0
	CL-2-12-W15	100	F5J5, F6D8, T6A23, MSA6, F4C21, F16J13, F8D20, F2P16, MNJ8	0
	LC-2-13-W8	100	F23M2, T14G11, T6A23, T11I18, MSA6, T6H20, F2P16, MNJ8	0
	LC-2-13-W10	106	F15H11, F23M2, T14G11, T6A23, T11I18, MSA6, T6H20, F16J13, F2P16, MNJ8	0
	CL-3-12-W7	100	F12K11, F6D, F15H11, F23M2, T14G11, T6A23, T11I18, MSA6, F8D20	0
	CL-3-12-W11	100	F12K11, F23M2, T14G11, T6A23, F16J13, F8D20	0
	LC-3-2-W1	96	F6D8, F15H11, F23M2, T14G11, T6A23, F4C21, MGI19	0
	LC-3-2-W2	96	F6D8, F15H11, F23M2, T14G11, T6A23, F4C21, F16J13, MGI19	0
ROS	CL-2-12-C5	100	F5J5, F6D8, F15H11, MSA6, F4C21, F16J13, F8D20, F2P16	0
	CL-2-12-C15	96	F5J5, F6D8, T14G11, T6A23, MSA6, T6H20, F4C21, F8D20, F2P16, MNJ8	0
	LC-2-13-C1	96	F23M2, T6A23, T11I18, MSA6, T6H20, MGI19	0
	LC-2-13-C13	96	F23M2, T6A23, T11I18, MSA6, T6H20, F8D20, F2P16	1
	CL-3-12-C5	96	F12K11, F6D8, F15H11, F23M2, T6A23, F8D20	0
	CL-3-12-C15	96	F12K11, F23M2, T6A23, F8D20	0

Verification of ROS induction

Two different assays were used to verify that ROS were induced by the experimental treatment. First, seedling growth was compared to mutant *oxt1* seedlings that show improved tolerance to oxidative stress. Wild-type seedlings grown on ROS inducing media had comparable shoot growth to those grown on ½ MS media. The roots were significantly

shorter, however (Figure 13). The mean root lengths of the WT and hybrid seedlings grown on ROS media were reduced relative to the oxidative stress resistant *oxt1* seedlings at a 95% confidence level (Table 10; Figure 14).

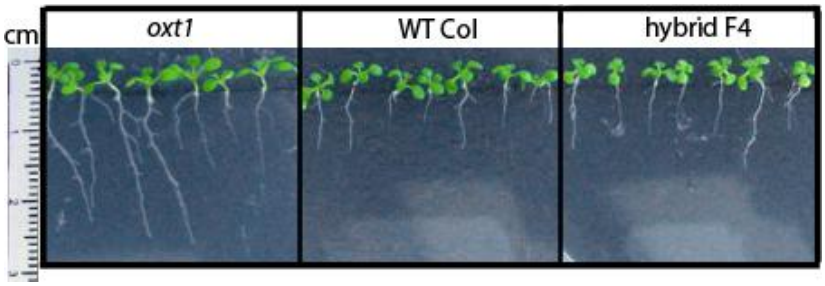


Figure 13: Comparison of seedlings grown on ROS inducing media (40µM BSO + 2µM AT). Seedlings were grown under the same growth conditions for 10 days. WT Col and hybrid F4 seedlings have inhibited growth compared to *oxt1* seedlings.

Table 10: Mean root lengths of 10 day old seedlings grown on ROS inducing media

	Seedling Line					
	Oxt1	**WT Col**	**WT Ler**	**CL-3-12-C6**	**CL-2-12-C5**	**CL-2-12-C15**
Mean root length (mm)	14.09	6.23	7.52	8.03	7.00	8.09
Standard deviation	4.93	1.42	2.44	2.41	1.65	1.24
P Value	---	1.278×10^{-17}	2.41×10^{-7}	3.71×10^{-8}	8.54×10^{-10}	2.63×10^{-7}

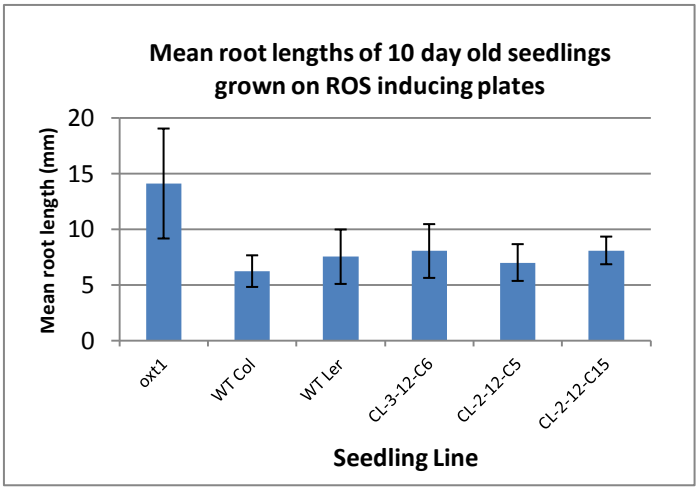


Figure 14: Mean root lengths of 10 day old seedlings grown on ROS plates (½ MS + Suc + 40 µM BSO + 2 µM AT) with standard deviations.

DAB staining was used as a second means of establishing that treatments induced ROS (Figure 15). The staining pattern of seedlings grown on ROS inducing plates is consistent with H_2O_2 accumulation in leaf cells, and more noticeably, in vascular tissue (Figure 15A). Younger leaves tended to stain more than older leaves. Trichomes, where present, were also deeply stained. The whole root was stained darkly and relatively uniformly. The seedlings grown on the ½ MS plates, on the other hand, showed no areas of dark staining (Figure 14B).

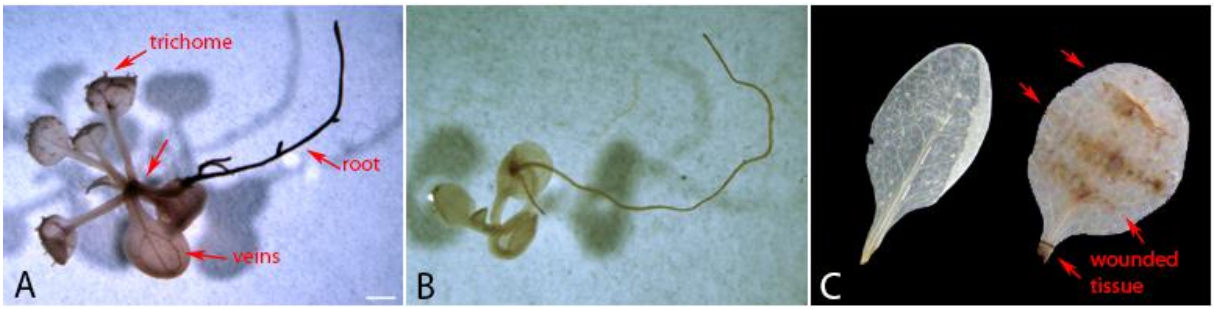


Figure 15: DAB staining of A. thaliana leaves and whole seedlings. Dark brownish-purple stained areas indicate high levels of H_2O_2 (red arrows). A: WT Col seedling grown on ROS inducing media. B: WT Ler seedling grown on ½ MS. C: non-wounded leaf (left) and mechanically damaged leaf (right).

To avoid staining due to tissue damage, seedlings were removed with care from the growth media. The areas where the tweezers inadvertently damaged the seedling or where the roots had been torn, however, became stained. Wounded leaves had small deposits of dark stain around the damaged tissues (Figure 15C). There is also some dark stain at the end of the stem where the leaf had been excised. The untreated leaves also showed slight staining near the incision, with some deposition in the vascular tissue. The rest of the untreated leaf, however, showed no staining reaction above background levels. Background stain appears as light beige in colour.

Assessing marker profiles using quantitative PCR analysis

F4 plants from the LC-2-13 wild-type hybrid lineage were assayed by qPCR using three different indel markers, T6H20, F23M2 and T14G11. Out of the three experimental groups (250 plants in total), a novel insertion sequence was detected in 41 untreated control F4s grown on ½ MS agar and one wounded F4 descendant plant grown on ½ MS agar. These 42 positives shared sufficient homology to an internal primer to direct PCR amplification and generated product in all three technical replicates. An example of qPCR products size separated on an agarose gel is shown in Figure 16.

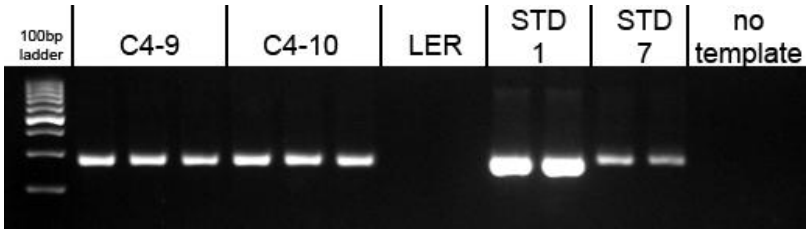


Figure 16 Digital image of an ethidium bromide stained agarose gel showing 151bp qPCR products obtained when using one primer homologous to the insertion sequences for marker T6H20. Samples C4-9 and C4-10 (LC-2-13 lineage) have the correct sized products in all three technical replicates. The Ler DNA controls and the no template controls did not produce PCR products. STD 1 and 7 represent two DNA dilutions used as standards to calibrate the amplification curve.

The C_t value is defined as the cycle in which the amplification curve crosses the threshold. The amplification curve correlates directly with the increase of the fluorescent signal. The threshold is the level of fluorescence above the baseline, where the signal is not considered background. The baseline is the average background noise level calculated using the early cycles when there is no detectable fluorescent signal increase due to the synthesis of double stranded PCR products (Eurogentec, 2008). Samples with higher starting concentrations of template DNA will have amplification curves that cross the threshold at earlier cycles.

For the 41 individuals scoring positive for an insertion in the untreated group, only the marker at the T6H20 locus showed evidence of new insertion sequences. The calculated copy number of insertion sequences relative to external reference sequence varied greatly, ranging from as low as 0.61 copies to as high as 729.6 copies per 100,000 external reference

sequences. Out of the 41 plants, 17 plants had a mean C_t that was 35 or higher. 19 plants had a mean C_t that was between 30 and 35, and only 3 plants had a mean C_t below 30.

One individual in the wounded group designated W1-14 was found to be positive for several novel insertion sequences and tested positive for insertions within three markers. For the markers F23M2, T14G11, and T6H20, there were 49.7, 771.0, and 1402.4 copies per 100,000 external reference sequences, respectively. A summary of copy numbers of insertion and external reference sequences can be found in Appendix C. A graphical representation of the data for those plants is presented in Figure 16. In all cases the mean C_t value was 35 and below.

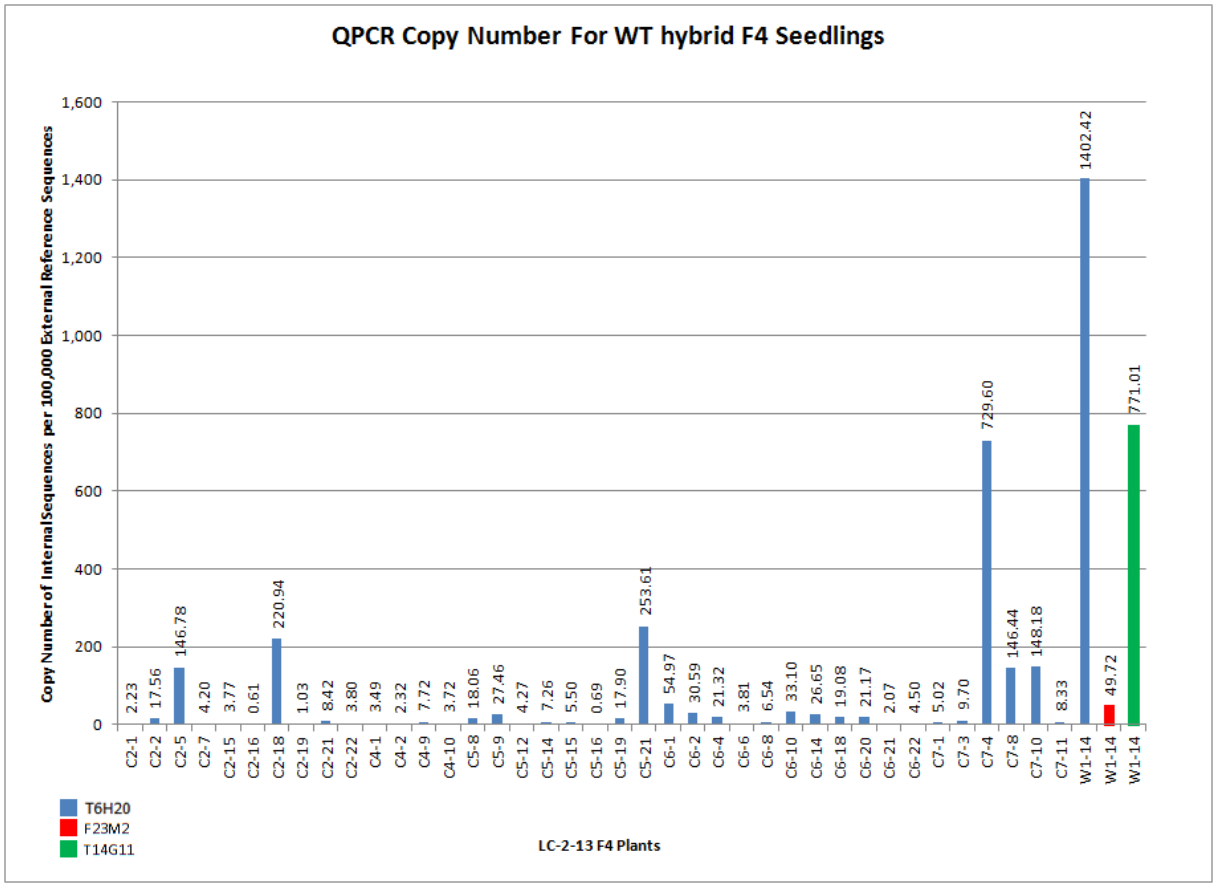


Figure 17: Relative copy number of novel insertion sequences for WT hybrid F4 seedlings (number of internal sequences per 100,000 external reference sequences). Three markers are shown, T6H20 (blue bars), F23M2 (red bars) and T14G11 (green bars).

DNA sequence analyses

Sequence data for W1-14 and the corresponding F3 parent plant (LC-2-13-W1) was obtained for genomic regions approximately 600 base pairs on either side of the indels F23M2, T14G11, and T6H20. A comparative analyses of these sequences revealed that the parental sequence is not identical to W1-14 at any of the loci analyzed (Figure 18). W1-14 sequences at these loci shared homology with the Col-0 reference genome but included 6 nucleotides polymorphisms (arrowheads, Figure 18) plus a 28 base pair insertion downstream of the original F23M2 marker that matched Columbia sequences precisely. At the T14G11 indel locus, W1-14 has acquired a 73 base pair insertion that is identical to the 74 base pair insertion found in Columbia with one exception. The full sequence alignments can be found in Appendix D, E and F.

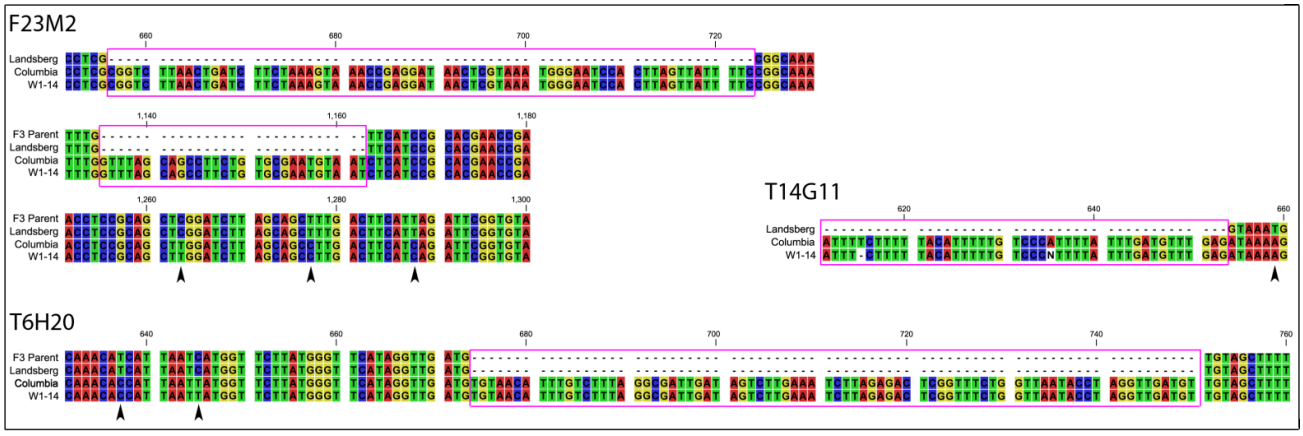


Figure 18: DNA sequence alignments comparing Landsberg, Columbia, LC-2-13-W1-14 and its F3 parent at three loci (F23M2, T6H20 and T14G11). Pink boxes indicate indel sites and arrows indicate sites of single nucleotide polymorphisms. The LC-2-13-W1-14 (shortened to W1-14) DNA sequence is identical to the Columbia sequence.

Sequencing was not completed for the plants LC-2-13-C2-18, LC-2-13-C5-21 and LC-2-13-C7-4 due to technical difficulties. The three samples were genotyped using the indel marker primer for T6H20 as well as the qPCR internal primers (Appendix C, Table 14). No Col sized bands were seen when analyzing the PCR products from the reaction using the indel marker primer set (Figure 19A). Only bands corresponding to the Ler allele were seen. However, Col sized PCR products were made when using the qPCR internal primer set (Figure 19B). The band was faint for LC-2-13-C2-18 and LC-2-13-C5-21 but strong for LC-

2-13-C7-4 and is consistent with the quantitative data obtained by qPCR (Figure 17). The internal sequencing primers, for example T6H20_seqint_R1, are homologous to the insertion found in Col. The PCR reactions using the sequencing primer sets for T6H20 (Appendix C, Table 14) did not produce any bands.

