5-2004

Isolation and Characterization of T-DNA Insertional Mutations in Arabidopsis thaliana and Its Use in Determining SIAMESE

Taylor S. Gwin

Isolation and Characterization of T-DNA Insertional Mutations in *Arabidopsis thaliana* and Its Use in Determining *SIAMESE*

by

Taylor S. Gwin

Undergraduate honors thesis under the direction of

Dr. John C. Larkin

Department of Biological Sciences

Submitted to the LSU Honors College in partial fulfillment of the Upper Division Honors Program.

May 2004

Louisiana State University

INTRODUCTION:

Arabidopsis thaliana is a small, dicotyledonous plant in the mustard family (Brassicaceae) that is endogenous to most temperate regions of the world (see Figure 1). Because of its size, rapid growth rate, and small genome (~125 mb on 5 chromosomes), *Arabidopsis* has become a standard for plant biology. Sequencing of the genome for the Columbia (Col) ecotype was finished in 2000, although over 750 ecotypes exist worldwide.

Trichomes (from the Greek *trichos*, meaning hair) are leaf hairs that arise on most plant epidermises. In *Arabidopsis*, wild-type trichomes develop from a single epidermal cell into much larger cell with a vertical stalk and two to four branches. Wild-type trichomes are found on the adaxial surface of leaves, sepals, petioles, and stems (see Figure 1).

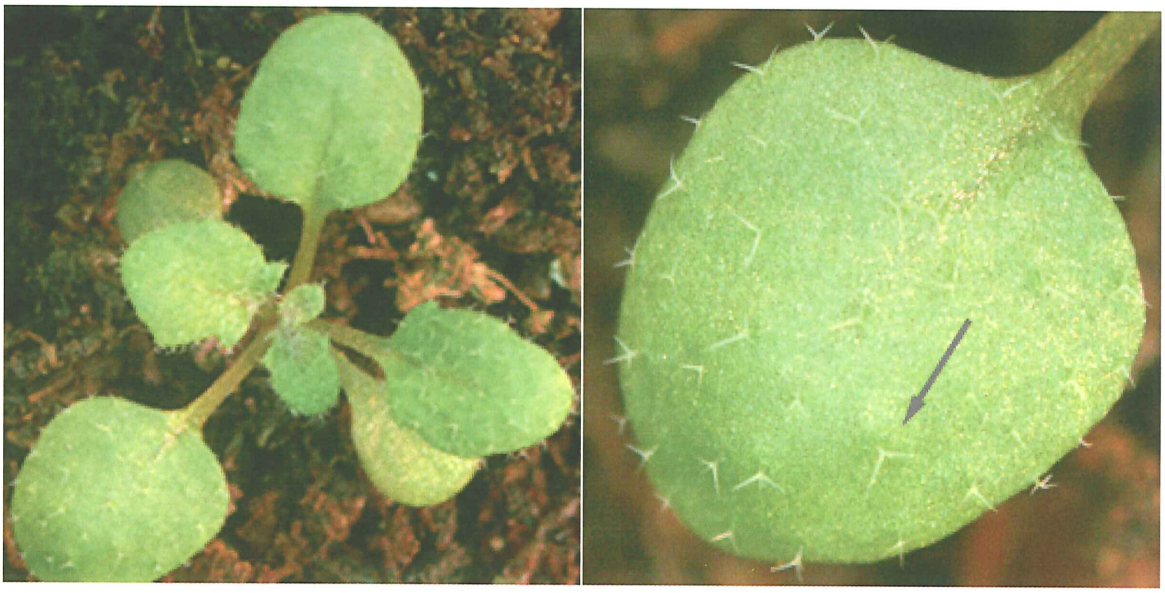

Figure 1. *Arabidopsis thaliana* and trichomes. This figure is of a plant approximately 2 weeks after germination. The arrow shows an example of a trichome.

All cells must decide what mature cell type they will become. Some cells, such as trichomes, become highly specialized compared to other plant cells, and thus are excellent models for cell differentiation. Moreover, in *Arabidopsis*, trichomes are not essential for plant viability, and mutations or absence of trichomes does not affect the plant. A great deal of research

has been conducted on the development of trichomes and their relation to other epidermal cells (see Larkin *et.al.*, 2003). Of particular interest is the spacing of trichomes as an example of developmental patterning and the regulation cell cycle in trichomes. The spacing of wild-type trichomes is regulated by lateral inhibition by a developing trichome. According to Larkin *et.al.* (1996), if trichomes were randomly developing on the leaf epidermis, in their study the probability of having two trichomes adjacent to each other was approximately 21%. However, no trichomes were found adjacent to each other indicating that a patterning mechanism must exist.

In a normal mitotic cell, the cell cycle replicates DNA and cellular materials once then undergoes mitosis to divide into daughter cells. Once a cell begins to develop as a trichome, however, the cell cycle switches to a modified version called endoreplication, in which mitosis is omitted and the genome is repeatedly replicated (see Figure 2). Thus wild-type trichomes have a nuclear DNA content, on average, of 32C but remain uni-cellular.

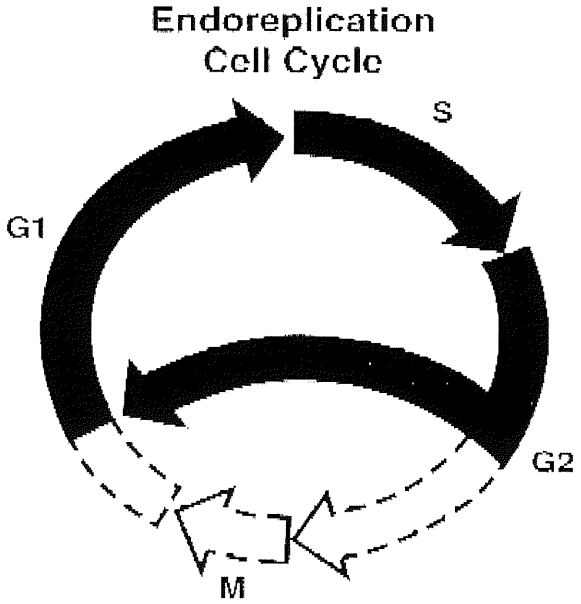

Figure 2. The cell cycle and endoreplication. When the cell cycle switches to endoreplication, mitosis is skipped, and G2 leads directly to G1. This results in a build-up of DNA in the cell and larger cell size. (figure courtesy of Walker *et.al.*, 2000)

Recessive mutations in the gene *SIAMESE* (*SIM*) have been shown to cause an increased number of adjacent trichomes, giving a characteristic "twin" phenotype, multiple cells per trichomes, and a decrease in the amount of endoreplication (Walker *et.al.*, 2000). They showed that in *sim* plants, the number of trichomes per leaf increases while the number of trichome initiation sites is unchanged. Furthermore, they showed that the DNA content reduces dramatically to approximately 8C in *sim* mutants. This data supports the idea that mutations in *SIAMESE* cause cell division after differentiation has already been initiated, resulting in multi-celled and/or twin trichomes (see Figure 3).