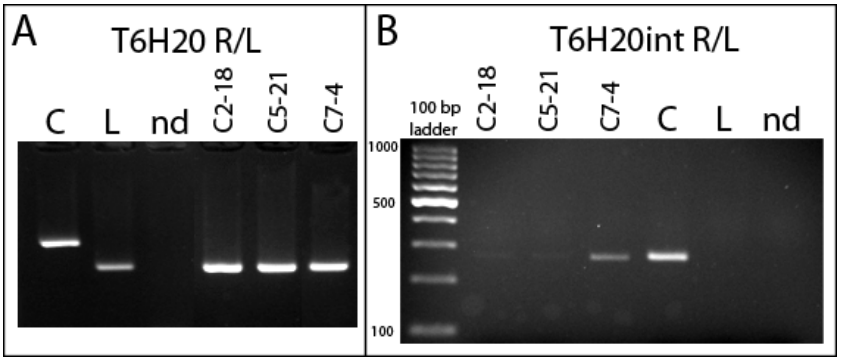


Figure 19: Agarose gel electrophoresis of PCR products for three LC-2-13 F4 DNA samples using indel marker primers (A) and qPCR internal primer (B) sets. Columbia (C), Landsberg (L), no template (nd).

Assessing marker stability in *A. thaliana* in response to tissue culture

The following results were obtained by Chris Hammill as part of his undergraduate thesis project. Callus tissue and regenerated plant tissue were derived from seedlings belonging to LC-2-13 F4 populations and genotyped. Approximately 2/3 of root explants produced callus tissue, an example of which can be found in Figure 20A. The callus tissue was maintained over five transfers to new growth media and about ¾ of the fifth generation callus tissue had shoot regeneration. The types of tissue induced on shoot inducing media varied from whole rosettes with multiple sets of true leaves (Figure 20B) to unidentifiable green pigmented structures (Figure 20D). On shoot-bearing tissue, root development was induced (Figure 20C). In total, whole plant regeneration was achieved in eight tissue cultures. Five of the resulting plantlets survived following transfer to soil and only three of the regenerated plants flowered and set seed.

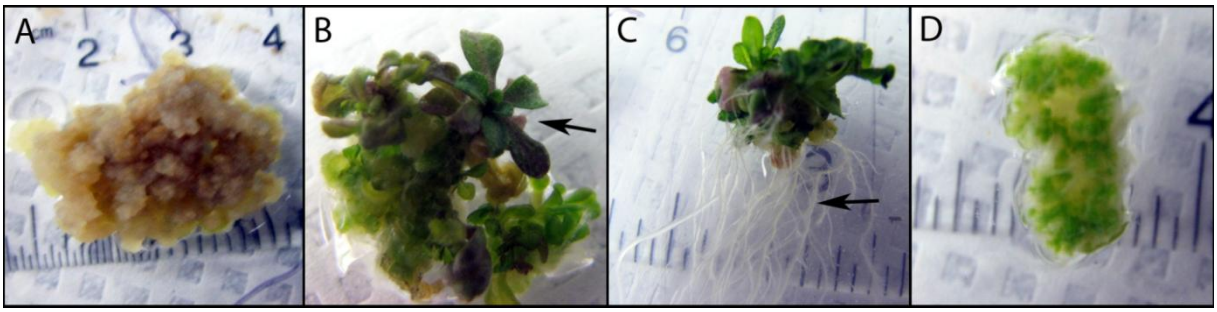


Figure 20: Callus induction using A. thaliana tissue. A: root-derived callus tissue, following one month in culture. B: shoot induction from root-derived callus on SIM media, rosette development is indicated by an arrow. C: root growth, indicated by an arrow, initiated following transfer of callus to hormone-free media. D: green tissue masses derived from root-callus on SIM media.

Molecular genotypes were determined for each plant line prior to callus induction, during progression through tissue culture and following plant regeneration. DNA profiles were determined for the seedlings used to initiate callus formation (to verify starting molecular profiles), for tissue prior to each of five tissue transfers following shoot induction and for seedlings obtained from fully regenerated plants. No differences in indel marker profiles were found in the callus tissue in any of subcultured generations. Also, of the 30 seed progeny tested that were obtained from the regenerated plants, none showed deviation in their indel marker profiles (Table 11).

Table 11: Summary table comparing the genotypes of each tissue sample profiled from the three experimental treatments. Genotypes for each marker were either homozygous Landsberg (Ler) or there was no amplification (No amp; genotype unknown).

		Genotype at Marker									
		T6H20		T14G11		T11I18		MSA6		F23M2	
Experiment	**Total samples**	**Ler**	**No amp.**	**Ler**	**No amp.**	**Ler**	**No amp.**	**Ler**	**No amp.**	**Ler**	**No amp.**
Sub-culturing	59	58	1	49	10	55	4	57	2	57	2
Shoot induction	20	20	0	20	0	19	1	20	0	19	1
Regenerated plant seedlings	30	30	0	30	0	30	0	29	1	30	0

Discussion

Traditionally, intraorganismal genetic heterogeneity (IGH) is thought to occur infrequently with genetic homogeneity being the norm. There is, however, increasing evidence that IGH is more widespread than previously thought. Taking the form of mosaicism or chimerism, IGH has been documented in bacteria, protists, fungi, vertebrates, and plants (Pineda-Krch and Lehtila, 2004). Also, modular organisms, such as plants, have a tendency towards IGH. We see this in long-lived trees as well as in short-lived plants like *Arabidopsis.* In *Arabidopsis* we see a form of IGH arising from a mechanism termed *restoration*, where individuals had reacquired previously lost ancestral genetic sequences. The mechanism behind this newly discovered inducer of IGH is not currently known. We also do not know how widespread restoration is in the plant kingdom. The potential value of this mechanism cannot be fully appreciated until we further our knowledge of it.

Finding triggers that would up-regulate the frequency of restoration events might shed some light on the mechanism. In order to accomplish that goal the effect of stresses were examined in *Arabidopsis.* It is hypothesized that the following stresses are triggers that elevate restoration frequency: metabolic stress, temperature stress, mechanical wounding and elevated ROS exposure. The aforementioned stresses were applied to two transgenic ADK silencing lines and a wild-type hybrid line. The genomic-targeted effects, in the form of restoration events, were measured using molecular genotyping techniques.

These methods yielded an outcome contrary to that predicted by the initial hypothesis and show that stress is perhaps an inhibitor of restoration. Non-stressed plants appear to have a higher frequency of restoration in comparison to stressed plants. This was especially apparent when qPCR, a sensitive and accurate technique, was used for genotyping plants in the mechanical wounded experiments. qPCR revealed that there was a high frequency of restoration in the progeny of un-treated control plants, but this was not the case for the progeny of the wounded plants. We can speculate about what these results might mean in the evolutionary sense, however, its true significance is still unclear.

Restoration as a Mechanism for Mosaicism

The first experiments explored restoration frequency on temperature stressed ADK-deficient lines. The cold stress experiments were done on seedlings from the line amiL-5-1. This line is derived from a single self-fertilized plant (derived from an initial amiADK x Ler cross; see). In this experiment, cold stress did not increase restoration frequency. In fact, no revertant seedlings were found in the stressed groups. Two revertant seedlings were found, however these were in the control group.

The two revertants, numbered 7 and 55, had different genetic profiles than their parent (Table 3). Seedling #7 differed from its parent at three indel markers. Two markers scored as homozygous for the insertion and one marker scored as heterozygous (Table 3). Since the parent was homozygous for the deletion for those three markers, #7's profile suggests that it could have double reversions and single reversions of alleles. Seedling #55 was also remarkable. Plant #55 was found to be a mosaic as its root and shoot had differing genetic profiles (Figure 9, Table 3, marker MSA6). Both alleles in the root carried the insertion, while the shoot was heterozygous. As its parent was also heterozygous at that marker, it is not clear if there was a reversion to the insertion allele in the root or a reversion to a deletion allele in the shoot. It is also possible that there was a double reversion in one tissue section and only a single reversion in the other.

Identification of revertants in seedlings grown at control temperatures but not at the other two temperatures could be due to either a real biological phenomenon or could be attributed to technical errors. The quality of the seedling DNA preparations was variable, resulting in poor or no PCR amplification for some samples. Although over 100 seedlings in each temperature group were genotyped, not all PCR amplifications generated usable data. Table 2 provides a summary of the number of usable data points. In addition, certain PCR primer sets that did not drive good amplification, notably, T11I18. This technical problem led to an incomplete genetic profiling of the amiL-5-1 F3 population. As such, it is possible that reversion events were missed.

Plants from the other ADK-deficient line where the original stocks harboured an ADK::GFP construct were examined for evidence of restoration of the transgene. The plants were the progeny of non-transgenic parents derived from segregation in the hybrid ADK-GFP lines (Hadk). Thus, fluorescent sectors in these plants can be used as indicators of restoration.

Molecular genotypes for these same lines were determined concurrently. The genetic profiles of two seedlings did not match the profiles of their respective parents based on Mendelian segregation. These will be designated simply as the fluorescent seedling and seedling #12 (Table 6). In both cases, the seedlings had heterozygous alleles for markers that were homozygous in the parental generation. Restoration could have caused the allele change to heterozygosity in both seedlings as well as the reappearance of the transgene in the fluorescent seedling. The sequence information would have been cached for at least one generation, the parental generation, as the restored alleles were indeed present in the grandparental generation. If the plant had restored the transgene, this result would show that the restoration mechanism can cache new genetic information and re-integrate it into the genome with minimal or no mutations.

The results can be explained in other ways, however. It is possible that the DNA preparations were contaminated with Col or Ler DNA. The contaminating DNA would therefore be the source of template that produced the aberrant PCR products seen in the agarose gel. It is also possible that the revertants are actually the result of cross-pollination with foreign pollen. Although plants were grown in a manner that minimized cross-pollination, the plants were not totally isolated from one another in the growth chamber. Also, seed pools may have been cross-contaminated with stray seeds from other plants. For the majority of aforementioned cases the alternate explanations cannot totally be ruled out. However, there is a robust example of IGH found in seedling #55 (amiL-5-1 line). Here, the root and shoot sections differ in genetic profiles, thus ruling out pollen cross-contamination and seed cross-contamination. It is the most compelling example of IGH presented so far.

Wild-type hybrid lines were also examined for restoration. Each generation of plants were profiled using molecular genotyping. Adult plants (F2 and F3 generations) as well as seedlings (F4 generation) were genotyped. F3 and F4 plants belonged to the wounding, no treatment, or ROS groups. No sequence aberrations were found in any samples in any of the groups, as summarized in Tables 7, 8 and 9.

The level of sensitivity of the PCR assay could have resulted in a reduced ability to detect restoration events. The PCR genotyping method used in the majority of experiments relies on sequences that flank the indel site. These regions are homologous in both Col and Ler genomes. Insertion sequences may not be amplified to a degree that can be visualized by

agarose gel electrophoresis if the sequences are at a very low abundance level. Also, flanking Col sequences may also be competing less effectively with Ler sequences for primers, thereby further reducing the amplification of the Col sequence.

Experimental evidence comparing PCR products obtained using flanking indel primers as opposed to primers where one primer hybridizes within the insertion sequence suggest that the actual number of insertion events may have been underestimated. As shown in Figure 18A, no insertion is detected at the locus when the flanking indel marker primer set was used. However, a product corresponding in size to the Col product is amplified when one of the two primers is homologous to a region within the insertion sequence (Figure 18B). This finding suggests restoration events may have been missed because of the sensitivity of the size-based PCR assay used.

Quantitative PCR (qPCR) is a sensitive, real-time method that can be used to accurately quantify DNA. QPCR was used to identify possible revertant hybrid plants in the control, wounded and ROS experimental groups. In addition to having increased sensitivity, this method allowed quantification of sector size in an individual sample by comparing the number of insertion sequences to the copy number of a genomic region immediately flanking the indel locus serving as a reference sequence. Plants in the wounded and ROS experimental groups were assayed using qPCR and compared to control (non-treated) plants at three marker loci; T6H20, F23M2 and T14G11. No marker changes were detected following ROS treatments. However, in the wounded group, one plant, designated LC-2-13-W1-14, scored positive for insertions at all three markers. Interestingly, in the control group 41 plants scored positive for insertions at the T6H20 marker (Figure 16, Appendix B).

The calculated copy number of insertion sequences for LC-2-13-W1-14 varied from marker to marker. In addition to showing robust amplification, a PCR product of the correct size was produced making it unlikely that the products are the result of non-specific binding. Of the three markers, the T6H20 locus showed the largest number of insertion copies. It was found that there were 1402 copies of insertion sequences per 100 000 external reference sequences, or in other words about 14 in 1000 DNA strands, carried the insertion.

If all insertion events for LC-2-13-W1-14 occurred concurrently in the same target cells, the copy numbers should be identical for all three markers. However, copy numbers differed between markers. This suggests that the restoration events did not arise at the same

frequency for each locus. One possible explanation for this disparity is that DNA sequence changes occur independently at each locus and may not be conditioned by events at other loci. It may be that each genetically distinct sector detected using qPCR resides in different parts of the plant. Because DNA was isolated from samples consisting of relatively large amounts of uniformly homogenized tissue we cannot detect sectors separately. Based on the data presented here, TH620 appears to be a hotspot for genomic sequence changes.

DNA sequence analysis of LC-2-13-W1-14 for the insertion site and regions flanking it revealed homology to wild-type Col sequences and not the F3 parent sequence. Also, upon further investigation using PCR molecular genotyping, it was found that LC-2-13-W1-14 scored positive for the insertion sequence for markers MSA6, F16J13, F2P16, and MGI19, all markers that were homozygous Ler in the F3 parent (Figure 18). Finding insertion sequences for all markers expected to be homozygous for the deletion could mean that the sample was contaminated with wild-type Col DNA. Also, polymorphisms found in LC-2-13-W1-14 such as single nucleotide polymorphisms and indels were also identical to wild-type Col, not the parental DNA. It is however possible that that the restored sequence tract covered the entire area sequenced given that the extent of a restored sequence tract is not yet determined.

The results for LC-2-13-W1-14 may suggest that the restoration events occurred on a global genomic scale over multiple unlinked loci. This could mean that there were targeted sequence reversions that occurred at the examined loci, or there could have been larger sections of the genome acquired sequence reversions. These reversions could have been limited to one sector in the plant, or there could be several small sectors dispersed throughout the plant with each having reversions occurring at different loci. As each plant was examined as a mixture of total DNA, the size of this sector, or where it arose on the plant could not be determined.

Due to the nature of the qPCR assay, it is not clear if what reversions occurred within a given chimeric sector. Also, the reversions may have arisen at any point in time of the plant's development. A sector arising early in development, for example, will be larger than those developing at a later stage as the initial cell or cells that experienced restoration would have had more cell divisions before the leaf matured. The sector of cells that carry the trait will be larger. It is likely that the restoration events occurred fairly late in development based on the low copy number of sequences that carry the insertion. IGH was detected at the

seedling stage as well as older plants. So, it is likely that genome fluctuations are ongoing and dynamic and not limited to specific developmental stages.

Although the copy number of sequences carrying the insertion was normalized relative to the reference, there were still differences in apparent restoration frequency between loci for LC-2-13-W1-14. One trivial explanation is that the differences were caused by uneven degradation of DNA. DNA can be bound by histones as nucleosomes. In theory, DNA associated with nucleosomes are less prone to degradation because they are protected from nucleases by bound histones (Thanakiatkrai and Welch, 2010). The differences in histone configuration across the genome could account for the differences in degraded DNA. For example, the calculated copy numbers of external reference sequences differ significantly between markers, with the amount detected for the T6H20 locus being almost double that for the F23M2 locus (Appendix B, W1-14 data). This discrepancy may indicate unequal DNA degradation, differences in qPCR amplification efficiency between the different primer sets, differences in input amount or a combination of factors.

The T6H20 indel has proven to be an interesting marker for testing restoration. There appears to be a fair amount of activity at that locus as indicated by the 41 positive plants in the qPCR assay for the no-treatment group alone. The copy number of restored sequences was very low and as such, these restoration events would not have been captured by the conventional PCR molecular genotyping method only utilizing flanking primers. The products can be reproduced using the insertion anchored primer set. The result, however, is a faintly visible product in an agarose gel (Figure 19B). Given that the other positive samples (excluding W1-14) had insertion copy numbers lower than the three samples in Figure 19, it is likely that many positives would still be missed.

Mosaicism at T6H20 has also been observed in an *hth-7* plant (Figure 1) at T6H20. Branches from this plant were assayed separately by qPCR using four indel markers, F8D6, F15H11, T14G11, and T6H20. Multiple marker changes were detected in each branch, and the copy number of insertions varied from marker to marker, reminiscent of the qPCR results for LC-2-13-W1-14. The T6H20 locus was found to be particularly active, with the highest copy number of insertion sequence per reference sequence for the *hth-7* plant. The copy number per reference sequence for the *hth-7* plant was smaller than that found in the control plants, up to 0.8/1000 and up to 7.3/1000 respectively (Hopkins et al., 2011). The discrepancy

could be due to the differences in sampling methods of the two experiments. DNA had been collected from only a portion of the *hth-7* branch, whereas the whole F4 wild-type hybrid plant was homogenized for DNA extraction. It is possible that only part of the restored sector of the *hth-7* plant was used for the qPCR. Alternatively there could be a naturally wide range of restoration frequency within and between populations and genetic backgrounds.

The *hth-7* qPCR experiments also revealed restoration at the locus T14G11. However, positive restoration events were not found at that locus in the qPCR using the F4 wild-type hybrid plants (the only exception being LC-2-13-W1-14). The F23M2 site also did not have much activity. The T6H20 locus had a very high frequency of restoration events and could potentially be a hotspot for restoration, although the reasons are still unclear.

The T6H20 indel is located on chromosome 3 in the first intron of Type 1 serine/threonine protein phosphatase 5 (*TOPP5*, AT3G46820.1). Protein phosphatases are involved in regulation of various processes in Arabidopsis such as abscisic acid signalling, auxin transport, and receptor-like protein-signalling (Lin et al., 1997). An integral part of cell-signalling events, protein phosphatases coordinate for protein phosphorylation and dephosphorylation with protein kinases (Wang et al., 2007). The *TOPP5* gene encodes the catalytic subunit of one of eight identified Type 1 protein phosphatases (PP1) in *Arabidopsis* (Kerk et al., 2002). Knowledge of PP1 genes is limited at this point in time, however, it is known that PP1 catalytic subunits are ubiquitously expressed based on expression pattern analysis. The PP1 genes have been located to 4 of the 5 Arabidopsis chromosomes. Phylogenetic analysis of the predicted amino acid sequences of the *A. thaliana* PP1 cDNA clones reveal that they are highly conserved. They are also similar to the amino acid identities of PP1 proteins found in other plant species, as well as fungi and animals. *TOPP2* is the closest related PP1, having a 92.0% identity when comparing cDNA clones. So, it is possible that some percentage of the insertions may have arisen from gene conversion with other genes in the genome (Lin et al., 1998). However, the indel sequence in question did not have 100% homology with other PP1 genes.

The high conservation of the PP1 catalytic subunit primary structures across different phyla is likely to have been maintained under strong selective pressure through evolution (Lin et al., 1998). It seems counterintuitive to have genetic variation in the form of sequence reversions at one of the PP1 genes as the PP1 amino acid primary structures are so essential

for their function. The T6H20 indel, however, is located within an intron, not the coding sequence of *TOPP5*. Introns in plant genes have been known to contain regulatory elements such as enhancer and silencer elements (Reddy and Reddy, 2004; Kim et al., 2006). Also, many important regulatory elements have been identified in large introns (over 500 bp), such as in the MADS-domain transcription factor genes *FLOWERING LOCUS C* and *SEED-STICK* (Sheldon et al., 2002; Kooiker et al., 2005). It is possible that the large first intron of *TOPP5*, measuring a little more than 500 bp, contains an as of yet unidentified regulatory region. It may be that the variation of expression of *TOPP5* provides an evolutionary advantage rather than the variation of the amino acid sequence. The role of the *TOPP5* gene in restoration would be an interesting avenue for further investigation.

Figure 21: TOPP5 gene model (AT3G46820.1). The qPCR T6H20 indel amplicon is located within intron 1 of the TOPP5 gene. Dark shaded areas indicate protein-coding regions, light shaded areas indicate untranslated regions.

The Effect of Stress on Restoration Frequency

Temperature stress was imposed on an ADK silencing line in an attempt to increase restoration frequency. The low growth temperature was used as a source of environmental stress. Although it was hypothesized that low growth temperatures would stress the plants and trigger restoration, there were no cases of restoration in plants grown at the lower temperatures. The only cases of restoration were found in the control group (21^{o}C growth temperature). Two of 105 seedlings in the 21^{o}C group were found to have either a different genotype from the parent or found to have differing genotypes in their roots and shoots. With these results, it appears that lower temperatures may inhibit restoration in seedlings.

The wild-type hybrid lines were examined more closely using qPCR. The assay revealed that the plants from the control, un-treated lines had low but detectable levels of non-parental sequence insertions while the plants in the wounding and ROS groups did not. Forty-one individuals were scored as positive for an aberrant insertion in the control lines, whereas

in the stressed lines there was only one (found within the wounding group). We can argue that the genetic background of the plants may be the cause of the differences in genetic stability found between the wounding and un-treated control lines. However, the plants in the ROS group were derived from the same population as the un-treated control plants. They have the same genetic background which supports the conclusion that ROS exposure not genetic background was the factor affecting IGH. These results are similar to that found in the temperature stress studies. It appears to contradict the hypothesis that stress elevates restoration frequency.

Plant responses to the external environment are complex. As such, it can be difficult to elucidate the factors governing a particular mechanism. There are changes at transcriptional, protein, metabolite, and epigenetic levels. Much is still unknown about the individual contributing factors in various systems. Even in this post-genomic era, there are a multitude of genes and proteins with unknown functions. Also, many, if not most, genes respond to endogenous and exogenous stimuli. If the plastic and modular natures of plants are taken into account as well, it can be even more difficult to understand the components and processes that contribute to a particular response. Epigenetics also plays a large role in the expression of genes.

It was observed that there was generational variation in phenotypes in the ADK-silencing pure lines. It is possible that this variation is controlled epigenetically. RNA silencing is thought to have developed as a defense against infection with viruses (Baulcombe, 2002). ADK has been found to be necessary for viral defence in tobacco (Wang et al., 2005), and perhaps the variation of phenotype and expression of ADK is affected by the levels of ADK itself. Gene silencing has been shown to be relieved by stress in *Arabidopsis*. This effect was transmissible over a few generations and was able to be reset upon certain conditions (Lang-Mladek et al., 2010). Perhaps a similar mechanism gave rise to phenotypically wild-type plants that were able to produce progeny with severely silenced phenotypes.

In these experiments, the ADK-silenced lines were crossed with wild-type Ler giving rise to large, seemingly healthy F1 progeny with no visible ADK silencing. The loss of ADK silencing in the F1 generation could be due to a reduced expression of the transgene given that the F1 plants have only one copy of the transgene in the genome. On the other hand, the loss

of silencing appears similar to heterosis. Heterosis or hybrid vigor has been documented with *Arabidopsis* ecotype hybrids. Factors that may contribute to hybrid vigor include epigenetic changes in levels of 24nt siRNA, DNA methylation and expression levels (Groszmann et al., 2011).

A large part of this project investigated the effect of mechanical wounding and ROS on IGH. In response to mechanical wounding a multitude of genes involved in defense and repair are up-regulated (Reymond et al., 2000). There is also large overlapping of gene expression that responds to a wide range of environmental stresses (Walley et al., 2007). Perhaps this increase of resource utilization and allocation for use in stress response indirectly influences restoration by limiting the resources available for the restoration mechanism. This theory fits with the observations found in the qPCR assays. A remarkable number of plants in the control, un-treated group were found to score positive for the acquisition of insertion sequences whereas the plants in the stressed groups had little to no evidence of genomic sequence changes.

A similar observation was found in soybean (K. Espinosa and R. Palmer, personal communication). Soybean plants were grown in a low-density honeycomb configuration. Half the plants were subjected to simulated hail, which consisted of defoliation of two thirds of the leaves. The other half served as the non-hailed controls. In one lineage, several of the plants showed within plant variation at different indel markers that also deviated from parental plant molecular genotypes (R. Palmer and K. Espinosa, personal communication). Variation was found only in the control group, a finding that is consistent with the results from the *Arabidopsis* wounding experiments discussed here.

Growing crops in ultra-low density, a method employed in the soybean experiments, has been shown to reveal genetic variability in highly inbred lines (Fasoula and Boerma, 2007). Growing individual plants at ultra-low densities essentially eliminates competition. Defined as the plant-to-plant interference with the equal use of density-limited aboveground and underground growth resources; competition results in unequal use of resources due to competitive advantages of some plants over others. The various competitive advantages and disadvantages can be induced genetically or environmentally. Competition results in differences in growth and development of plants due to uneven growth suppression (Fasoula and Fasoula, 2002). Removing the need for resource allocation for competition may unburden

plants enough to unmask variation. In keeping with this line of thought, we would expect to detect DNA sequence changes in those plants that do not have to cope with stress, such as the un-treated *Arabidopsis* and soybean plants.

Removing competition may have the effect of unveiling variation, but there are other unexplored factors such as synergism that may also play a role in this process. For expression of some genes, regulation of transcription requires the synergistic binding of transcription factors (Michel, 2010). Synergistic phenotypes arise from the activity of two different genes and do not resemble either (Martienssen and Irish, 1999). It is not unthinkable that restoration could rely on a similar regulation mode where two or more factors cooperatively control the restoration mechanism. There may be genetic and environmental factors that could, when combined, significantly increase restoration frequency. For example, an experiment could be designed to examine an *hth* mutant line grown in a low-density honeycomb pattern. These factors have been shown to be associated with increased genetic variability and together the effect may be amplified.

Similarly, tissue culture propagation has been associated with higher rates of mutation in regenerated plants (Jiang et al., 2011). It was our expectation that restoration frequency would likewise be increased. Our genotyping callus tissue and regenerated plants did not reveal any DNA sequence reversions. However, the number of samples, loci and plant lines examined was relatively small. Furthermore, sensitive quantitative methods were not employed. Perhaps detecting sequence reversion requires a more in-depth investigation. It is also possible that there were reversions in a small population of cells, but the molecular genotyping method used to detect reversions was not sensitive enough. A highly comprehensive method such as the massively parallel Illumina sequencing utilized by Jiang et al. (2011) would allow for whole genome coverage, thereby drastically increasing the likelihood of reversion detection. Such an approach would be ideal for the detection of restoration in general.

As a final note, it is an intriguing fact that the variation in the control lines in the mechanical wounding experiment was only found at the T6H20 loci. As previously mentioned, the T6H20 indel is within the first intron of the *TOPP5* gene. As for the other indels, F23M2 is not found within a gene and T14G11 is located within an unknown protein. The significance of *TOPP5* in restoration has not yet been determined. There are however,

many directions for speculation about *TOPP5* as well as the role of the various stresses on the restoration mechanism. *TOPP5* is one of the genes that are up-regulated in response to genotoxic stress (Chen et al., 2003). PP1 proteins, such as *TOPP5*, are involved in the regulation of membrane channels, cell cycle control and developmental regulation in plants (Smith and Walker 1996; Luan, 2003) and it is likely that they are up-regulated in response to stress. In fact, many genes are up-regulated in response to stress. Perhaps by examining these response genes in future experiments, we can find loci with high restoration activity.

Conclusions

There were a few cases of restoration or possible restoration in the ADK-silencing lines. The possibility that the revertants found in the Hadk and amiADK lines are actually the result of out-crossing or seed contamination cannot be definitively ruled out. However, mosaicism within individuals is the strongest evidence for restoration. Mosaicism was observed in the amiADK F3 seedling amiL-5-1 #55 in which the root and shoot had different genotypes. Most importantly, mosaicism in an individual excludes out-crossing and contamination of the seed pool by errant seeds. Even more striking evidence of restoration was found when qPCR was employed to assay DNA sequence reversions in wild-type F4 plants in the wounding and ROS experiments. QPCR allowed for the detection of minute copies of DNA strands containing non-parental insertion sequences among the vastly more numerous non-reverted alleles. The reversions in the wild-type hybrid F4 seedlings were not able to be detected using conventional size-based PCR genotyping. So, it logically follows that many restoration events may have been missed due to use of insufficiently sensitive genotyping techniques. As for whether environmental or metabolic stress serves to increase restoration frequency, the evidence provided here points tentatively to the opposite hypothesis. The cases of restoration in the temperature, wounding and ROS experiments were found in the control groups. In those experiments, cold-stress and wounding seemed to turn off the restoration mechanism as no revertants were found in any of the stressed groups. On another note, there is insufficient evidence to conclude that ADK-deficiency increases restoration frequency. However, increasing sample size in all experiments would facilitate in obtaining a robust statistical evaluation.

Investigating restoration can be an inherently difficult process. It can be difficult to use the acquisition of non-parental DNA as an indicator of restoration. There were several cases where progeny were found to harbour unexpected alleles. But, the evidence needed to be examined with possible sources of error in mind. Contamination is an issue that cannot be avoided as it can confound the results of each experiment. Sources of contamination include DNA from parental plant, wild-type, or other sources. Out-crossing and seed contamination could explain why some of the progeny had the rare single and double reversions. With this in mind, it may be challenging to accept the outcomes presented in this thesis as examples of true restoration events. However, the accumulation of evidence for restoration working as a mechanism for generating IGH is compelling. It is my belief that with the use of more sensitive and robust technologies and methodologies, the mystery that is the restoration mechanism will slowly but surely be revealed. In this project there were several cases of genetic variances that could be attributed to restoration. However, in some cases, the suspect results could be explained by more mundane means. Even so, this research has ultimately revealed that intraorganismal genetic heterogeneity occurs in Arabidopsis plants, primarily when plants experience good growth conditions and are not resource limited.

References

Abbott RJ and Gomes MF. 1989. Population genetic structure and outcrossing rate of *Arabidopsis thaliana* (L.) Heynh. Heredity 62: 411-418.

Alcazar RM, Lin R and Fire AZ. 2008. Tranmission dynamics of heritable silencing induced by double-stranded RNA in *Caenorhabditis elegans*. Genetics 180: 1275-1288.

Apel K and Hirt H. 2004. Reactive oxygen species: metabolism, oxidative stress, and signal transduction. Annu Rev Plant Biol 55: 373–399.

Bakker EG, Stahl EA, Toomajian C, Nordborg M, Kreitman M, and Bergelson J. 2006. Distribution of genetic variation within and among local populations of *Arabidopsis thaliana* over its species range. Mol Ecol 15: 1405-1418.

Baulcombe D. 2002. Viral suppression of systemic silencing. Trends in Microbiology 10: 306-308.

Bergelson J, Stahl E, Dudek S, Kreitman M. 1998. Genetic variation within and among populations of *Arabidopsis thaliana*. Genetics 148: 1311-1323.

Browse J. 2005. Jasmonate: an oxylipin signal with many roles in plants. Vitam. Horm. 72: 431-456.

Burch L. R. and Stuchbury T. 1987. Activity and distribution of enzymes that interconvert purine bases, ribosides and ribotides int he tomato plant and possible implications for cytokinin metabolisms. Physiol Plant 69: 283-288.

Chaudhury A. 2005. Hothead healer and extragenomic information. Nature. 473: E1-E2.

Chen I, Haehnel U, Altschmied L, Schubert I and Puchta H. 2003. The transcriptional response of *Arabidopsis* to genotoxic stress - a high-density colony array study (HDCA). The Plant Journal 35: 771-786.

Chen J, Cooper DN, Chuzhanova N, Ferec C, and Patrinos GP. 2007. Gene conversion: mechanisms, evolution and human disease. Nature Reviews Genetics. 8: 762-775.

Chen Y, Schneeberger RG, Cullis CA. 2005. A site-specific insertion sequence in flax genotrophs induced by the environment. New Phytologist. 167: 171-180.

Comai L. 2005. A toxic mutator and selection alternative to the non-Mendelian RNA cache hypothesis for *hothead* reversion. The Plant Cell. 17: 2856-2858.

Cullis CA. 2005. Mechanisms and control of rapid genomic changes in flax. Annals Bot 95: 201-208.

Cullis CA and Chralton L. 1981. The induction of ribosomal DNA changes in flax. Plant Sci Lett 20: 213-217.

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 19:1349.

Eurogentec. 2008. QPCR guide. http://www.eurogentec.com/uploads/QPCR-guide.pdf

Fagerstrom T, Briscoe DA and Sunnucks P. 1998. Evolution of mitotic cell-lineages in multicellular organisms. Trends Ecol Evol 13: 117-120.

Fasoula VA and Boerma HR. 2007. Intra-cultivar variation for seed weight and other agronomic traits within three elite soybean cultivars. Crop Science 47: 367-373.

Fasoula VA and Fasoula DA. 2002. Principles underlying genetic improvement for high and stable crop yield potential. Field Crops Research 75: 191-209.

Folse HJ III and Roughgarden J. 2011. Direct benefits of genetic mosaicism and intraorganismal selection: modeling coevolution between a long-lived tree and a short-lived herbivore. Evolution doi:10.1111/j.1558-5646.2011.01500.x

Fujita M, Fujita Y, Noutoshi Y, Takahashi F, Narusaka Y, Yamaguchi-Shinozaki K and Shinozaki K. 2006. Crosstalk between abiotic and biotic stress responses: a current view from the points of convergence in the stress signaling networks. Curr Opin Plant Biol 9: 436-442.

Gachev T, Gadjev I, Van Breusegem FV, Inze D, Dukiandjiev S, Toneva V and Minkov I. 2002. Hydrogen peroxide protects tobacco from oxidative stress by inducing a set of antioxidant enzymes. CMLS, Cell. Mol. Life Sci. 59: 708-714.

Glauser G, Dubugnon L, Mousavi SAR, Rudaz S, Wolfender J and Farmer EE. 2009. Velocity estimates for signal propagation leading to systemic jasmonic acid accumulation in wounded *Arabidopsis*. The Journal of Biological Chemistry. 284: 34506-34513.

Griffith OW and Meister A. 1979. Potent and specific inhibition of glutathione synthesis by buthionine sulfoximine (S-*n*-butylhomocysteine sulfoximine). J Biol Chem. 254: 7558-7560.

Groszmann M, Greaves IK, Albertyn ZL, Scofield GN, Peacock WJ and Dennis ES. 2011. Changes in 24-nt siRNA levels in Arabidopsis hybrids suggest an epigenetic contribution to hybrid vigor. PNAS 108: 2617-2622.

Gundlach H, Muller MJ, Kutchan TM and Zen MH. 1992. Jasmonic acid is a signal transducer in elicitor-induced plant cell cultures. Proc. Natl Acad. Sci. 89: 2389-2393.

Hopkins MT, Author AMK, Chang P-C, Author KCV, Lai D, Doerr MD & Lolle SJ. 2011. *De Novo* Genetic Variation Revealed in Somatic Sectors of Single *Arabidopsis* Plants. [Internet]. In *Plant-a-Seed*. Available at http://www.plant-a-seed.com/#research-paper

Humpherys D, Eggan K, Akutsu H, Hochedlinger K, Rideout WM 3rd, Biniszkiewicz D, Yanagimachi R and Jaenisch R. 2001. Epigenetic instability in ES cells and cloned mice. Science 293: 95-97.

Jiang C, Mithani A, Gan X, Belfield EJ, Klingler JP, Zhu JK, Ragoussis J, Mott R and Harberd NP. 2011. Regenerant *Arabidopsis* lineages display a distinct genome-wide spectrum of mutations conferring variant phenotypes. Current Biology 21: 1385-1390.

Koo AJK, Gao X, Jones AD and Howe GA. 2009. A rapid wound signal activates the systemic synthesis of bioactive jasmonates in Arabidopsis. The Plant Journal. 59: 974-986.

Kooiker M, Airoldi CA, Losa A, Manzotti PS, Finzi L, Kater MM and Colombo L. 2005. *BASIC PENTACYSTEINE1*, a GA binding protein that induces conformational changes in the regulatory region of the homeotic Arabidopsis gene *SEEDSTICK*. Plant Cell 17: 722-729.

Koornneef M, Hanhart C and Thiel F. 1989. A genetic and phenotypic description of *Eceriferum (cer)* mutants in *Arabidopsis thaliana* . Journal of Heredity 80: 118-122.

Lang-Mladek C, Popova O, Kiok K, Berlinger M, Rakic B, Aufsatz W, Jonak C, Hauser M and Luschnig C. 2010. Transgenerational inheritance an resetting of stress-induced loss of epigenetic gene silencing in *Arabidopsis*. Molecular Plant 3: 594-602.

Lin Q, Li J, Smith RD and Walker JC. 1998. Molecular cloning and chromosomal mapping of type one serine/threonine protein phosphatases in *Arabidopsis thaliana*. Plant Molecular Biology 37: 471-481.

Lolle SJ, Victor JL, Young JM, and Pruitt RE. 2005. Genome-wide non-mendelian inheritance of extragenomic information in *Arabidopsis*. Letters to nature. Nature. 434: 505-509.

Lolle SJ, Victor JL, Young JM and Pruitt RE. 2006. Increased outcrossing in hothead mutants. Nature 443: E8.

Luan S. 2003. Protein phosphatases in plants. Annu Rev Plant Biol 54: 63-92.

Martienssen R and Irish V. 1999. Copying out our ABCs the role of gene redundancy in interpreting genetic hierarchies. TIG 15: 435-437.

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: 2295-2297.