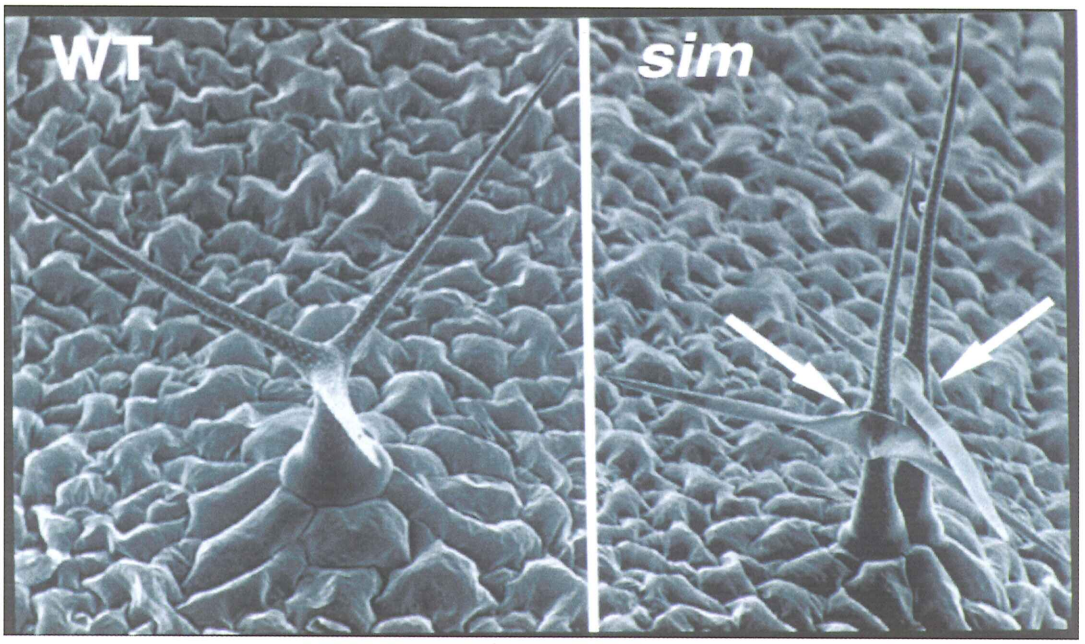

Figure 3. SEM images of wild-type and *siamese* trichomes. WT trichomes are single-celled and separated from each other. *Siamese* trichomes are multi-cellular and often found in pairs. The arrows show cell junctions in *siamese* trichomes. (figure courtesy of Walker *et.al.*, 2000).

Transferred-DNA (T-DNA) insertional mutagenesis is a process that allows random mutations to be induced in an organism's genome by the insertion of foreign DNA. The *sim*-2 plant line from Martin Hülskamp is a family of Wassileskjia (WS) background that had undergone T-DNA insertional mutagenesis and showed a *siamese*-like phenotype. This mutagenesis was

accomplished by treating the plants with *Agrobacterium tumefaciens* that contained the pD991 T-DNA and using vacuum filtration to transform the plants. Transformation with pD991 not only causes random mutation, but, due to the *β-glucuronidase* (GUS) reporter gene contained within the T-DNA, it also creates an enhancer trap. If the T-DNA inserts near a gene's regulatory region (enhancer), the GUS gene will be expressed and it is possible to visualize a putative expression pattern for whatever gene the enhancer is naturally associated with. The *sim*-2 line, as mentioned before, showed a *siamese*-like phenotype and thus was studied in hopes of finding *SIM*. A second insert was present as well, and because initial analysis (Jason Walker, unpublished observations) did not appear to show an insert on the same chromosome as *SIM*, interest shifted to the second insert, at that time thought to be in the *Calreticulin 1* gene, and a possible mutant floral phenotype.

Calreticulin (*CRT*) is a calcium binding protein found in the endoplasmic reticulum primarily used for the storage of calcium and as a molecular chaperone. *CRT* is found in all tissues of *Arabidopsis*, but it is found in especially high quantities in seeds and floral tissues. No role has yet been determined for the higher quantity, although it has been proposed that *CRT* may have a function in the synthesis of enzymes required for the degradation of sucrose for nectar. (Nelson *et.al.*, 1997) There are three *CRT* genes in *Arabidopsis*, with *CRT1* being the focus of this study. No visible phenotype had been determined for mutations in *CRT1*, and we wished to investigate whether there was a floral mutant phenotype associated with the putative *CRT* insertion.

This study analyzed the segregation of T-DNA insertions within various plant lines derived from the initial *sim*-2 line to determine what, if any, mutations are caused by the presence of the T-DNA, and locate the insertions within the *Arabidopsis* genome. Ultimately, I was able to show that, in

contrast to Jason Walker's initial work, that this T-DNA line contained an insertion tightly linked to the *sim*-2 trichome phenotype. These data and the establishment of a family homozygous for the single insert linked to *sim* directly led to the finding of the insert and the discovery of the molecular identity of *SIM*.

MATERIALS AND METHODS:

pD991. The insert, pD991, used in this study was created in the laboratory of Thomas Jack (Campisi *et.al.* 1999). pD991 contains a kanamycin resistance gene (bacterial *nptII*) under the control of the 5′ mannopine synthase promoter and a ***β-glucuronidase*** gene (bacterial *uidA*) with a minimal promoter (truncated 35S cauliflower mosaic virus promoter) (see Figure 4). Expression of kan^r in transgenic plants is always observed, but only if pD991 is inserted near a genomic enhancer will expression of GUS be observed. This allows the GUS activity to serve as an indicator for the tissues in which the gene in question may function. (Campisi *et.al.* 1999)

DNA extraction. DNA extraction was performed by grinding the tissue and treating with either Edward's extraction buffer (0.2M Tris pH 7.5, 0.25M NaCl, 25mM EDTA, 0.05% SDS) or Edward's and proteinase K. Purification of DNA was performed using phenol:chloroform:isoamyl alcohol (25:24:1). DNA was precipitated with 95% ethanol and 3M NaOAc pH 5.2, resuspended in dH_2O, and stored at -20° C.