Michel D. 2010. How transcription factors can adjust the gene expression floodgates. Progress in Biophysics and Molecular Biology 101: 16-37.

Moffatt BA, Wang L, Allen MS, Stevens YY, Qin W, Snider J, and von Schwartzenberg K. 2000. Adenosine kinase of *Arabidopsis*. Kinetic properties and gene expression. Plant Physiol. 124(4): 1775-1785.

Murashige T and Skoog F. 1962. A revised medium for rapid growth and bioassays for tobacco tissue cultures. Physiol Plant 15(3): 473-497.

Nowacki M, Vijayan V, Zhou Y, Schotanus K, Doak TG and Landweber LF. 2008. RNA-mediated epigenetic programming of a genome-rearrangement pathway. Nature 451: 153-159.

Nowacki M, Haye JE, Fang W, Vijayan V and Landweber LF. 2010. RNA-mediated epigenetic regulation of DNA copy number. PNAS 107: 22140-22144.

Pineda-Krch M and Lehtila K. 2004. Costs and benefits of genetic heterogeneity within organisms. J Evol Biol 17: 1167-1177.

Probst AV, Dunleavy E, Almouzni G. 2009. Epigenetic inheritance during the cell cycle. Nature Reviews Molecular Cell Biology. 10: 192-206.

Rassoulzadegan M, Grandjean V, Gounon P, Vincent S, Gillot I and Cuzin F. 2006. RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. Nature 441(25): 469-474.

Raven, Evert, Eichhorn. 1999. Biology of Plants sixth edition. W.H. Freeman and Company/Worth Publishers. pp. 192-193.

Ray A. 2005. RNA cache or genome trash? Nature. 437: E1-E2.

Reymond P, Weber H, Damond M and Farmer EE. 2000. Differential gene expression in response to mechanical wounding and insect feeding in Arabidopsis. The Plant Cell 12: 707-719.

Richards EJ. 2006. Inherited epigenetic variation – revisiting soft inheritance. Nature Reviews Genetics. 7: 395-401.

Russell PJ. 2005. iGenetics: A Mendelian approach. Benjamin Cummings. pp. 4-5.

Santelices B. 1999. How many kinds of individuals are there. Trends Ecol Evol 4: 152-155.

Santelices B. 2004. Mosaicism and chimerism as components of intraorganismal genetic heterogeneity. Commentary J Evol Biol 17:1187-1188.

Schomberg D, Stephan D. 1997. Adenosine kinase. *In* D Schomberg, D Stephan, eds, Enzyme Handbook, Vol 13. Springer-Verlag, Berlin 2.7.1.20, pp 1-9

Schoor S, Farrow S, Blaschke H, Lee S, Perry G, von Schwartzenberg K, Emery N and Moffatt B. 2011. Adenosine kinase contributes to cytokinin interconversion in Arabidopsis. Plant Physiology 157: 659-672.

Sheldon CC, Conn AB, Dennis ES and Peacock WJ. 2002. Different regulatory regions are required for the vernalization-induced repression of *FLOWERING LOCUS C* and for the epigenetic maintenance of repression. Plant Cell 14: 2527-2537.

Smith RD and Walker JC. 1996. Plant protein phosphatases. Annu Rev Plant Mol Biol 47: 101-125.

Testerink C and Munnik T. 2005. Phosphatidic acid: a multifunctional stress signaling lipid in plants. Trends Plant Sci 10: 368–375.

Thanakiatkrai P and Welch L. 2010. Evaluation of nucleosome forming potentials (NFPs) of forensically important STRs. Forensic Science International: Genetics. 5: 285-290.

Wang H, Buckley KJ, Yang X, Buchmann RC and Bisaro DM. 2005. Adenosine kinase inhibition and suppression of RNA silencing by geminivirus AL2 and L2 proteins. J Virol 79: 7410-7418.

Wang H, Chevalier D, Larue C, Cho SK and Walker JC. 2007. The protein phosphatases and protein kinases of Arabidopsis thaliana. The Arabidopsis Book. The American Society of Plant Biologists. doi: 10.1199/tab.0106

Xu P, Yuan, S, Li Y, Zhang H, Wang X, Lin H and Wu X. 2007. Genome-wide high-frequency non-Mendelian loss of heterozygosity in rice. Genome 50: 297-302.

Appendix A

Transgene constructs

Hybrid lines were generated from crossing transgenic ADK deficient lines in the Columbia background to Landsberg *erecta*. Subsequent generations were produced through self-fertilization.

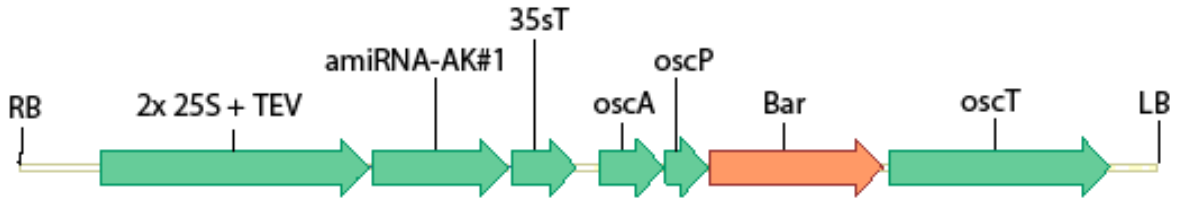


Figure 21: Transgene construct used to generate the amiADK 7-7 ADK deficient Arabidopsis thaliana lines. The artificial microRNA sequence targets ADK1 and ADK2 genes (Schoor et al., 2011).

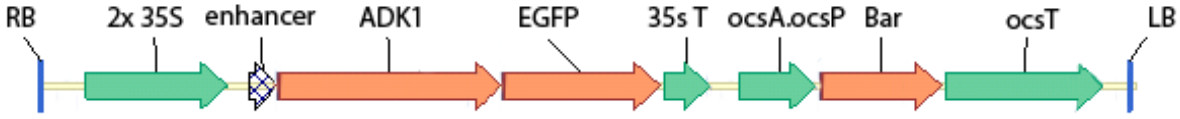


Figure 22: Transgene construct used to generate the fluorescent ADK-deficient *A. thaliana* line ADK1-GFP. An ADK1 cDNA and EGFP fusion protein is expressed in tissues.

Appendix B

QPCR copy number calculations

Table 12: QPCR Copy Number Calculations for LC-2-13 F4 Plants

Sample	Marker	STD Curve (External) [x=copy #, y=C(t)]	R2 value	STD Curve (Internal) [x=copy #, y=C(t)]	R2 value	C(t) Mean (External	C(t) Mean (Internal)	Copy # (External)	Copy # (Internal)	Copy # Int /100,000 Ext
C2-1	T6H20	y = -1.45ln(x) + 38.08	0.996	y = -1.45ln(x) + 36.67	0.996	22.58	36.7	43909.03	0.98	2.23
C2-2	T6H20	"	"	"	"	21.93	33.06	68668.02	12.06	17.56
C2-5	T6H20	"	"	"	"	21.67	29.72	82214.69	120.68	146.78
C2-7	T6H20	"	"	"	"	22.67	35.87	41296.83	1.74	4.20
C2-15	T6H20	"	"	"	"	21.48	34.84	93720.68	3.53	3.77
C2-16	T6H20	"	"	"	"	21.35	37.36	102296.04	0.62	0.61
C2-18	T6H20	"	"	"	"	21.13	28.59	119065.66	263.07	220.94
C2-19	T6H20	"	"	"	"	20.90	36.14	139729.61	1.44	1.03
C2-21	T6H20	"	"	"	"	22.77	34.96	38609.95	3.25	8.42
C2-22	T6H20	"	"	"	"	22.21	35.56	56588.59	2.15	3.80
C4-1	T6H20	y = -1.42ln(x) + 36.66	0.994	y = -1.46ln(x) + 36.17	0.998	20.95	35	63895.29	2.23	3.49
C4-2	T6H20	"	"	"	"	22.81	37.51	17241.25	0.40	2.32
C4-9	T6H20	"	"	"	"	22.44	35.37	22412.47	1.73	7.72
C4-10	T6H20	"	"	"	"	21.06	35.02	59094.10	2.20	3.72
C5-8	T6H20	y = -1.42ln(x) + 36.78	0.995	y = -1.49ln(x) + 35.12	0.999	21.87	32.32	36249.83	6.55	18.06
C5-9	T6H20	"	"	"	"	24.24	34.18	6858.82	1.88	27.46
C5-12	T6H20	y = -1.45ln(x) + 38.08	0.996	y = -1.45ln(x) + 36.67	0.996	20.86	34.04	143750.13	6.13	4.27
C5-14	T6H20	"	"	"	"	21.22	33.63	112153.48	8.14	7.26
C5-15	T6H20	"	"	"	"	22.23	35.04	55930.72	3.08	5.50
C5-16	T6H20	"	"	"	"	21.24	37.05	110743.61	0.77	0.69
C5-19	T6H20	y = -1.45ln(x) + 37.57	0.991	y = -1.52ln(x) + 37.25	0.998	23.24	35.34	19628.43	3.51	17.90
C5-21	T6H20	"	"	"	"	23.03	31.09	22692.06	57.55	253.61
C6-1	T6H20	y = -1.36ln(x) + 39.31	0.991	y = -1.58ln(x) + 40.21	0.999	20.99	30.79	706553.34	388.40	54.97
C6-2	T6H20	"	"	"	"	22.16	33.07	299893.79	91.74	30.59
C6-4	T6H20	"	"	"	"	21.48	32.85	494692.52	105.45	21.32
C6-6	T6H20	"	"	"	"	21.38	35.46	531061.02	20.21	3.81
C6-8	T6H20	"	"	"	"	20.77	33.89	834886.34	54.60	6.54
C6-10	T6H20	"	"	"	"	22.19	32.98	293384.02	97.12	33.10
C6-14	T6H20	y = -1.42ln(x) + 36.78	0.995	y = -1.49ln(x) + 35.12	0.999	23.09	33.02	15393.60	4.10	26.65
C6-18	T6H20	"	"	"	"	23.88	34.35	8815.78	1.68	19.08
C6-20	T6H20	"	"	"	"	22.65	32.90	20920.85	4.43	21.17
C6-21	T6H20	"	"	"	"	23.06	36.79	15750.61	0.33	2.07
C6-22	T6H20	"	"	"	"	23.07	35.65	15615.50	0.70	4.50
C7-1	T6H20	y = -1.45ln(x) + 37.57	0.991	y = -1.52ln(x) + 37.25	0.998	22.43	36.43	34156.60	1.72	5.02
C7-3	T6H20	"	"	"	"	21.86	34.83	50664.73	4.91	9.70
C7-4	T6H20	"	"	"	"	23.38	29.85	17833.80	130.12	729.60
C7-8	T6H20	"	"	"	"	24.05	33	11185.43	16.38	146.44
C7-10	T6H20	"	"	"	"	23.42	32.32	17291.09	25.62	148.18
C7-11	T6H20	"	"	"	"	24.34	37.66	9172.04	0.76	8.33
W1-14	T6H20	y = -1.45ln(x) + 37.57	0.991	y = -1.52ln(x) + 37.25	0.998	23.17	28.64	20567.00	288.44	1402.42
W1-14	F23M2	y = -1.35ln(x) + 37.44	0.995	y = -0.92ln(x) + 30.55	0.927	24.69	28.86	12625.41	6.28	49.72
W1-14	T14G11	y = -1.67ln(x) + 42.19	0.998	y = -1.62ln(x) + 40.32	0.995	25.65	32.16	19975.37	154.01	771.01

Appendix C

PCR primers

PCR primers used in this project are listed within this section. All primers were synthesized by Sigma-Aldrich Canada Ltd., Oakville, ON.

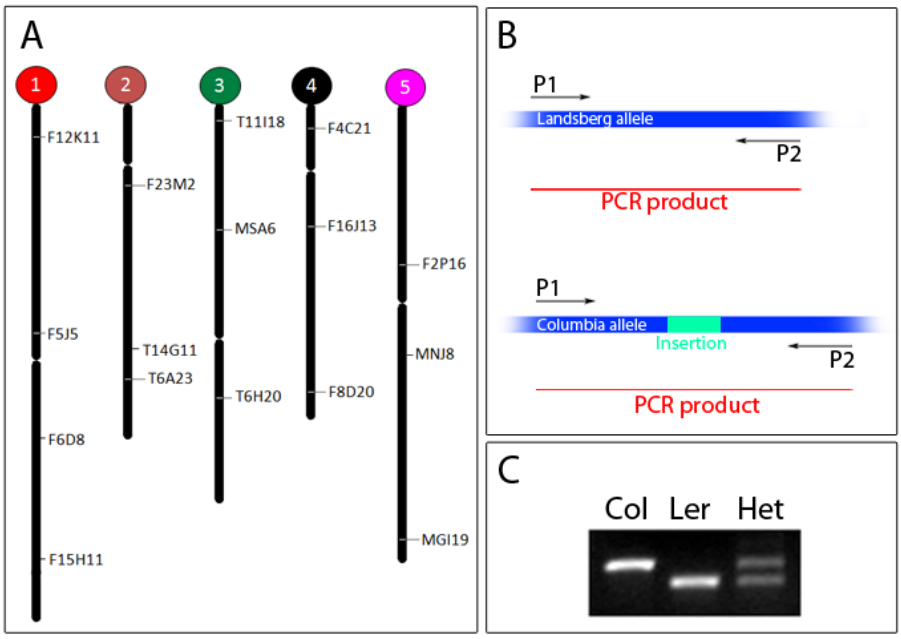


Figure 23: Size-based PCR genotyping; A: approximate location of indel sites on *A. thaliana* chromosomes; B: PCR primers sit outside the insertion site. The amplified product of the Columbia allele is larger than that for the Landsberg allele due to the insertion sequence; C: agarose gel electrophoresis of PCR products for homozygous Col, Ler, and heterozygous alleles.

Table 13: List of primer sequences for size-based PCR genotyping

Primer	Sequence 5' to 3'	Insertion size (bp)	Primer	Sequence 5' to 3'	Insertion size (bp)
F12K11 L	CCATATCTTGGAGTTGGCAGA	45	MSA6 L	CTGGGGTGTTCTCACAGGAT	54
F12K11 R	TGTCTTCAGGAACACAACCA		MSA6 R	CGTTGGAGGTGGTCTTAGGT	
F5J5 L	TGAAGATTTCGTGGAAGCAA	75	T6H20 L	TGCATTGGTTTCTCTGCTTG	77
F5J5 R	CTCATGGATGCCTAATACCG		T6H20 R	GGGAAACCTCCATACTCGAA	
F6D8 L	CTCCGTCTTCCAGAGTTTGA	94	F4C21 L	TGGTTAGGGTTCTGGTCAGG	82
F6D8 R	TTCGGGTGATTAGTACGGAAA		F4C21 R	AGTGGCTCATCGTTCGAGAT	
F15H11 L	ATTTGCGGCTGAAAGACAAG	76	F16J13 L	GAAGCATGTTTTGTGTATCTTGC	80
F15H11 R	TGAGTGTGTCATGAGTGTTTGTTT		F16J13 R	CCGCATCTCCACATTTCATT	
F23M2 L	TAAAGTTGTTGGCCGAGGAG	68	F8D20 L	CACCAGACGGTGATGAAGAG	84
F23M2 R	TCGGAGATACCCGAGCTAAA		F8D20 R	CATTCGCGCATTTATTGTTG	
T14G11 L	CCTATGTGTCAAGAGAGATTTCCA	73	F2P16 L	AAAATGGTTTACCACATGGACA	48
T14G11 R	TTTGTTCCATTTATAAGCGTTTCTC		F2P16 R	TCCCAAATCAATTCAAGGAAA	
T6A23 L	AACACCAAGTCAACTGTTTTTGTT	61	MNJ8 L	CATGGATCAAAGATGATCTCCA	51
T6A23 R	TCAAAATAAACACCCCCAACT		MNJ8 R	TTCGCTTTTCGTGTTTCTGA	
T11I18 L	CCCCAATTCGAAATGTAAGG	74	MGI19 L	TGCACATGACTTCAACAGAAAA	47
T11I18 R	CGCTCCTTGACAGTTTTCCT		MGI19 R	ATGTGGGTGGGTGTTGATTT	

Table 14: Primers used in quantitative PCR

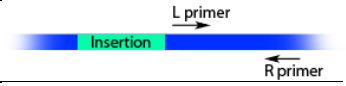
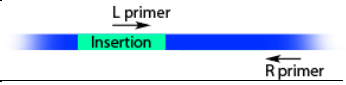
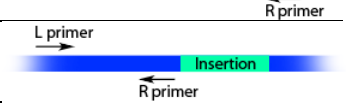
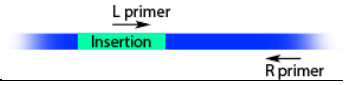
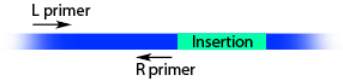
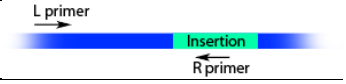
Marker	Primer sequences 5' to 3'		Location of primers with respect to insertion
F23M2	F23M2_extref_L CGAGCAGGGAACCAACAAGG	F23M2_extref_R GCCTCCTCGGCCAACAACTT	
	F23M2_int_L_2 TCCCATTTACGAGTTATCCTCGGTTT	F23M2_extref_R_4 CCGCACTTCGGGTTCAGTCT	
T14G11	T14G11_extref_L CACAAAAATTAAGGAATAATAAATG TTCTC	T14G11R TTTGTTCCATTTATAAGCGTTTCTC	
	T14G11int TTGTCCCATTTTATTTGATGTTTG	T14G11R TTTGTTCCATTTATAAGCGTTTCTC	
T6H20	T6H20_extref_L TGGGCTTACCCTGTTCATGGAG	T6H20_extref_R GCAGAGAAACCAATGCATTTTCA	
	T6H20_exrref_L TGGGCTTACCCTGTTCATGGAG	T6H20_int CCAGAAACCGAGTCTCTAAGATTTC A	

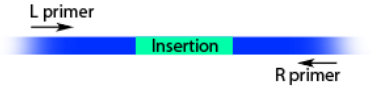
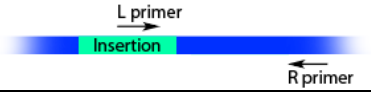
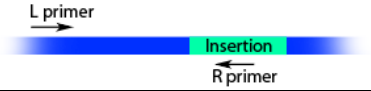
Table 15: List of primer pairs