TAIL-PCR. TAIL (thermal asymmetric interlaced) PCR was performed as described by Liu *et.al.* (1995). Anonymous degenerate (AD) primers used were (5′-3′) NTCGASTWTSGWGTT (AD1), NGTCGASWGANAWGAA (AD2), and WGTGNAGWANCANAGA (AD3). Specific primers designed from the right and left pD991 border sequences were GCATGCAAGCTTGGCACTGG (Oligo123), TGAGACCTCAATTGCGAGC (124), TGATCCATGTAGATTTCCCG (154), ATAACGCTGCGGACATCTAC (155), CCTATAAATACGGATCG (156), and TCGGGCCTAACTTTTGGTG (86). Primary, secondary, and tertiary

reactions were performed to amplify nested fragments. Since each new product should be slightly smaller than the previous product, if all three are visualized using agarose gel electrophoresis, a characteristic "step-down" banding pattern should be observed. For each nested reaction, product from the prior reaction is diluted 1:50 and used as the template.

Adapter-PCR. Purified DNA is first digested with one of *Dra* I, *EcoR* V, *Pvu* II, and *Stu* I, which do not cut in pD991. Oligos of different lengths are mixed together, heated to 94° C, and allowed to cool gradually. This process generates the double stranded adapter. The adapter is then ligated to the ends of the digested DNA. Reactions were performed using the nested primers designed from the adapter DNA, GTAATACGACTCACTATAGGGC (AP1) and ACTATAGGGCACGCGTGGT (AP2) and a primer designed from the border sequence of pD991, CACCAAAAGTTAGGCCCGA (Oligo86inv). For the secondary reaction, product from the primary reaction is diluted 1:50 and used as the template.

DNA sequencing. Reactions were performed using the protocol for PE Applied Biosystems' BigDye™ and were sent to Pennington Biomedical Research Center for sequencing.

Planting. Plants were grown on either soil or on agar plates containing Murashige and Skoog (MS) salts. All plants were kept in an environmental growth chamber at a constant 22° C and constant illumination. Soil was layered with one inch of vermiculite and one inch of sterile soil. Plants were watered with 1/3 dH_2O, 1/3 nutrient solution (10mM KNO_3, 5mM KH_2PO_4 pH 5.6, 4mM $MgSO_4$, 4mM $Ca(NO_3)_2$, 0.01% Sequestrene, 70μM H_3BO_3, 14μM $MnCl_2$, 0.5μM $CuSO_4$, 1μM $ZnSO_4$, 0.2μM $NaMoO_4$, 10μM NaCl, 0.01μM $CoCl_2$), and 1/3 BT (0.6% *Bacillus thuringiensis* subs. *israelensis*). Some plates contained, in addition to MS salts, kanamycin (1mg/L) or sucrose (1% or 4% by mass). Plates were kept either horizontally for screening or vertically for root analysis.

GUS staining. The presence of the pD991 insert in various plant tissues was established by testing for activity of the *β-glucuronidase* gene (GUS) using the substrate 5-bromo-4-chloro-3-indole-β-D-glucuronide (X-Gluc). Initially plants were incubated for 15 minutes in a pre-fix solution (0.1M $NaHPO_4$ pH 7, 0.1% formaldehyde, 0.1% Triton X-100 in dH_2O). Plants were then washed with 0.05M $NaHPO_4$ pH 7 and incubated overnight at 37° C in an X-Gluc solution (0.05M $NaHPO_4$ pH 7, 0.1% Triton X-100, 1mM $K_3Fe(CN)_6$, 1mM X-Gluc (in DMF) in dH_2O). Plants were then incubated for 30 minutes in FAA (18:1:1 EtOH:HOAc:Formalin) and decolored using an ethanol series. Staining patterns were visualized under a dissection microscope. Photographs were taken using a SPOT INSIGHT camera attached to the microscope.

***CRT1* amplification.** Primers used for the amplification of *Calreticulin 1* were designed from the f14g9 BAC clone and are as follows: f14g9-15585l GCAATGGCGAAACTAAACCC (I), 16665l CCTCCGTCCAGATGCTACTTA (II), 17399l GTTGGAGTTGAGTTGTGGCAG (III), 17954l GATGATGACAATGAAGGAGATGAC (IV), 16686r GTAAGTAGCATCTGGACGGAGG (V), 17419r CTGCCACAACTCAACTCCAAC (VI), 18129r CACAGAGTTAGTCATACAGCCC (VII), and 19996r CACACAAACACAGACAAAAAA (VIII). Primers were also used from the pD991 borders as follows: Oligo-38rb TAACTTTTGGTGTGATGATG, 98rb CACCTGAATGGCGAATGAG, 73lb GACCATCATACTCATTGCT, and 138lb CTGCGGACATCTACATTTT.

t19n18/t32m21 amplification. Primers used for the amplification of the t19n18 and 132m21 BAC clones were designed approximately 2kb apart in an overlapping fashion. There were 87 primer sets designed to cover the 160 kb region.

Planting numbers. Families are named with respect to year and chronological order, but in this paper all families are simply numbered for clarity as follows:

Number	Reference Name	Number	Reference Name	Number	Reference Name
1	01-545-1	10	01-545-29	19	01-545-67
2	01-545-2	11	01-545-37	20	01-545-68
3	01-545-4	12	01-545-39	21	01-545-69
4	01-545-5	13	01-545-40	22	01-545-70
5	01-545-7	14	01-545-42	23	01-545-71
6	01-545-9	15	01-545-43	24	01-545-72
7	01-545-10	16	01-545-47		
8	01-545-12	17	01-545-54	25	01-545-68-1
9	01-545-27	18	01-545-55		

ORIGINAL LINE:

Martin Hülskamp isolated a plant line resulting from T-DNA insertional mutagenesis that showed a *siamese*-like mutant phenotype. This family became the original *sim*-2 plant line (in a WS background), and Jason Walker performed the initial analysis. Jason found that there were at least two inserts present in the family (Jason Walker, unpublished observations), and due to the *siamese*-like phenotype, an insert was thought to be in the *SIM* gene. TAIL-PCR, if successful, yields fragments of decreasing sizes due to the nested primers. So if the products from each of the TAIL reactions are run adjacently on an agarose gel, the bands seem to "step down" in size (see also Materials and Methods). Results showed the step-down pattern suggesting the presence of an insert. Jason's sequencing of the tertiary PCR product yielded sequence from the terminal exon of *CRT1* (Jason Walker, unpublished observations); however, no sequence from the T-DNA insert itself was detected in the sequence, and thus the presence of a T-DNA insertion in *CRT1* had not been fully confirmed. A second putative insertion on chromosome 3 was also detected, and both predicted insertions were thought to have been found (Jason Walker, unpublished observations). In addition to the *siamese*-like phenotype, Jason also saw a mutant floral phenotype in a single experiment and hypothesized that this phenotype might be due to an insertion in *CRT1* (Jason Walker, unpublished observations).

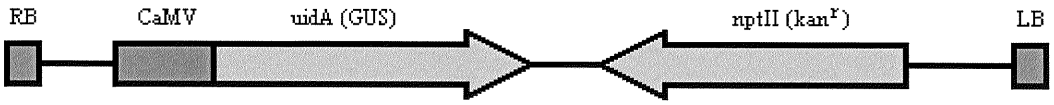

Figure 4. Structural map of pD991. The GUS reporter gene with its truncated 35S CaMV promoter, the kan^r gene, and the left and right borders are noted.

Jason Walker crossed *sim*-2 to plants of the Col ecotype yielding the family 01-545, which was allowed to self-fertilize and showed a segregating *siamese*-like phenotype. The goal of this experiment was to identify families that segregated each insertion individually, as indicated by 3:1 kan^r:kan^s segregation, to allow any mutant phenotypes to be associated with a single specific insertion. Thus the individual F2 plants were allowed to self-fertilize and 24 F3 families were planted on agar plates containing kanamycin to score both mutant phenotypes and kanamycin resistance. Since the original *sim*-2 plant line was expected to have two inserts, the F2 population should have segregated the two inserts in a 9:3:3:1 ratio. Thus, 9/16 of the F2 families should contain both inserts, with both inserts either heterozygous or homozygous. 3/16 of the families should contain the first insert only, and 3/16 should contain the second insert only, with either insert as heterozygous or homozygous. 1/16 of the families should contain neither insert, and should thus be fully sensitive to kanamycin. If either insert is homozygous, the family will be fully resistant to kanamycin, regardless of the presence or absence of a second insert. Since our objective was to separate the two inserts, all families that showed full kanamycin resistance were ignored, as it is typically difficult to determine whether two different inserts are present without additional self-fertilization. If a family segregates 3:1 kan^r:kan^s, it is expected to be heterozygous for a single insertion, and any additional mutant phenotype should segregate 3:1 WT:mutant in the total population. Furthermore, since we expected any mutant phenotype to be linked to an insertion, the mutant phenotype should segregate 2:1

WT:mutant among the kanamycin resistant plants, because all mutant plants would contain the kan^r insertion. If the mutant phenotype segregates 3:1 WT:mutant in the kanamycin resistant plants, it is most likely that the mutant phenotype and the insertion are not linked.

Of the 24 F2 families, one was entirely sensitive to kanamycin, revealing the absence of any insertions. 11 families were entirely resistant to kanamycin, showing that they were at least homozygous for one insert and may or may not have contained a second insert. The other 12 families showed segregating kanamycin resistance, which corresponds to one or two different heterozygous insertions (see Table 1).

Two families, 5 and 16, initially showed segregation of kanamycin resistance in a 3:1 kan^r:kan^s ratio which is consistent with one heterozygous insertion but did not segregate the *sim*-like phenotype (see Table 1). If a mutation in *SIM* was present, we would expect the plants to segregate 3:1 WT:*sim* in the total population, and if the *sim*-like phenotype was linked to the insertion, we would expect the plants to segregate 2:1 WT:*sim*, as explained above. Since no *sim* phenotype segregated at all, it is most likely that the insert present in families 5 and 16 is not linked to *SIM*.

Two other families, 13 and 20, also showed segregation that is consistent with one heterozygous insertion, but these families showed segregation of the *sim*-like phenotype, as well (see Table 1). The *sim*-like phenotype segregated in a 2:1 WT:*sim* ratio in the kan^r plants, which suggests linkage between the insert and the *sim*-like phenotype (see Table 3). If the kan^r and the *sim*-like phenotype were two unlinked traits segregating in the populations, one would expect a 3:1 WT:*sim* ratio in the kan^r plants, as explained above. Thus we were ultimately able to show that the *sim*-like phenotype was, in fact, linked to this T-DNA insertion.

Table 1. Segregation of Kanamycin resistance in F3 families.

Family Number	Kanamycin Segregation		Segregating *siamese*?
	Sensitive	Resistant	
1	15		No
2	4	24	Yes
3		27	Yes
4	1	27	Yes
5	3	10	No
6	3	16	Yes
7		21	Yes
8		45	Yes
9		10	Yes
10	2	15	Yes
11		17	No
12		23	Yes
13	2	13	Yes
14		10	Yes
15		20	Yes
16	2	3	No
17	5	51	Yes
18		53	Yes
19		22	No
20	15	26	Yes
21		22	Yes
22	1	21	Yes
23		22	Yes
24		28	Yes

Shown are the numbers of plants that were either kanamycin sensitive or resistant and whether or not the family showed segregation of *sim*. The boxed families are those that were expected to be heterozygous for a single insertion and were tested further.

***CALRETICULIN 1*:**

After Jason Walker's TAIL results showing insertions in *CRT1* and chromosome 3, it was initially thought that no insert was present in *SIM*, and the *siamese*-like phenotype was a random mutation independent of the T-DNA insertional mutagenesis. For this reason, we turned our attention to confirming that there was an insertion in the *CRT1* gene and whether there was any linkage between the insert and the putative floral phenotype. Due to the

low number of plants in the first screening, additional screenings were performed on families 5 and 16, those expected to be heterozygous for the single insertion not linked to *sim*. In the additional test, family 16 failed to segregate 3:1 kan^r:kan^s ($P < 0.005$), but instead segregated much closer to 15:1 ($P > 0.1$), the expected ratio for two heterozygous insertions. Furthermore, family 16 segregated *sim* in this second test, verifying that two insertions were present. The 3:1 kan^r:kan^s ratio was maintained in family 5 ($P > 0.1$), verifying that the family only contains the single insertion ($P < 0.005$ for 15:1 segregation) (see Table 2).

Table 2. Additional test results for families 5 and 16.

Family Number	Kanamycin Segregation		P values		Segregating *siamese*?
	Resistant	Sensitive	(3:1 ratio)	(15:1 ratio)	
5	74	20	> 0.1	< 0.005	No
16	64	6	< 0.005	> 0.1	Yes

Shown are the number of plants that were kanamycin sensitive or resistant, the P values for the possible segregation ratios, and whether or not the family segregated *sim*.