Primer Pair (5' to 3')		Notes
pSAT-F CATTTGGAGAGGACGTCGAG	pXCS-R CTGGTGATTTTTGCGGACTC	Used for detection of amiADK transgene
adktestp-RF CTCTGGTTGAGAAGGCCAAG	EGFPm-R GAACTTCAGGGTCAGCTTGC	Used for detection of ADKGFP transgene
adktestp-RF CTCTGGTTGAGAAGGCCAAG	adktestp-R AGCTTCTCTTTGGGGAGAGG	Used for detection of genomic and cDNA ADK1 gene
T14G11-IndMid L GAGTTGTGTTCCAGGGCCTA	T14G11-IndMid R TTTGTTGTGCGAATTCATTG	Used for synthesizing qPCR standard DNA , T14G11 locus
F23M2-IndMid L GCACAGAAGGCTGCTAAACC	R-F23M2-IndMid R ATGGAAGGCAAAACAGTTCG	Used for synthesizing qPCR standard DNA, F23M2 locus
T6H20-IndMid L TTTCCTGTTTGGGATCTGAG	T6H20-IndMid R TCAGGAGATAGTCCACCATGC	Used for synthesizing qPCR standard DNA, T6H20 locus

Table 16: List of sequencing primers

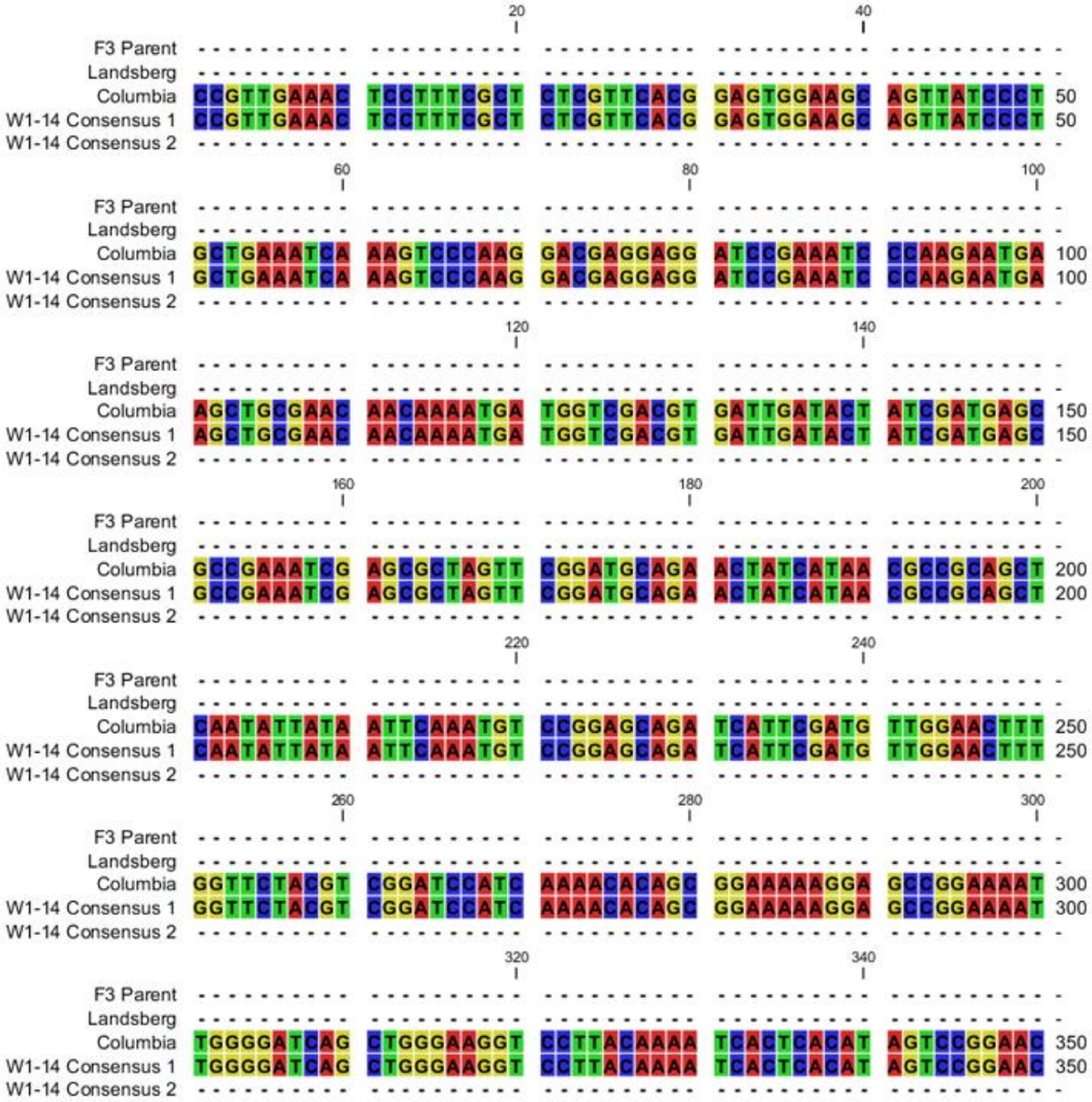
Primer name	Sequence (5’ to 3’)	Primer name	Sequence (5’ to 3’)
F23M2_seqext_L2	GACGGCCATTGATAATGAAA	T14G11_seqext_L2	CTCCGGTATTACGATGCTTTT
F23M2_seqext_R1	CCAAGAAGAGCTACCGTTGA	T14G11_seqint_R2	CAAACATCAAATAAAATGGGACA
F23M2_seqint_L1	TTGCCGGAAAATAACTAAGTG	T6H20_seqint_L2	CATTTGTCTTTAGGCGATTGAT
F23M2_seqint_R1	TCGCGGTCTTAACTGATCTT	T6H20_seqext_R1	CCCATCCTCTACAACCTGTG
T14G11_seqint_L1	TGCAATCCAAGTATTTTCTTTTT	T6H20_seqext_L2	TTCTTATCTTGGCGATCGAA
T14G11_seqext_R1	GAATTTTCTAGCTCTTCACAAAGC	T6H20_seqint_R1	CAACATCAACCTAGGTATTAACCA

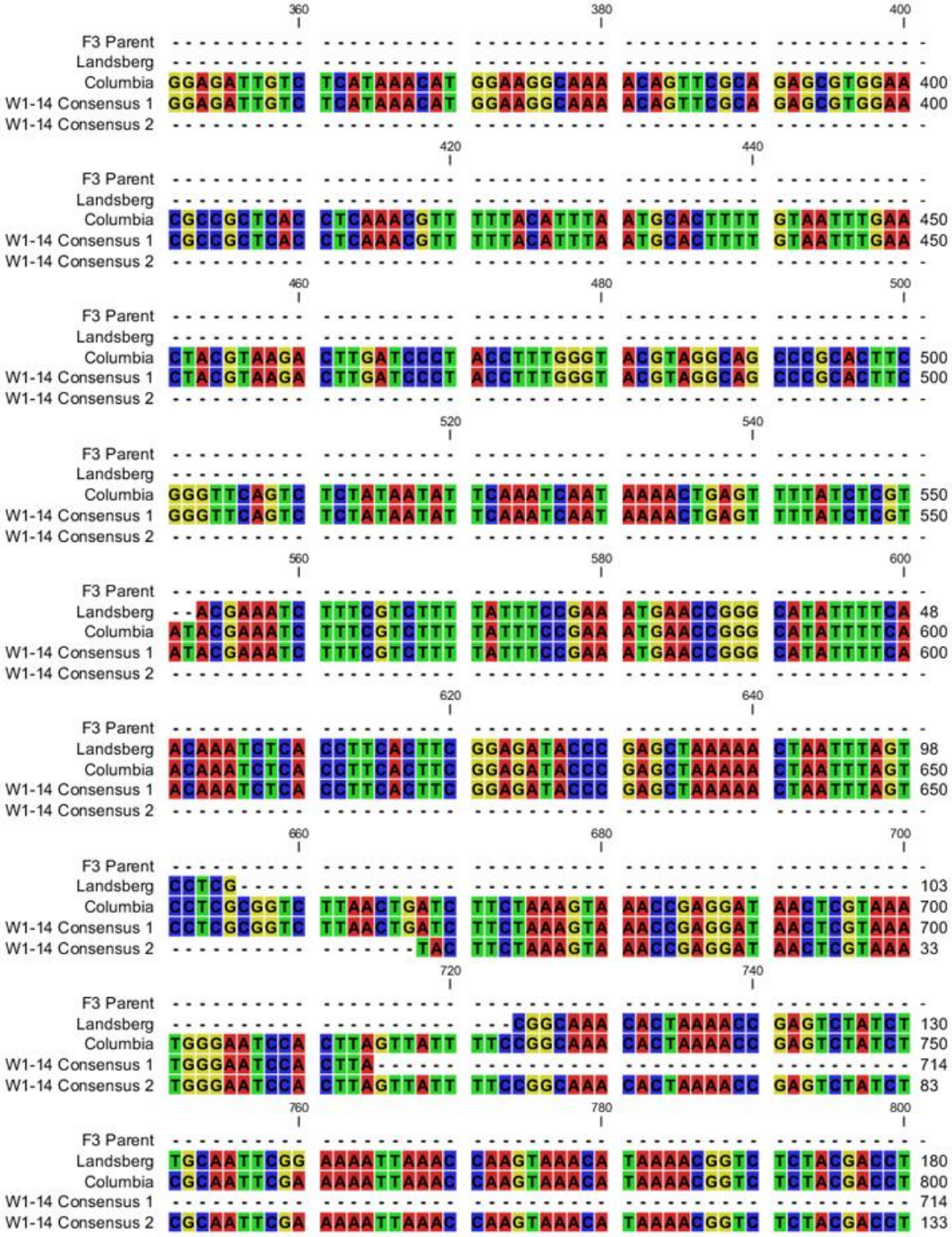
Table 17: Primer pairs used for synthesizing PCR products for sequencing

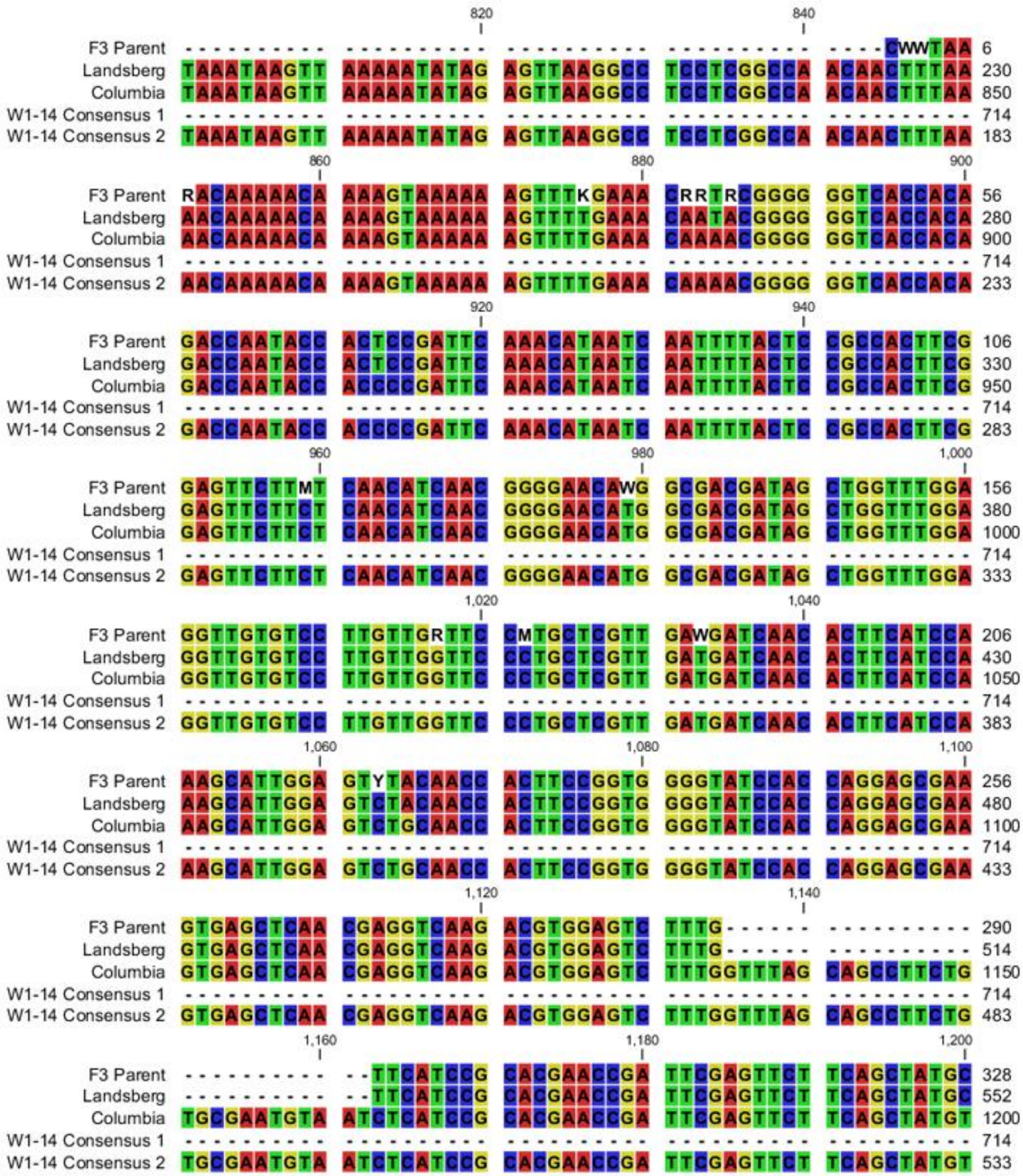
Sample	Primer pair		Location of primers with respect to insertion
LC-2-13 #W1	F23M2_seqext_L2	F23M2_seqext_R1	
	T14G11_seqext_L2	T14G11_seqext_R1	
	T6H20_seqext_L1	T6H20_seqext_R1	
LC-2-13-W1 #14	F23M2_seqint_L1	F23M2_seqext_R1	
	T14G11_seqint_L1	T14G11_seqext_R1	
	T6H20_seqint_L2	T6H20_seqext_R1	
	F23M2_seqext_L2	F23M2_seqint_R1	
	T14G11_seqext_L2	T14G11_seqint_R2	
	T6H20_seqext_L2	T6H20_seqint_R1	