The mutant floral phenotype should have segregated 2:1 WT:phenotype in family 5 (assuming linkage), however, no mutant floral phenotype was ever witnessed, nor was segregation of the phenotype in the original *sim*-2 population ever observed again after Jason Walker's initial report, and thus it was not pursued further. GUS staining, which shows the expression of the *β-glucuronidase* gene (GUS) present in the T-DNA insert, of family 5 showed GUS expression in the stipules, roots, and flower buds, with no staining in the trichomes, which is consistent with this insert not being related to *SIM* (see Figure 5).

Figure 5. GUS expression in plants heterozygous for a single insert not linked to *SIM*. These images are of plants from family 5. Note the staining in flowers, stipules, and roots.

My interest now shifted to confirming that there was a T-DNA insertion in *CRT1*, as indicated by Jason Walker's data. TAIL-PCR was performed again on the initial *sim*-2 line, but results were not found to be consistent with those previously seen. Under the same conditions, no step-down amplification was seen. Primers were then designed that span the *CRT1* gene and amplification of these primers was attempted to show that no insert was present. Primer pairs for the gene amplification were as follows: I with V, II with VI, III with VII, IV with VII, and IV with VIII (see Figure 6). Also, PCR reactions with all possible pairs of the spanning primers and the border sequence primers from the pD991 insert were attempted to see if any amplification occurred. From Jason Walker's results, we expected the reactions with primers III with VII and IV with VII to fail, suggesting an insertion in that region. However, all spanning primer sets amplified across the gene, yielding fragments of expected lengths, and no border sequence primers amplified consistently with spanning primers (with one exception, as noted below). If an insert had been present, at least one set of spanning primers would not have amplified (as the insert is ~2kb in length), or the resulting fragment would have been at least 2kb larger than expected. Also, consistent amplification between at least one border primer and a spanning

primer should have occurred, and this could have been sequenced to determine the insertion site.

As mentioned before, Jason Walker's TAIL results did not yield the exact location of the insert, only sequence from the 3′ end of *CRT1* (see Figure 6). Interestingly, the only consistent amplification with a border primer from pD991 was with 179541, a primer near the terminal exon of *CRT1*. This amplification yielded a PCR fragment with a size that would indicate that the insertion was approximately 1.5kb downstream of the gene, however. Since it is possible to have a downstream transcriptional enhancer, and since disruption of this could cause a mutant phenotype, additional primers were designed spanning the downstream region as well (179541 with 19996r). However, this primer set amplified a fragment of expected size from genomic DNA, suggesting that there was no insertion. Since all reactions amplified fragments of identical length in both *sim*-2 and WS, and each fragment length was the expected size from genomic DNA, it was concluded that no insert existed in or around the *CRT1* gene.

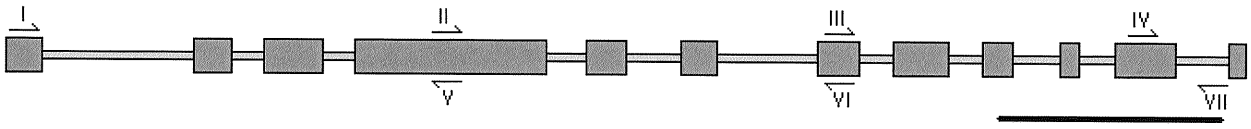

Figure 6. Schematic diagram of *Calreticulin 1* (5′ to 3′) showing primers and sequenced TAIL results. Exons are shown as large dark boxes; introns are shown as small light boxes. Primers are drawn as an arrow above (left primers) or below (right primers) the sequence along with numerals corresponding to their names. I is 155851, II is 166651, III is 173991, IV is 179541, V is 16686r, VI is 17419r, and VII is 18129r. The region amplified by Jason Walker in his TAIL reactions is in underlined (Jason Walker, unpublished observations). Primer F14G9-19996r (VIII) is not shown as it is ~1.7 kb downstream of *CRT1*.

After the failure of TAIL-PCR to amplify any insert, it was thought that the insert might be present multiple times in succession with possible concatamers. Since an insert was still present in the families, Adapter-PCR was attempted (with help from Michelle Speckhart) to discover its location. The primary Adapter reaction showed amplification of fragments in all four of the enzyme digestions that were not present in the controls. However, the

secondary reaction, which was run with the secondary Adapter primer, AP2, and the same insert border primer, showed no amplification in any samples or controls. Dilutions of 1:25 and 1:12 also showed no amplification in any samples or controls (see also Materials and Methods). Adapter-PCR was not attempted again with other insert border primers, but since the primary reaction worked in all enzyme digestions yielding fragments of different lengths, errant primer binding was not considered a problem. Further attempts at a successful secondary reaction have not yet been performed. However, since the mutant floral phenotype has not been able to be reproduced, and no other mutant phenotype is visible, it is not likely that further analysis will be attempted.

***SIAMESE*:**

As mentioned before, after Jason Walker's initial TAIL-PCR results, linkage between a T-DNA insert and the *siamese*-like phenotype was far from certain. However, the families 13 and 20, those expected to be heterozygous for a single insertion linked to *sim*, were studied to verify whether or not linkage did exist. Again, the original analysis of both families yielded only a few plants that germinated, so additional tests were performed. Family 13 was again grown on MS agar plates containing kanamycin, and again this family segregated 3:1 kan^r:kan^s ($P > 0.1$, opposed to $P < 0.005$ for the ratio of 15:1 for two inserts) and 2:1 WT:*sim* ($P > 0.9$, opposed to $P < 0.5$ for the ratio of 3:1 for non linkage) (see Table 3). This verifies that family 13 is heterozygous for a single T-DNA insert, and it suggests that this insert is linked to the *sim* phenotype.

Family 20 was grown on regular MS agar plates so that GUS staining patterns could be established without interference from kanamycin segregation. All plants that germinated were GUS stained, and trichome phenotypes were established afterwards. The expected ratio of plants that

stained to plants that did not stain was 3:1, corresponding to the 3:1 kan^r:kan^s ratio which designates those plants with an insert. Segregation of staining plants to those that did not stain did match the ratio of 3:1 ($P > 0.5$, opposed to $P < 0.005$ for the ratio of 15:1 for two inserts), verifying that the family was heterozygous for one insertion (see Table 3). Furthermore, the fact that there is a GUS pattern at all shows that the insertion is located near an enhancer region. If the insert is linked to *sim,* the trichome phenotype ratio was expected to be 2:1 WT:*sim* among the plants that showed GUS staining, corresponding to the ratio of WT:*sim* found in the original screening and that expected for the segregation of linked traits. Again, if the two traits were unlinked, the ratio should have been 3:1 WT:*sim* among the plants that showed GUS staining. The ratio of the segregation of wild-type to *sim*-like plants again suggests linkage between the insertion and the phenotype ($P > 0.5$ for 2:1, $P < 0.1$ for 3:1) (see Table 3). A third test was then performed in which the trichome phenotype was established prior to GUS staining. We expected that all plants showing the *sim* phenotype would have GUS expression, as the mutant phenotype should be caused by a homozygous insertion in the individual plant. Thus all plants showing the *sim* phenotype were stained, and all of the plants did show the GUS pattern. If the insertion and the *sim* phenotype were not linked, this sub-population should have segregated with a ratio of 3:1 GUS:WT, but if there was linkage between the two traits, all plants with the *sim* phenotype should have shown the GUS pattern. Results verify linkage between the insert and the phenotype ($P > 0.995$ for all GUS, $P < 0.005$ for 3:1 GUS:WT) (see Table 3). GUS staining showed expression of the GUS gene in the trichomes, as expected, as well as in the roots and vasculature (see Figure 7).

Plants showing the *siamese* phenotype from F3 family 20 should be homozygous for the single insert, so self-fertilization should yield a family homozygous for the insert. In order to establish a homozygous family, plants

with the *sim* phenotype were allowed to self-fertilize, and the F4 generation was grown on kanamycin plates. As expected, one F4 family, 25, showed all kanamycin resistance and all plants had the *sim* phenotype. Thus, family 25 is homozygous for the *sim* linked insert.

Table 3. Additional testing of families heterozygous for *sim* linked insert.

Family Number	Kanamycin Segregation		P values		*siamese* Segregation		P values	
	resistant	sensitive	(3:1 ratio)	(15:1 ratio)	WT	*sim*	(2:1 ratio)	(3:1 ratio)
13	57	24	> 0.1	< 0.005	27	13	> 0.9	< 0.5
	GUS Segregation		P values		*siamese* Segregation		P values	
	GUS	WT	(3:1 ratio)	(15:1 ratio)	WT	*sim*	(2:1 ratio)	(3:1 ratio)
20	156	49	> 0.5	< 0.005	49	107	> 0.5	< 0.1
	GUS Segregation in *sim*		P values					
	GUS	WT	(1:0 ratio)	(3:1 ratio)				
20	34	0	> 0.995	< 0.005				

Family 13 was grown on MS agar with kanamycin; Family 20 was grown on MS agar without any kanamycin. In the first screening for family 20, *sim* segregation was determined after GUS staining. In the second screening of family 20, GUS staining was performed only on plants showing the *sim* phenotype.

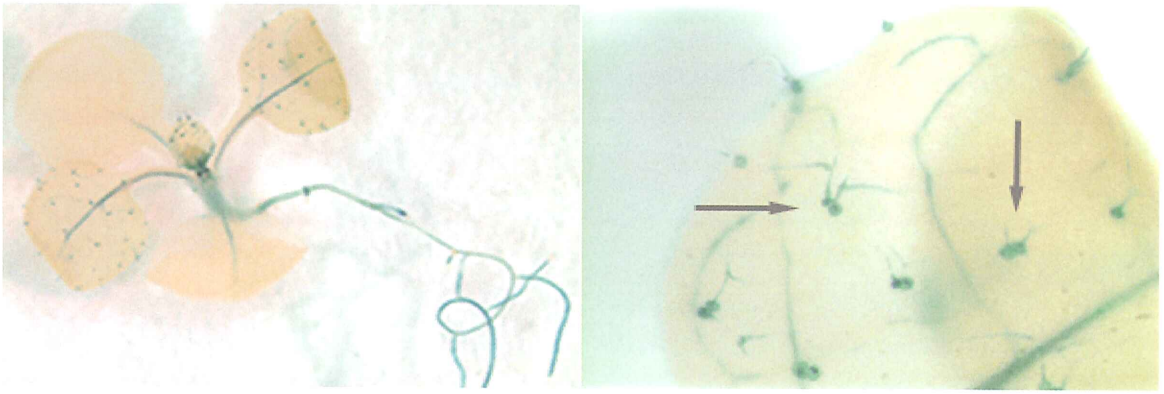

Figure 7. GUS expression pattern in families segregating the *sim* phenotype. These pictures are of a plant showing the *sim* phenotype of family 13. Note the staining in roots and trichomes. The arrows show pairs of trichomes characteristic of the *sim* phenotype.

Initial mapping of the *SIM* gene using the *sim*-1 allele performed by Jason Walker and Jason Churchman showed that the gene was located on chromosome 5, near the sequence corresponding to BAC clones t19n18 and

t32m21. Efforts to pinpoint the gene had not proven successful, however. So along with Alex Hellmann and Jason Churchman, the mapping was repeated, and the previous mapping results were confirmed (see Table 4).

Table 4. Recombinant data collected by Jason Walker and Alex Hellman.

	Number of Recombinants per BAC Clone					
	MED24	F8F6	T32M21	T1E3	MUK11	MUK11
J. Walker	13				11	
A. Hellman	4	2	0	0	3	4

Initial PCR mapping by Jason Walker showed that *SIM* was present on chromosome 5, but markers were only available for MED24 and MUK11, so the gene was not pinpointed further (Jason Walker, unpublished observations). Additional PCR mapping by Alex Hellman utilized markers on F8F6, T32M21, T1E3, and a second marker on MUK11 as well (Alex Hellman, unpublished data). Note that, as expected from this data, *SIM* is present on T32M21.

At this point TAIL-PCR and inverse-PCR (by Michelle Speckhart) were attempted on family 25, but TAIL-PCR yielded no consistent amplification and inverse-PCR only showed that the insert probably contained multiple copies of pD991 with possible concatamers. Since other methods to determine the location of the T-DNA insert had failed, with Alex Hellmann, amplification of the genome of family 25 corresponding to t19n18 and t32m21 was attempted. As with the amplification of *CRT1*, primers were designed and the absence of amplification due to the presence of the 2kb insert between primers was tested for. All primer pairs that successfully amplified a fragment of expected size in wild-type DNA also amplified the wild-type size band in family 25. With this method, approximately 50% of the 160 kb region was discarded as a possible location for the insert. This corresponds to 11 fully amplified open reading frames (ORF's) and 12 partially amplified ORF's of the 37 ORF's on the two BAC clones. Noting the time required to successfully test 160kb of DNA by amplifying 2kb fragments, Michelle Speckhart tried Adapter-PCR on family 25, since this method can be used on

concatamers and multiple copies of an insertion. With this method, the insertion was finally found in the putative protein At5g04470 (on t32m21). Michelle Speckhart verified this result by sequencing the ORF in *sim*-1 and showing that, in fact, there is a single base pair mutation (for *sim*-1) in the start codon. The pD991 insertion occurs upstream of *SIM*, and deletes the promoter and approximately 70% of the gene itself, resulting in no expression of the gene (see Figure 8).