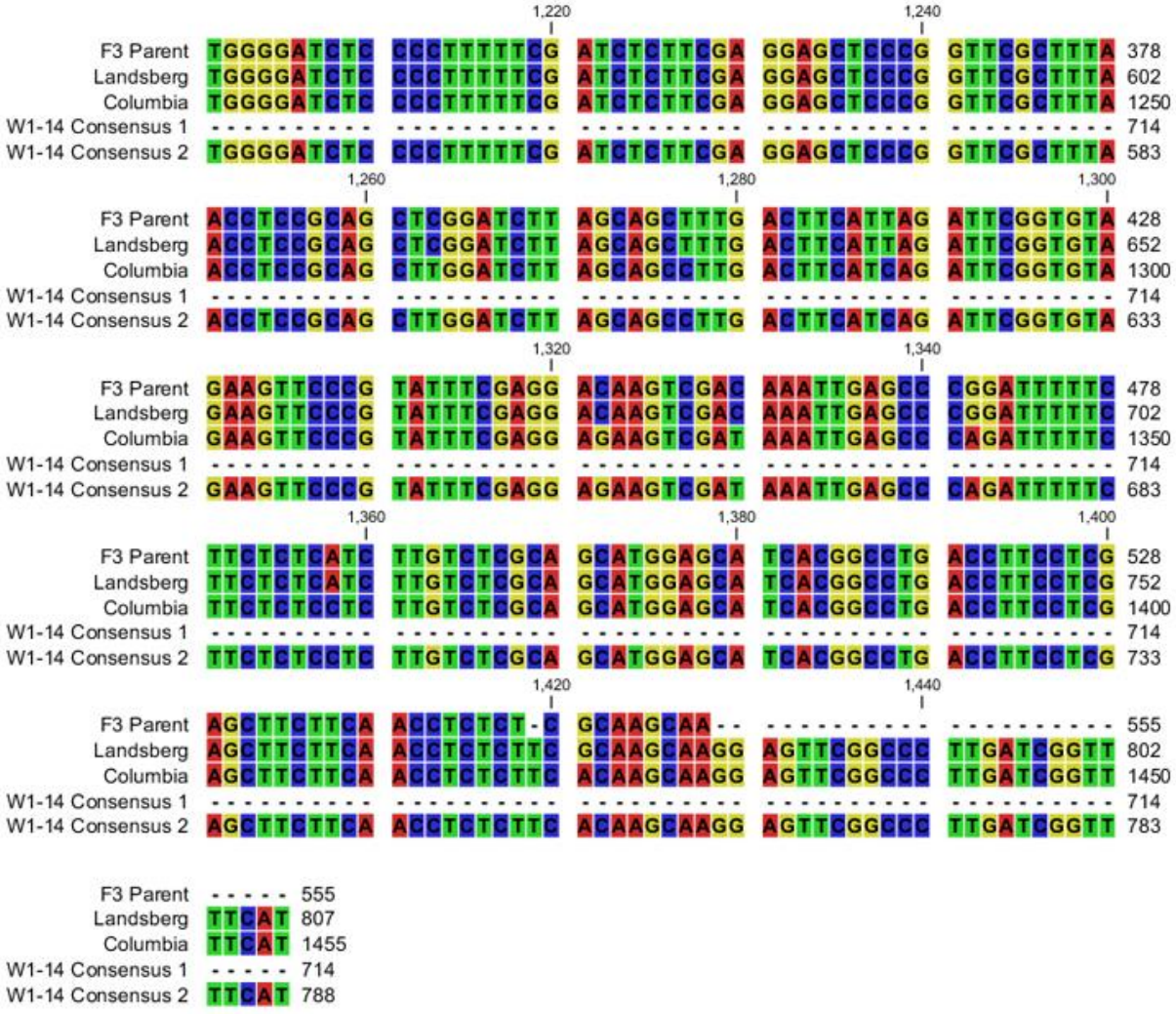
Appendix D

F23M2 sequence alignment



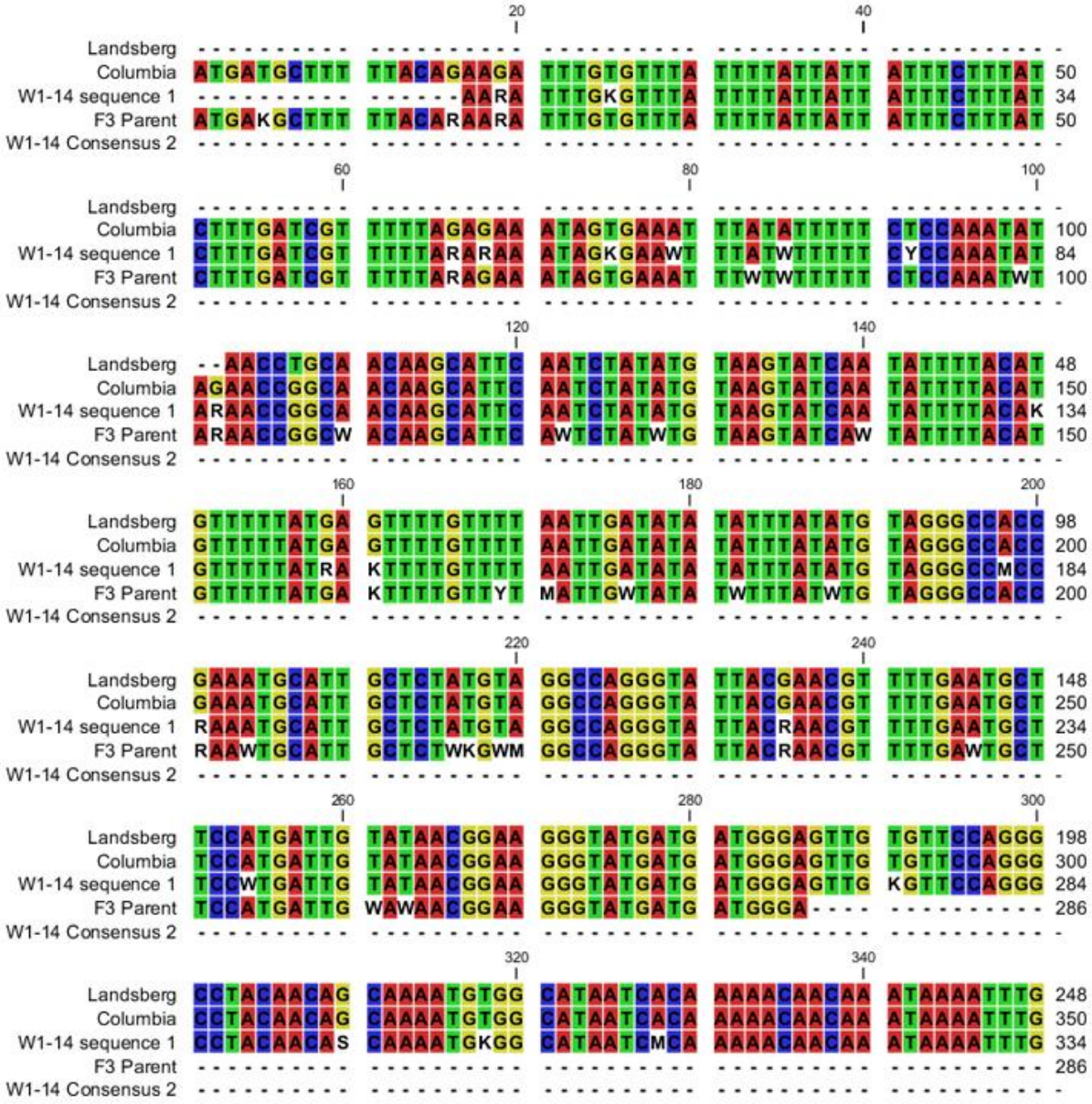

360
380
400
F3 Parent
Landsberg
Columbia GGAGATTGTC TCATAAACAT GGAAGGCAAA ACAGTTCGCA GAGCGTGGAA 400
W1-14 Consensus 1 GGAGATTGTC TCATAAACAT GGAAGGCAAA ACAGTTCGCA GAGCGTGGAA 400
W1-14 Consensus 2
420
440
F3 Parent
Landsberg
Columbia CGCCGCTCAC CTCAAACGTT TTTACATTTA ATGCACTTTT GTAATTTGAA 450
W1-14 Consensus 1 CGCCGCTCAC CTCAAACGTT TTTACATTTA ATGCACTTTT GTAATTTGAA 450
W1-14 Consensus 2
460
480
500
F3 Parent
Landsberg
Columbia CTACGTAAGA CTTGATCCCT ACCTTTGGGT ACGTAGGCAG CCCGCACTTC 500
W1-14 Consensus 1 CTACGTAAGA CTTGATCCCT ACCTTTGGGT ACGTAGGCAG CCCGCACTTC 500
W1-14 Consensus 2
520
540
F3 Parent
Landsberg
Columbia GGGTTCAGTC TCTATAATAT TCAAATCAAT AAAACTGAGT TTTATCTCGT 550
W1-14 Consensus 1 GGGTTCAGTC TCTATAATAT TCAAATCAAT AAAACTGAGT TTTATCTCGT 550
W1-14 Consensus 2
560
580
600
F3 Parent
Landsberg --ACGAAATC TTTCGTCTTT TATTTCCGAA ATGAACCGGG CATATTTTCA 48
Columbia ATACGAAATC TTTCGTCTTT TATTTCCGAA ATGAACCGGG CATATTTTCA 600
W1-14 Consensus 1 ATACGAAATC TTTCGTCTTT TATTTCCGAA ATGAACCGGG CATATTTTCA 600
W1-14 Consensus 2
620
640
F3 Parent
Landsberg ACAAATCTCA CCTTCACTTC GGAGATACCC GAGCTAAAAA CTAATTTAGT 98
Columbia ACAAATCTCA CCTTCACTTC GGAGATACCC GAGCTAAAAA CTAATTTAGT 650
W1-14 Consensus 1 ACAAATCTCA CCTTCACTTC GGAGATACCC GAGCTAAAAA CTAATTTAGT 650
W1-14 Consensus 2
660
680
700
F3 Parent
Landsberg CCTCG 103
Columbia CCTCGCGGTC TTAACTGATC TTCTAAAGTA AACCGAGGAT AACTCGTAAA 700
W1-14 Consensus 1 CCTCGCGGTC TTAACTGATC TTCTAAAGTA AACCGAGGAT AACTCGTAAA 700
W1-14 Consensus 2 TAC TTCTAAAGTA AACCGAGGAT AACTCGTAAA 33
720
740
F3 Parent
Landsberg CGGCAAA CACTAAAACC GAGTCTATCT 130
Columbia TGGGAATCCA CTTAGTTATT TTCCGGCAAA CACTAAAACC GAGTCTATCT 750
W1-14 Consensus 1 TGGGAATCCA CTTA 714
W1-14 Consensus 2 TGGGAATCCA CTTAGTTATT TTCCGGCAAA CACTAAAACC GAGTCTATCT 83
760
780
800
F3 Parent
Landsberg TGCAATTCGG AAAATTAAAC CAAGTAAACA TAAAACGGTC TCTACGACCT 180
Columbia CGCAATTCGA AAAATTAAAC CAAGTAAACA TAAAACGGTC TCTACGACCT 800
W1-14 Consensus 1 714
W1-14 Consensus 2 CGCAATTCGA AAAATTAAAC CAAGTAAACA TAAAACGGTC TCTACGACCT 133

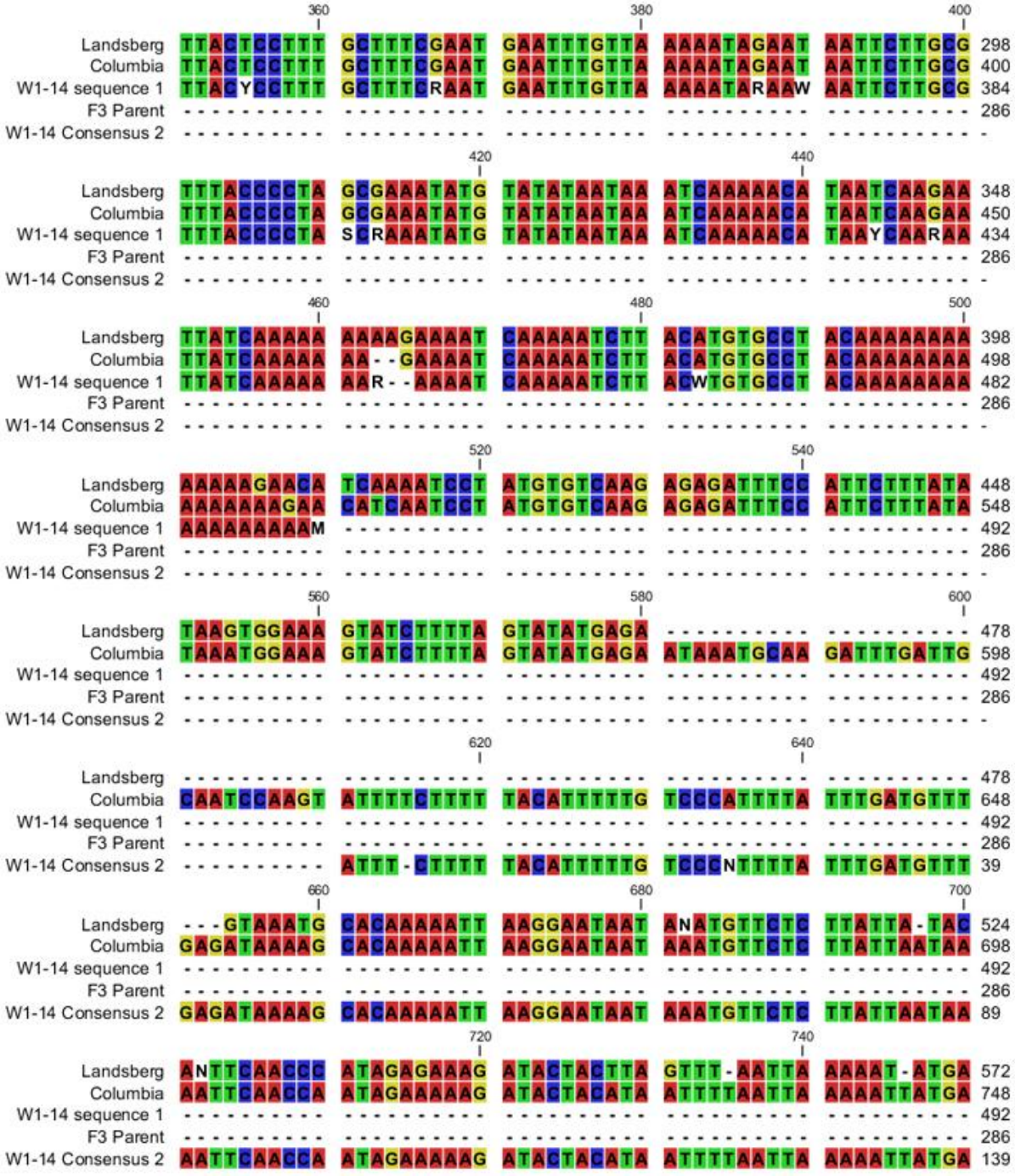

820
840
F3 Parent ---------- ---------- ---------- ---------- ----CWWTAA 6
Landsberg TAAATAAGTT AAAAATATAG AGTTAAGGCC TCCTCGGCCA ACAACTTTAA 230
Columbia TAAATAAGTT AAAAATATAG AGTTAAGGCC TCCTCGGCCA ACAACTTTAA 850
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 TAAATAAGTT AAAAATATAG AGTTAAGGCC TCCTCGGCCA ACAACTTTAA 183
860
880
900
F3 Parent RACAAAAACA AAAGTAAAAA AGTTTKGAAA CRRTRCGGGG GGTCACCACA 56
Landsberg AACAAAAACA AAAGTAAAAA AGTTTTGAAA CAATACGGGG GGTCACCACA 280
Columbia AACAAAAACA AAAGTAAAAA AGTTTTGAAA CAAAACGGGG GGTCACCACA 900
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 AACAAAAACA AAAGTAAAAA AGTTTTGAAA CAAAACGGGG GGTCACCACA 233
920
940
F3 Parent GACCAATACC ACTCCGATTC AAACATAATC AATTTTACTC CGCCACTTCG 106
Landsberg GACCAATACC ACTCCGATTC AAACATAATC AATTTTACTC CGCCACTTCG 330
Columbia GACCAATACC ACCCCGATTC AAACATAATC AATTTTACTC CGCCACTTCG 950
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 GACCAATACC ACCCCGATTC AAACATAATC AATTTTACTC CGCCACTTCG 283
960
980
1,000
F3 Parent GAGTTCTTMT CAACATCAAC GGGGAACAWG GCGACGATAG CTGGTTTGGA 156
Landsberg GAGTTCTTCT CAACATCAAC GGGGAACATG GCGACGATAG CTGGTTTGGA 380
Columbia GAGTTCTTCT CAACATCAAC GGGGAACATG GCGACGATAG CTGGTTTGGA 1000
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 GAGTTCTTCT CAACATCAAC GGGGAACATG GCGACGATAG CTGGTTTGGA 333
1,020
1,040
F3 Parent GGTTGTGTCC TTGTTGRTTC CMTGCTCGTT GAWGATCAAC ACTTCATCCA 206
Landsberg GGTTGTGTCC TTGTTGGTTC CCTGCTCGTT GATGATCAAC ACTTCATCCA 430
Columbia GGTTGTGTCC TTGTTGGTTC CCTGCTCGTT GATGATCAAC ACTTCATCCA 1050
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 GGTTGTGTCC TTGTTGGTTC CCTGCTCGTT GATGATCAAC ACTTCATCCA 383
1,060
1,080
1,100
F3 Parent AAGCATTGGA GTYTACAACC ACTTCCGGTG GGGTATCCAC CAGGAGCGAA 256
Landsberg AAGCATTGGA GTCTACAACC ACTTCCGGTG GGGTATCCAC CAGGAGCGAA 480
Columbia AAGCATTGGA GTCTGCAACC ACTTCCGGTG GGGTATCCAC CAGGAGCGAA 1100
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 AAGCATTGGA GTCTGCAACC ACTTCCGGTG GGGTATCCAC CAGGAGCGAA 433
1,120
1,140
F3 Parent GTGAGCTCAA CGAGGTCAAG ACGTGGAGTC TTTG------ ---------- 290
Landsberg GTGAGCTCAA CGAGGTCAAG ACGTGGAGTC TTTG------ ---------- 514
Columbia GTGAGCTCAA CGAGGTCAAG ACGTGGAGTC TTTGGTTTAG CAGCCTTCTG 1150
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 GTGAGCTCAA CGAGGTCAAG ACGTGGAGTC TTTGGTTTAG CAGCCTTCTG 483
1,160
1,180
1,200
F3 Parent ---------- --TTCATCCG CACGAACCGA TTCGAGTTCT TCAGCTATGC 328
Landsberg ---------- --TTCATCCG CACGAACCGA TTCGAGTTCT TCAGCTATGC 552
Columbia TGCGAATGTA ATCTCATCCG CACGAACCGA TTCGAGTTCT TCAGCTATGT 1200
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 TGCGAATGTA ATCTCATCCG CACGAACCGA TTCGAGTTCT TCAGCTATGT 533