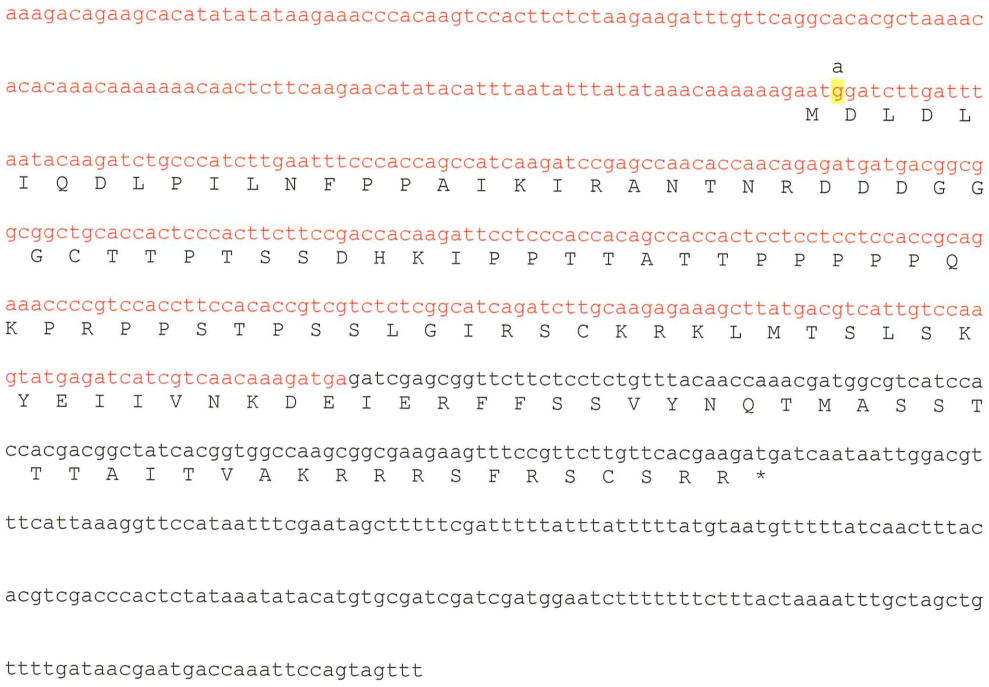

Figure 8. *SIAMESE* sequence without introns and amino acid translation with *sim*-1 and *sim*-2 mutations. The mutation causing *sim*-1 is marked yellow, with the single base pair substitution shown above. The region deleted by the T-DNA insertion causing the *sim*-2 mutation is in red.

In addition to the trichome phenotype in *sim* plants, it was expected that mutations in *SIM* would have an effect on other plant tissues, as well. Birnbaum *et.al.* (2003) have shown that *SIM* is expressed in significant amounts in roots, so we expected that *SIM* may have a role in root development. As *SIM* is known to have a role in shifting the cell cycle to endoreplication, it was proposed that a mutation in *SIM* would cause roots to fail to shift into endoreplication, and this could cause an increase in root

size and growth rates. Since there are other genes from the *SIM* family that are expressed in roots as well, this effect may only be seen if induced by the presence of sucrose, which stimulates cell division in roots and may reveal any rate limiting defects in root cell division. So families 20 and 25 as well as the ecotype WS were grown on regular MS agar plates, plates containing 1% sucrose, and plates containing 4% sucrose. However, no noticeable difference in root development was observed. The same test was repeated with *sim*-1 and Col plants grown on MS agar plates and plates containing 1% and 4% sucrose. As expected, the growth rates for roots of *sim*-1 plants grown on MS were similar to those of Col roots, whereas the root growth rates for *sim*-1 plants grown on 1% and 4% was higher than those of Col plants. The range of growth rates for *sim*-1 plants is considerable, though, yielding substantial standard deviations (see Figure 9). However the data shown is representative of a small population, and further study will be performed.

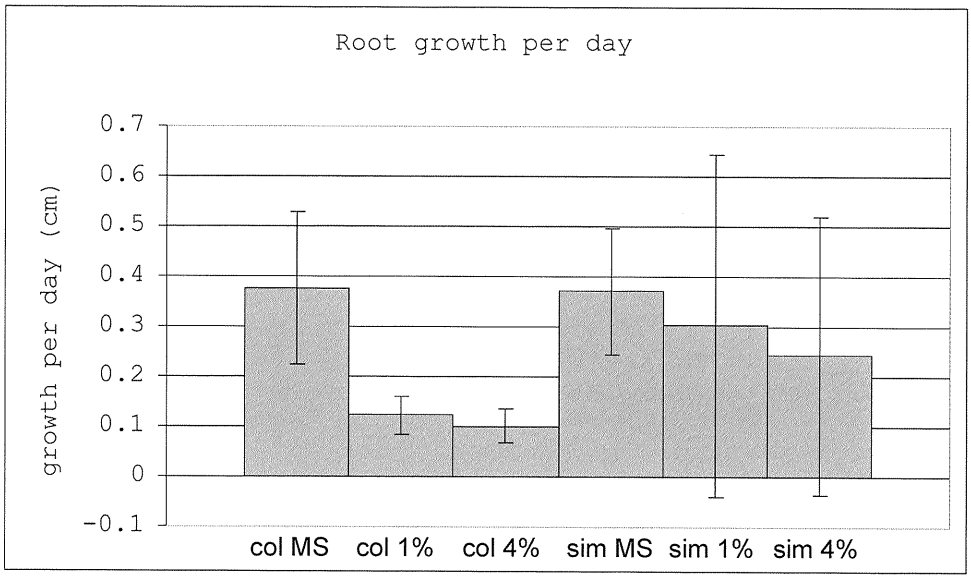

Figure 9. Average root growth per day for Col and *sim*-1 plants on MS media with different concentrations of sucrose. Sample sizes were 8 for Col and *sim*-1 on MS, 12 for Col on 1%, 13 for Col and *sim*-1 on 4%, and 15 for *sim*-1 on 1%. Standard deviations shown above are extensive for *sim*-1 grown on 1% and 4% sucrose, so further studies are underway.

CONCLUSIONS:

This research was originally started because of the *siamese*-like phenotype and the hope that the T-DNA was inserted into *SIM*. After initial setbacks with the *sim*-2 line (failure of TAIL to show a second insert outside of *CRT1*) and successes in mapping *SIM* (localization to two BAC clones), further study on *SIM* was dropped. The research towards the proposed insert in *CRT1* was an offshoot project initially designed to find some success in an apparent failure, given that the primary goal of the lab was to isolate *SIM*. Ironically, it was the original goal of using the T-DNA insert to isolate *SIM* that proved to be successful, not the offshoot project.

As mentioned above, neither insert was actually found in this study. The insert first thought to be in *CRT1* is now considered to be at an unknown location. It should be noted, however, that the GUS reporter gene can only be expressed if it is inserted near an enhancer, so it is possible that the insert is near enough to an enhancer region for the transcription of GUS but is not interfering with transcription of the gene. Perhaps more likely, the insert could still be within a gene, and the mutation simply has no visible phenotype due to functional redundancy, etc. Nonetheless, there exists no insert in *CRT1*, and still no visible mutant phenotype can be associated with the gene.

Michelle Speckhart found the *sim*-2-linked insert, as noted above, and consequently *SIM* was found as well. Several methods were simultaneously being used to find either the insert or the gene directly. Of these methods, the BAC clone amplification and the sequencing of ORF's (by Michelle Speckhart) would also have eventually yielded the gene. From Michelle Speckhart's analysis, *SIM* is thought to be a cyclin dependant kinase inhibitor (CDKi), and there are two proposed pathways for the regulation of the cell cycle by *SIAMESE*. *SIM* may have a function at either the G1 to S transition or at the G2 to M transition, binding to CDK's and deactivating

them (see Figure 10). The possibility that the *SIM* protein functions at the G2 to M transition is particularly exciting, as this would be the only known CDKi at that position.

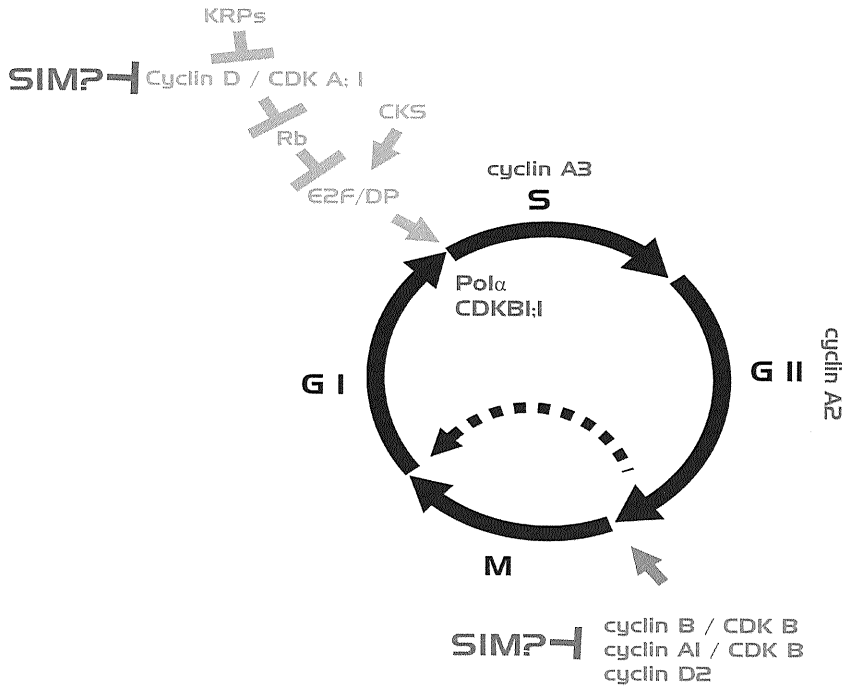

Figure 10. Possible roles of *SIM* in the cell cycle. *SIM* has two possible roles, to either inhibit the cyclins that promote the transition from G2 to M or to inhibit cyclins that indirectly inhibit the transition from G1 to S. (figure courtesy of Matthew Brown)

The expression of *SIM* in the roots of *Arabidopsis* suggests a role in root development, as well. Data has been collected that supports the proposition that mutations in *SIM* alter the growth rate in roots, but no study has yet been done to measure the DNA content of root cells in *sim* plants. Further studies on *SIAMESE*'s effect on growth rate are proposed, and additional tests are also being planned to establish the effect *SIM* specifically has on the cell cycle's shift to endoreplication in roots.

LITERATURE CITED:

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, and Benfey PN. (2003). A gene expression map of the Arabidopsis root. *Science* **302**, 1956-60.

Campisi L, Yang Y, Yi Y, Heilig E, Herman B, Cassista AJ, Allen DW, Xiang H, and Jack T. (1999). Generation of enhancer trap lines in *Arabidopsis* and characterization of expression patterns in the inflorescence. *Plant J.* **17**, 699-707.

Larkin JC, Brown ML, Schiefelbein J. (2003). How do cells know what they want to be when they grow up? Lessons from Epidermal Patterning in *Arabidopsis*. *Annu Rev Plant Physiol Plant Mol Biol.* **54**, 403-30.

Larkin JC, Marks MD, Nadeau J, and Sack F. (1997). Epidermal Cell Fate and Patterning in Leaves. *Plant Cell* **9**, 1109-20.

Larkin JC, Young N, Prigge M, and Marks MD. (1996). The control of trichome spacing and number in *Arabidopsis*. *Development* **122**, 997-1005.

Liu Y, Mitsukawa N, Oosumi T, and Whittier RF. (1995). Efficient isolation and mapping of *Arabidopsis thaliana* T-DNA insert junctions by thermal asymmetric interlaced PCR. *Plant J.* **8**, 457-63.

Nelson DE, Glaunsinger B, and Bohnert HJ. (1997). Abundant accumulation of the calcium-binding molecular chaperone calreticulin in specific floral tissues of Arabidopsis thaliana. *Plant Physiol.* **114**, 29-37.

Walker JD, Oppenheimer DG, Concienne J, and Larkin JC. (2000). *SIAMESE*, a gene controlling the endoreduplication cell cycle in *Arabidopsis thaliana* trichomes. *Development* **127**, 3931-40.