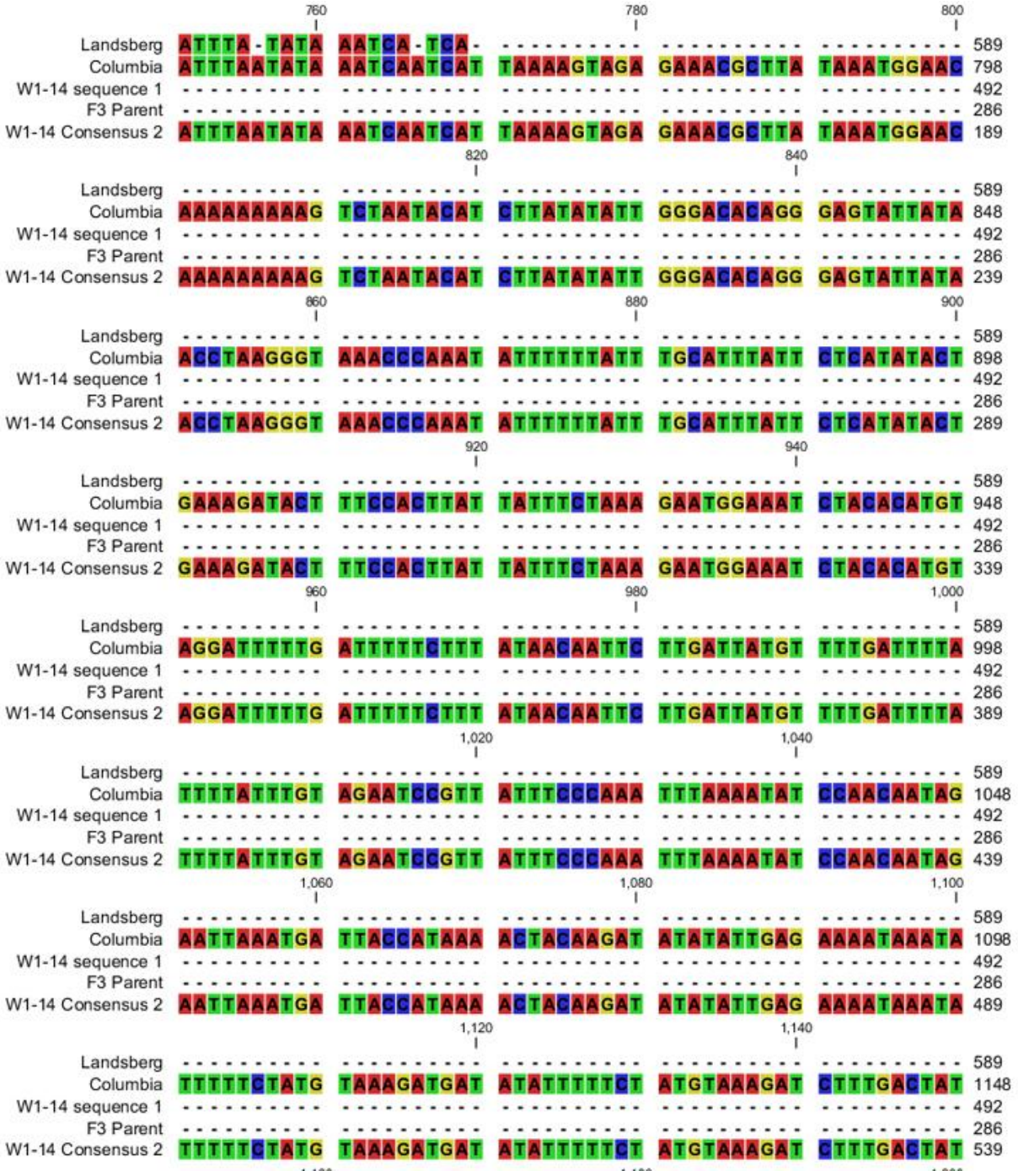

1,220
1,240
F3 Parent TGGGGATCTC CCCTTTTTCG ATCTCTTCGA GGAGCTCCCG GTTCGCTTTA 378
Landsberg TGGGGATCTC CCCTTTTTCG ATCTCTTCGA GGAGCTCCCG GTTCGCTTTA 602
Columbia TGGGGATCTC CCCTTTTTCG ATCTCTTCGA GGAGCTCCCG GTTCGCTTTA 1250
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 TGGGGATCTC CCCTTTTTCG ATCTCTTCGA GGAGCTCCCG GTTCGCTTTA 583
1,260
1,280
1,300
F3 Parent ACCTCCGCAG CTCGGATCTT AGCAGCTTTG ACTTCATTAG ATTCGGTGTA 428
Landsberg ACCTCCGCAG CTCGGATCTT AGCAGCTTTG ACTTCATTAG ATTCGGTGTA 652
Columbia ACCTCCGCAG CTTGGATCTT AGCAGCCTTG ACTTCATCAG ATTCGGTGTA 1300
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 ACCTCCGCAG CTTGGATCTT AGCAGCCTTG ACTTCATCAG ATTCGGTGTA 633
1,320
1,340
F3 Parent GAAGTTCCCG TATTTCGAGG ACAAGTCGAC AAATTGAGCC CGGATTTTTC 478
Landsberg GAAGTTCCCG TATTTCGAGG ACAAGTCGAC AAATTGAGCC CGGATTTTTC 702
Columbia GAAGTTCCCG TATTTCGAGG AGAAGTCGAT AAATTGAGCC CAGATTTTTC 1350
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 GAAGTTCCCG TATTTCGAGG AGAAGTCGAT AAATTGAGCC CAGATTTTTC 683
1,360
1,380
1,400
F3 Parent TTCTCTCATC TTGTCTCGCA GCATGGAGCA TCACGGCCTG ACCTTCCTCG 528
Landsberg TTCTCTCATC TTGTCTCGCA GCATGGAGCA TCACGGCCTG ACCTTCCTCG 752
Columbia TTCTCTCCTC TTGTCTCGCA GCATGGAGCA TCACGGCCTG ACCTTCCTCG 1400
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 TTCTCTCCTC TTGTCTCGCA GCATGGAGCA TCACGGCCTG ACCTTCCTCG 733
1,420
1,440
F3 Parent AGCTTCTTCA ACCTCTCT-C GCAAGCAA-- ---------- ---------- 555
Landsberg AGCTTCTTCA ACCTCTCTTC GCAAGCAAGG AGTTCGGCCC TTGATCGGTT 802
Columbia AGCTTCTTCA ACCTCTCTTC ACAAGCAAGG AGTTCGGCCC TTGATCGGTT 1450
W1-14 Consensus 1 ---------- ---------- ---------- ---------- ---------- 714
W1-14 Consensus 2 AGCTTCTTCA ACCTCTCTTC ACAAGCAAGG AGTTCGGCCC TTGATCGGTT 783
F3 Parent ----- 555
Landsberg TTCAT 807
Columbia TTCAT 1455
W1-14 Consensus 1 ----- 714
W1-14 Consensus 2 TTCAT 788

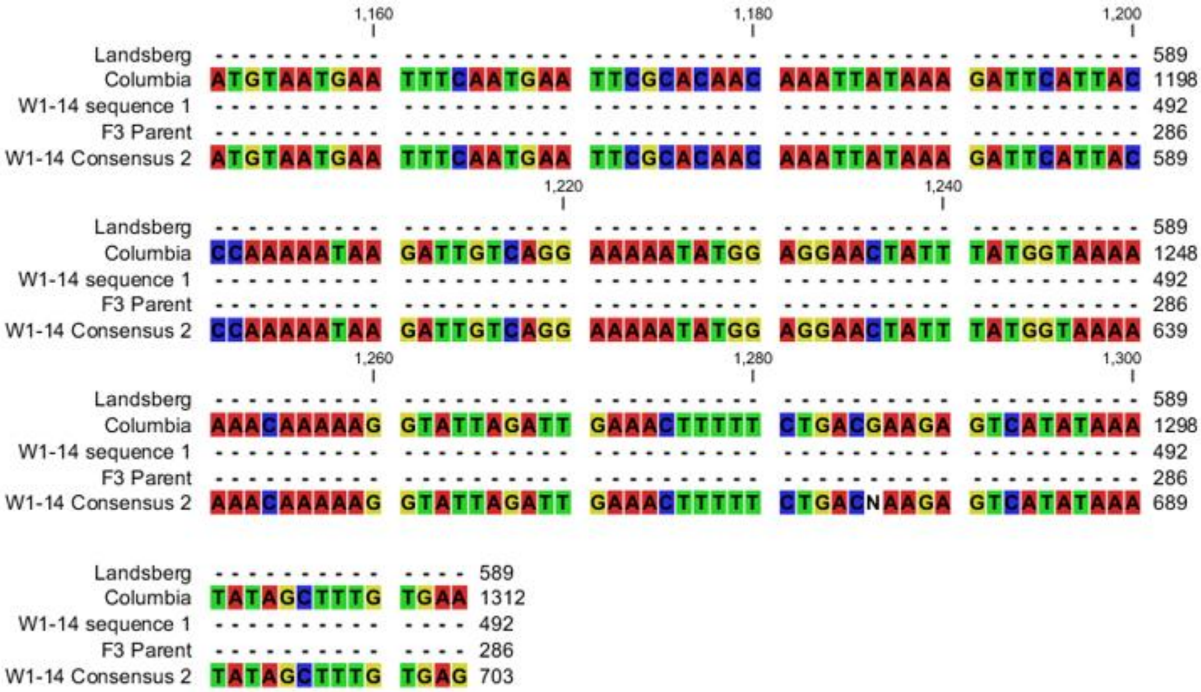
Appendix E

T14G11 sequence alignment



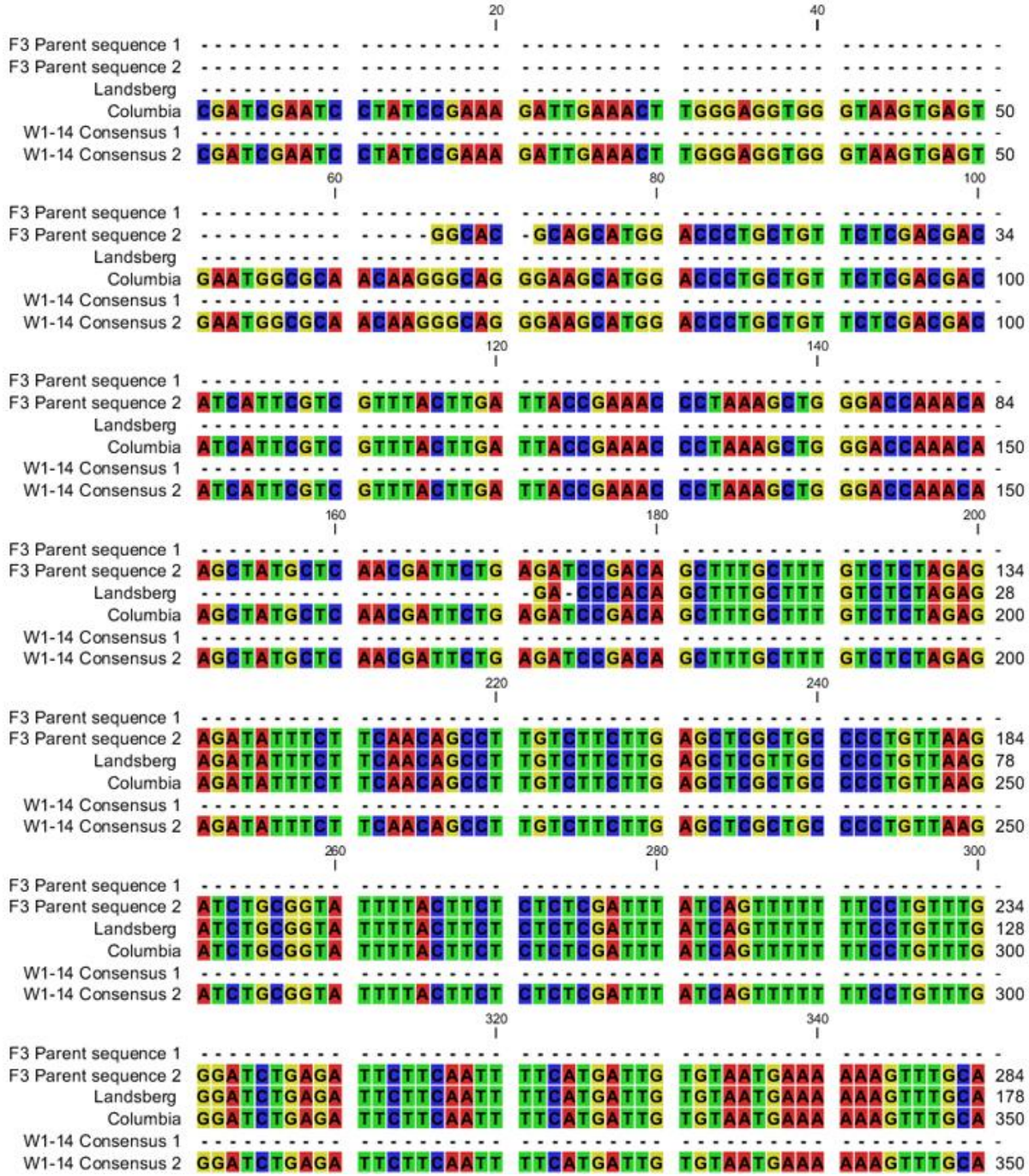

360
380
400
Landsberg TTACTCCTTT GCTTTCGAAT GAATTTGTTA AAAATAGAAT AATTGTTGCG 298
Columbia TTACTCCTTT GCTTTCGAAT GAATTTGTTA AAAATAGAAT AATTCTTGCG 400
W1-14 sequence 1 TTACYCCTTT GCTTTCRAAT GAATTTGTTA AAAATARAAW AATTCTTGCG 384
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- -
420
440
Landsberg TTTACCCCTA GCGAAATATG TATATAATAA ATCAAAAACA TAATCAAGAA 348
Columbia TTTACCCCTA GCGAAATATG TATATAATAA ATCAAAAACA TAATCAAGAA 450
W1-14 sequence 1 TTTACCCCTA SCRAAATATG TATATAATAA ATCAAAAACA TAAYCAARAA 434
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- -
460
480
500
Landsberg TTATCAAAAA AAAAGAAAAT CAAAAATCTT ACATGTGCCT ACAAAAAAAA 398
Columbia TTATCAAAAA AA--GAAAAT CAAAAATCTT ACATGTGCCT ACAAAAAAAA 498
W1-14 sequence 1 TTATCAAAAA AAR--AAAAT CAAAAATCTT ACWTGTGCCT ACAAAAAAAA 482
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- -
520
540
Landsberg AAAAAGAACA TCAAAATCCT ATGTGTCAAG AGAGATTTCC ATTCTTTATA 448
Columbia AAAAAAAGAA CATCAATCCT ATGTGTCAAG AGAGATTTCC ATTCTTTATA 548
W1-14 sequence 1 AAAAAAAAAM ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- -
560
580
600
Landsberg TAAGTGGAAA GTATCTTTTA GTATATGAGA ---------- ---------- 478
Columbia TAAATGGAAA GTATCTTTTA GTATATGAGA ATAAATGCAA GATTTGATTG 598
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- -
620
640
Landsberg ---------- ---------- ---------- ---------- ---------- 478
Columbia CAATCCAAGT ATTTTCTTTT TACATTTTTG TCCCATTTTA TTTGATGTTT 648
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ---------- ATTT-CTTTT TACATTTTTG TCCCNTTTTA TTTGATGTTT 39
660
680
700
Landsberg ---GTAAATG CACAAAAATT AAGGAATAAT ANATGTTGTC TTATTA-TAC 524
Columbia GAGATAAAAG CACAAAAATT AAGGAATAAT AAATGTTGTC TTATTAATAA 698
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 GAGATAAAAG CACAAAAATT AAGGAATAAT AAATGTTGTC TTATTAATAA 89
720
740
Landsberg ANTTCAACCC ATAGAGAAAG ATACTACTTA GTTT-AATTA AAAAT-ATGA 572
Columbia AATTCAACCA ATAGAAAAAG ATACTACATA ATTTTAATTA AAAATTATGA 748
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 AATTCAACCA ATAGAAAAAG ATACTACATA ATTTTAATTA AAAATTATGA 139

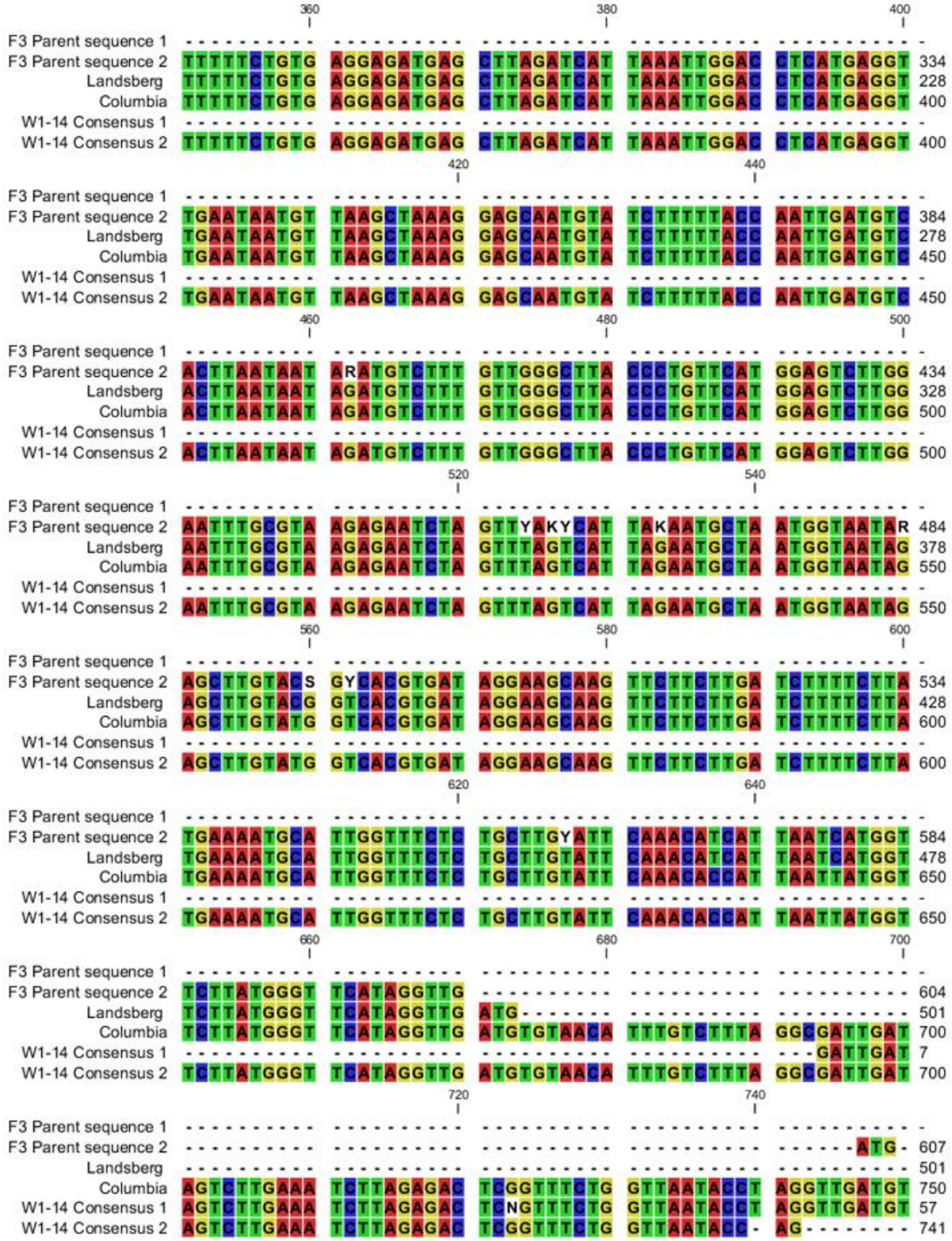

760
780
800
Landsberg ATTTA-TATA AATCA-TCA- ---------- ---------- ---------- 589
Columbia ATTTAATATA AATCAATCAT TAAAAGTAGA GAAACGCTTA TAAATGGAAC 798
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ATTTAATATA AATCAATCAT TAAAAGTAGA GAAACGCTTA TAAATGGAAC 189
820
840
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia AAAAAAAAAG TCTAATACAT CTTATATATT GGGACACAGG GAGTATTATA 848
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 AAAAAAAAAG TCTAATACAT CTTATATATT GGGACACAGG GAGTATTATA 239
860
880
900
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia ACCTAAGGGT AAACCCAAAT ATTTTTTATT TGCATTTATT CTCATATACT 898
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ACCTAAGGGT AAACCCAAAT ATTTTTTATT TGCATTTATT CTCATATACT 289
920
940
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia GAAAGATACT TTCCACTTAT TATTTCTAAA GAATGGAAAT CTACACATGT 948
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 GAAAGATACT TTCCACTTAT TATTTCTAAA GAATGGAAAT CTACACATGT 339
960
980
1,000
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia AGGATTTTTG ATTTTTCTTT ATAACAATTC TTGATTATGT TTTGATTTTA 998
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 AGGATTTTTG ATTTTTCTTT ATAACAATTC TTGATTATGT TTTGATTTTA 389
1,020
1,040
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia TTTTATTTGT AGAATCCGTT ATTTCCCAAA TTTAAAATAT CCAACAATAG 1048
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 TTTTATTTGT AGAATCCGTT ATTTCCCAAA TTTAAAATAT CCAACAATAG 439
1,060
1,080
1,100
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia AATTAAATGA TTACCATAAA ACTACAAGAT ATATATTGAG AAAATAAATA 1098
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 AATTAAATGA TTACCATAAA ACTACAAGAT ATATATTGAG AAAATAAATA 489
1,120
1,140
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia TTTTTCTATG TAAAGATGAT ATATTTTTCT ATGTAAAGAT CTTTGACTAT 1148
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 TTTTTCTATG TAAAGATGAT ATATTTTTCT ATGTAAAGAT CTTTGACTAT 539

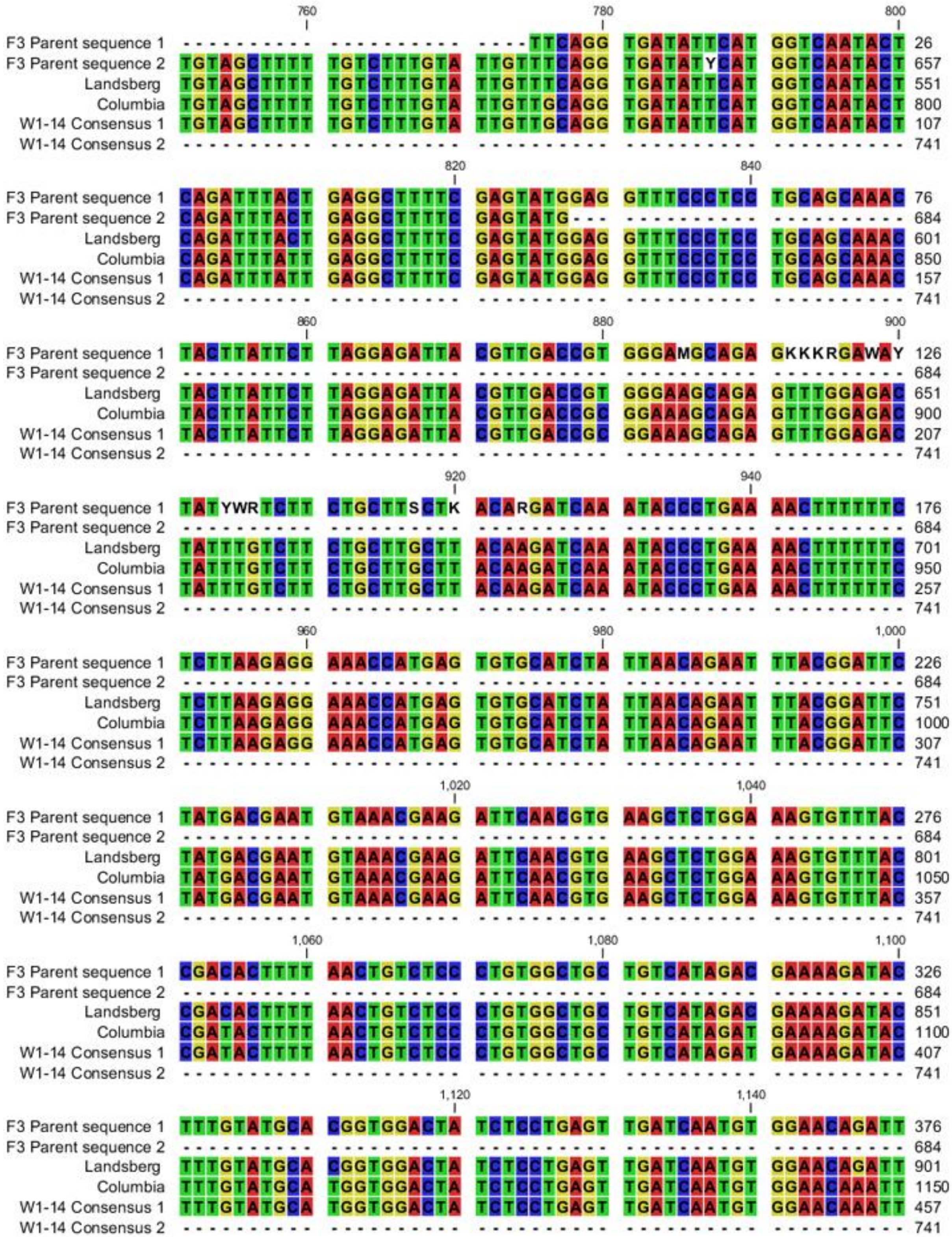

1,160
1,180
1,200
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia ATGTAATGAA TTTCAATGAA TTCGCACAAC AAATTATAAA GATTCATTAC 1198
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 ATGTAATGAA TTTCAATGAA TTCGCACAAC AAATTATAAA GATTCATTAC 589
1,220
1,240
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia CCAAAAATAA GATTGTCAGG AAAAATATGG AGGAACTATT TATGGTAAAA 1248
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 CCAAAAATAA GATTGTCAGG AAAAATATGG AGGAACTATT TATGGTAAAA 639
1,260
1,280
1,300
Landsberg ---------- ---------- ---------- ---------- ---------- 589
Columbia AAACAAAAAG GTATTAGATT GAAACTTTTT CTGACGAAGA GTCATATAAA 1298
W1-14 sequence 1 ---------- ---------- ---------- ---------- ---------- 492
F3 Parent ---------- ---------- ---------- ---------- ---------- 286
W1-14 Consensus 2 AAACAAAAAG GTATTAGATT GAAACTTTTT CTGACNAAGA GTCATATAAA 689
Landsberg ---------- ---- 589
Columbia TATAGCTTTG TGAA 1312
W1-14 sequence 1 ---------- ---- 492
F3 Parent ---------- ---- 286
W1-14 Consensus 2 TATAGCTTTG TGAG 703

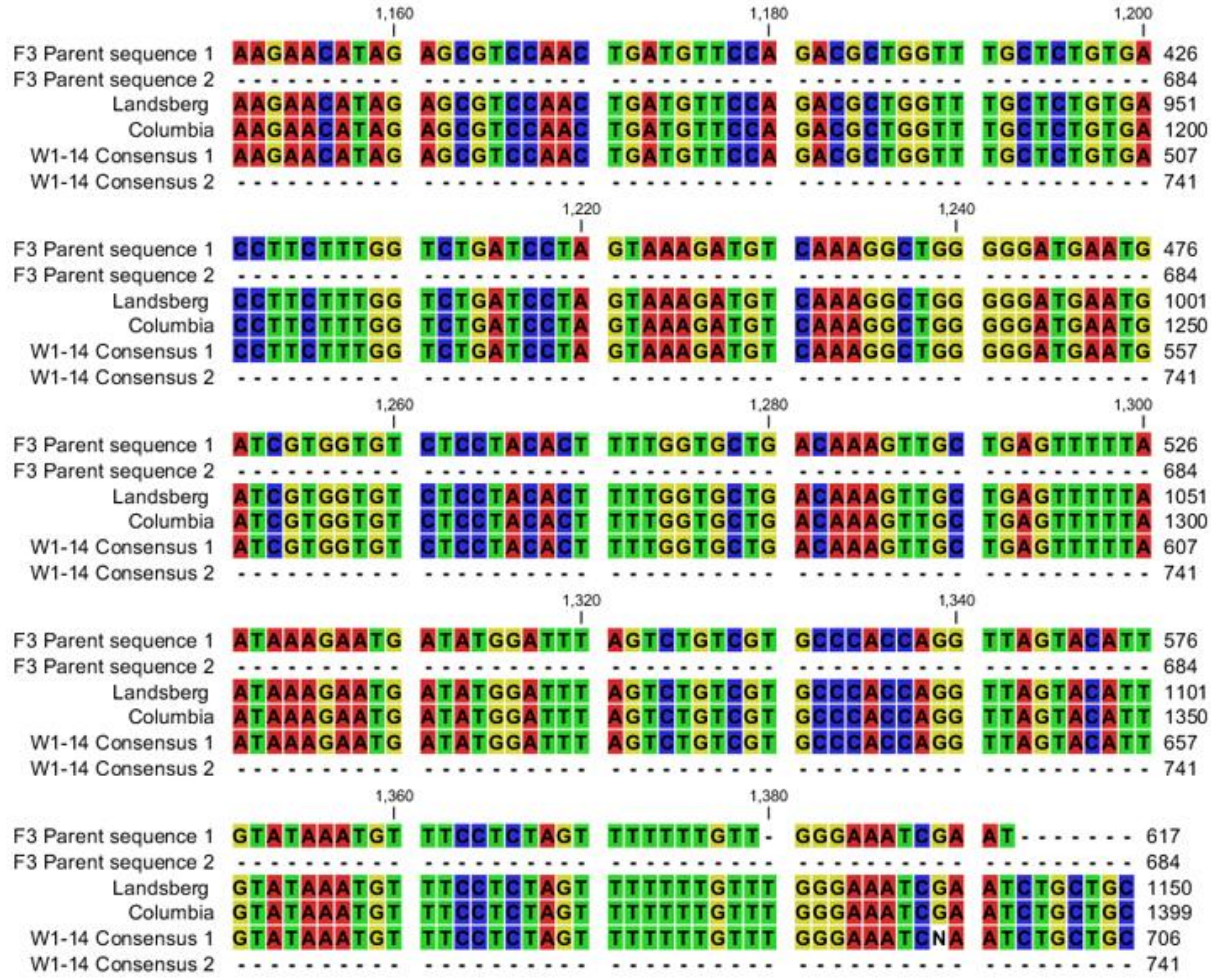
Appendix F

T6H20 sequence alignments




360
380
400
F3 Parent sequence 1
F3 Parent sequence 2
Landsberg
Columbia
W1-14 Consensus 1
W1-14 Consensus 2
420
440
460
480
500
520
540
560
580
600
620
640
660
680
700
720
740


760 780 800
F3 Parent sequence 1 ---------- ---------- ----TTCAGG TGATATTCAT GGTCAATACT 26
F3 Parent sequence 2 TGTAGCTTTT TGTCTTTGTA TTGTTTCAGG TGATATYCAT GGTCAATACT 657
Landsberg TGTAGCTTTT TGTCTTTGTA TTGTTTCAGG TGATATTCAT GGTCAATACT 551
Columbia TGTAGCTTTT TGTCTTTGTA TTGTTGCAGG TGATATTCAT GGTCAATACT 800
W1-14 Consensus 1 TGTAGCTTTT TGTCTTTGTA TTGTTGCAGG TGATATTCAT GGTCAATACT 107
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741
820 840
F3 Parent sequence 1 CAGATTTACT GAGGCTTTTC GAGTATGGAG GTTTCCCTCC TGCAGCAAAC 76
F3 Parent sequence 2 CAGATTTACT GAGGCTTTTC GAGTATG--- ---------- ---------- 684
Landsberg CAGATTTACT GAGGCTTTTC GAGTATGGAG GTTTCCCTCC TGCAGCAAAC 601
Columbia CAGATTTATT GAGGCTTTTC GAGTATGGAG GTTTCCCTCC TGCAGCAAAC 850
W1-14 Consensus 1 CAGATTTATT GAGGCTTTTC GAGTATGGAG GTTTCCCTCC TGCAGCAAAC 157
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741
860 880 900
F3 Parent sequence 1 TACTTATTCT TAGGAGATTA CGTTGACCGT GGGAMGCAGA GKKKRGAWAY 126
F3 Parent sequence 2 ---------- ---------- ---------- ---------- ---------- 684
Landsberg TACTTATTCT TAGGAGATTA CGTTGACCGT GGGAAGCAGA GTTTGGAGAC 651
Columbia TACTTATTCT TAGGAGATTA CGTTGACCGC GGAAAGCAGA GTTTGGAGAC 900
W1-14 Consensus 1 TACTTATTCT TAGGAGATTA CGTTGACCGC GGAAAGCAGA GTTTGGAGAC 207
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741
920 940
F3 Parent sequence 1 TATYWRTCTT CTGCTTSCTK ACARGATCAA ATACCCTGAA AACTTTTTTC 176
F3 Parent sequence 2 ---------- ---------- ---------- ---------- ---------- 684
Landsberg TATTTGTCTT CTGCTTGCTT ACAAGATCAA ATACCCTGAA AACTTTTTTC 701
Columbia TATTTGTCTT CTGCTTGCTT ACAAGATCAA ATACCCTGAA AACTTTTTTC 950
W1-14 Consensus 1 TATTTGTCTT CTGCTTGCTT ACAAGATCAA ATACCCTGAA AACTTTTTTC 257
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741
960 980 1,000
F3 Parent sequence 1 TCTTAAGAGG AAACCATGAG TGTGCATCTA TTAACAGAAT TTACGGATTC 226
F3 Parent sequence 2 ---------- ---------- ---------- ---------- ---------- 684
Landsberg TCTTAAGAGG AAACCATGAG TGTGCATCTA TTAACAGAAT TTACGGATTC 751
Columbia TCTTAAGAGG AAACCATGAG TGTGCATCTA TTAACAGAAT TTACGGATTC 1000
W1-14 Consensus 1 TCTTAAGAGG AAACCATGAG TGTGCATCTA TTAACAGAAT TTACGGATTC 307
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741
1,020 1,040
F3 Parent sequence 1 TATGACGAAT GTAAACGAAG ATTCAACGTG AAGCTCTGGA AAGTGTTTAC 276
F3 Parent sequence 2 ---------- ---------- ---------- ---------- ---------- 684
Landsberg TATGACGAAT GTAAACGAAG ATTCAACGTG AAGCTCTGGA AAGTGTTTAC 801
Columbia TATGACGAAT GTAAACGAAG ATTCAACGTG AAGCTCTGGA AAGTGTTTAC 1050
W1-14 Consensus 1 TATGACGAAT GTAAACGAAG ATTCAACGTG AAGCTCTGGA AAGTGTTTAC 357
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741
1,060 1,080 1,100
F3 Parent sequence 1 CGACACTTTT AACTGTCTCC CTGTGGCTGC TGTCATAGAC GAAAAGATAC 326
F3 Parent sequence 2 ---------- ---------- ---------- ---------- ---------- 684
Landsberg CGACACTTTT AACTGTCTCC CTGTGGCTGC TGTCATAGAC GAAAAGATAC 851
Columbia CGATACTTTT AACTGTCTCC CTGTGGCTGC TGTCATAGAT GAAAAGATAC 1100
W1-14 Consensus 1 CGATACTTTT AACTGTCTCC CTGTGGCTGC TGTCATAGAT GAAAAGATAC 407
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741
1,120 1,140
F3 Parent sequence 1 TTTGTATGCA CGGTGGACTA TCTCCTGAGT TGATCAATGT GGAACAGATT 376
F3 Parent sequence 2 ---------- ---------- ---------- ---------- ---------- 684
Landsberg TTTGTATGCA CGGTGGACTA TCTCCTGAGT TGATCAATGT GGAACAGATT 901
Columbia TTTGTATGCA TGGTGGACTA TCTCCTGAGT TGATCAATGT GGAACAAATT 1150
W1-14 Consensus 1 TTTGTATGCA TGGTGGACTA TCTCCTGAGT TGATCAATGT GGAACAAATT 457
W1-14 Consensus 2 ---------- ---------- ---------- ---------- ---------- 741


1,160
1,180
1,200
F3 Parent sequence 1 AAGAACATAG AGCGTCCAAC TGATGTTCCA GACGCTGGTT TGCTCTGTGA 426
F3 Parent sequence 2 684
Landsberg AAGAACATAG AGCGTCCAAC TGATGTTCCA GACGCTGGTT TGCTCTGTGA 951
Columbia AAGAACATAG AGCGTCCAAC TGATGTTCCA GACGCTGGTT TGCTCTGTGA 1200
W1-14 Consensus 1 AAGAACATAG AGCGTCCAAC TGATGTTCCA GACGCTGGTT TGCTCTGTGA 507
W1-14 Consensus 2 741
1,220
1,240
F3 Parent sequence 1 CCTTCTTTGG TCTGATCCTA GTAAAGATGT CAAAGGCTGG GGGATGAATG 476
F3 Parent sequence 2 684
Landsberg CCTTCTTTGG TCTGATCCTA GTAAAGATGT CAAAGGCTGG GGGATGAATG 1001
Columbia CCTTCTTTGG TCTGATCCTA GTAAAGATGT CAAAGGCTGG GGGATGAATG 1250
W1-14 Consensus 1 CCTTCTTTGG TCTGATCCTA GTAAAGATGT CAAAGGCTGG GGGATGAATG 557
W1-14 Consensus 2 741
1,260
1,280
1,300
F3 Parent sequence 1 ATCGTGGTGT CTCCTACACT TTTGGTGCTG ACAAAGTTGC TGAGTTTTTA 526
F3 Parent sequence 2 684
Landsberg ATCGTGGTGT CTCCTACACT TTTGGTGCTG ACAAAGTTGC TGAGTTTTTA 1051
Columbia ATCGTGGTGT CTCCTACACT TTTGGTGCTG ACAAAGTTGC TGAGTTTTTA 1300
W1-14 Consensus 1 ATCGTGGTGT CTCCTACACT TTTGGTGCTG ACAAAGTTGC TGAGTTTTTA 607
W1-14 Consensus 2 741
1,320
1,340
F3 Parent sequence 1 ATAAAGAATG ATATGGATTT AGTCTGTCGT GCCCACCAGG TTAGTACATT 576
F3 Parent sequence 2 684
Landsberg ATAAAGAATG ATATGGATTT AGTCTGTCGT GCCCACCAGG TTAGTACATT 1101
Columbia ATAAAGAATG ATATGGATTT AGTCTGTCGT GCCCACCAGG TTAGTACATT 1350
W1-14 Consensus 1 ATAAAGAATG ATATGGATTT AGTCTGTCGT GCCCACCAGG TTAGTACATT 657
W1-14 Consensus 2 741
1,360
1,380
F3 Parent sequence 1 GTATAAATGT TTCCTCTAGT TTTTTTGTTT- GGGAAATCGA AT------- 617
F3 Parent sequence 2 684
Landsberg GTATAAATGT TTCCTCTAGT TTTTTTGTTT GGGAAATCGA ATCTGCTGC 1150
Columbia GTATAAATGT TTCCTCTAGT TTTTTTGTTT GGGAAATCGA ATCTGCTGC 1399
W1-14 Consensus 1 GTATAAATGT TTCCTCTAGT TTTTTTGTTT GGGAAATCNA ATCTGCTGC 706
W1-14 Consensus 2 741

Appendix G

Sequencing: nucleic acid symbols

These Rules are as close as possible to the published version [see *Biochem. J.*, 1985, **229**, 281-286; *Eur. J. Biochem.*, 1985, **150**, 1-5; *J. Biol. Chem.*, 1986, **261**, 13-17; *Mol. Biol. Evol.*, 1986, **3**, 99-108; *Nucl. Acids Res.*, 1985, **13**, 3021-3030; *Proc. Nat. Acad. Sci. (U. S.)*, 1986, **83**, 4-8; and in *Biochemical Nomenclature and Related Documents*, 2nd edition, Portland Press, 1992, pp 122-126.

Table 18: Nucleic acid symbols

G	Guanine
A	Adenine
T	Thymine
C	Cytosine
R	Purine (adenine or guanine)
Y	Purimidine (thymine or cytosine)
W	Adenine or thymine
S	Guanine or cytosine
M	Adenine or cytosine
K	Guanine or thymine
H	Adenine or thymine or cytosine
B	Guanine or cytosine or thymine
V	Guanine or adenine or cytosine
D	Guanine or adenine or thymine
N	Guanine or adenine or thymine or cytosine