ABSTRACT

ZHOU, RONGRONG. Recombineering-based Gene Tagging in *Arabidopsis*. (Under the direction of Jose Alonso).

One of the main challenges of the post-genome era is deciphering the function of the over 30,000 genes encoded by the *Arabidopsis* genome. Translational fusions between the genes of interest (in their complete genomic context) and tags such as GFP provide the most reliable and detailed information of *in vivo* gene expression, a key step for understanding gene function. Classical molecular genetic approaches involving *in vitro* manipulation of large genomic DNA fragments are extremely labor-intense and unsuitable for whole-genome scale projects. Therefore, it would be highly desirable to develop alternative high-throughput methods to generate new tools that would enable the plant community to easily obtain accurate information of *in vivo* gene expression and protein localization for any given *Arabidopsis* gene.

Recent advances in *E. coli*-based homologous recombination systems (recombineering) allow for the rapid and precise modification of artificial chromosomes (BACs) without the need for restriction enzymes and classical cloning techniques. In *Arabidopsis*, where homologous recombination is highly inefficient, gene modifications in their chromosomal context can be achieved by recombineering of transformation-ready BAC clones (TACs). Using this system, tags can be precisely fused to the gene of choice in any given position, and subsequently reintroduced into the *Arabidopsis* genome by *Agrobacterium*-mediated transformation.

In this study, we proposed a pipeline suitable for recombineering-based gene tagging in *Arabidopsis*: 1) transformation of target gene containing TACs from *JAtY* clone to recombineering strain SW102, 2) two-steps recombination with *galK* positive/negative

selection, 3) transformation of engineered TACs into *Agrobacterium* and 4) transforming *Arabidopsis* plants to examine expression pattern of a gene and the corresponding protein subcellular localization. 41 genes were selected to evaluate the efficiency and fidelity of each step in the pipeline as well as some of the factors that may affects such parameters. Our results indicate that tags can be efficiently and precisely inserted to a desired location in the gene of interests (GOI) and the modified TAC can be easily transformed into plant. Although deletions of these large T-DNAs are not uncommon during their integration in the plant genome, it is relatively easy to identify transgenic plants carrying a complete copy of the T-DNA. The practical utility of the system has been experimentally demonstrated and the expression pattern and protein localization of several genes examined. Importantly, this system was successfully used to make all the sequence modifications required to observe low-expressed and short-live proteins such as those belonging to the *Arabidopsis* Aux/IAA gene family. Tagged Aux/IAA proteins were proved to be nuclear-localized and trangenes were shown to display similar response to auxin treatment as endogenous genes. Finally, this procedure was optimized by: 1) eliminating the need for the low efficient second recombination step (replacement of the selectable marker by the fluorescent protein) using instead a much more efficient flipase-dependent step, 2) by removing bottle necks such as the preparation of competent cells for each individual clone using instead a pooling strategy, and 3) by developing a web tool to facilitate oligo design and clone selection. These optimizations were critical to quickly generate the constructs for 24 Aux/IAAs and setting the bases for a future whole genome tagging strategy.

Our results have demonstrated that the recombineering-based system can be adopted as a robust, efficient and reliable strategy for the generation of whole-gene reporter fusion

constructs. This work has built a solid foundation for development of 96-well format recombineering pipeline to generate genome-wide gene-tag collection and provide detailed map of the *Arabidopsis* localizome.

Recombineering-based Gene Tagging in *Arabidopsis*

by
Rongrong Zhou

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Functional Genomics

Raleigh, North Carolina

2010

APPROVED BY:

Jose Alonso Committee Chair	David Bird Committee Co-Chair
Robert Franks	Steffen Heber
Wenbin Lu	

DEDICATION

I dedicate this dissertation to my mother and wife.

BIOGRAPHY

Rongrong Zhou was born in Gansu, China in 1979. In 2002, he graduated from Department of Biophysics at Nankai University with a Bachelor of Science degree, and began to pursue a Master degree in the Department of Biological Science & Biotechnology at Tsinghua University. In August 2005, he began to work on his Ph.D in Functional Genomics Program at North Carolina State University. In 2007, he received Master of Bioinformatics degree from Bioinformatics Program at North Carolina State University.

ACKNOWLEDGMENTS

This work has been performed in the laboratory of Dr. Jose Alonso. I would first thank Dr. Alonso for creating an excellent scientific environment, providing brilliant directions and bringing state-of-the-art technologies. Without his encouragement, guidance and support, finishing this work would have not been possible. Additionally, I am very grateful to Dr. Anna Stepanova for being so gracious to guide me with her outstanding experimental skills and knowledge during this work. From initial to end, this work has been influenced by members of Alonso laboratory, as well as those in the laboratory of Dr. Robert Franks, who together provided incredible scientific atmosphere, enlightening discussion, and of course many great helps with my experiments.

I would also like to thank my committee, Dr. David Bird (Co-chair), Dr. Robert Franks, Dr. Steffen Heber and Dr. Wenbin Lu for direction and guidance to my research.

Finally, the most important contributors to my successes have been my family and friends who shared the best and worst moments in my life. My mother, Fangqin Niu and my wife, Huimin Li encouraged me to pursue my graduate career in the United States and put my life into perspective. They love me completely, support me unconditionally, and bring me happiness beyond expectation. Also, I am fortunate to have an amazing group of friends: Jonathan Parham, Segundo Garcia, Eric Nelson and Sean Howell to share my joys and tribulations.

TABLE OF CONTENTS

LIST OF TABLES

Chapter 1

Chapter 2

LIST OF FIGURES

Introduction

Chapter 1

Chapter 2

Chapter 3

INTRODUCTION

Recombineering-based Gene Tagging in *Arabidopsis*: a New Tool for Functional Genomics

With the rapid progress of next generation DNA sequencing technology, there is no doubt that even the 'thousand-dollar genome' could be achieved in the near future (Mardis, 2006). However, information regarding the functions of thousands of genes identified lags far behind the genome sequence with low throughput and quality. In the post-genome era, plant biologists have to face two major challenges: to identify all of the genes encoded by the sequenced genome and to decipher their functions (Figure 1). Upon the availability of genome sequence and powerful bioinformatics tools, localization of the genes in the genome and their basic structural elements (introns, exons and putative regulatory sequences) can be predicted (*Arabidopsis* Genome Initiative, 2000). However, further experimental refinement and validation of these predictions is necessary since gene-finding algorithms are imperfect (Yamada et. al. 2003; Haas et. al., 2005). Clearly, reliable and standardized methods are required to explore gene function at genome scale. Up to date, large-scale phenotype analysis including generation of loss-of-function (Parinov et. al., 2000; Ostergaard et. al.,2004) or gain-of-function mutants (Weigel et. al., 2000; Ichikawa et. al., 2006) has been used for functional research into plant genomes. However, detectable phenotypes could not be generated for many proteins in loss-of-function studies due to genetic redundancy and phenotypes of gain-of-function mutants obtained from ectopic over-expression may reflect patterns more than the real gene function. A parallel approach to complement phenotypic data and generate functional information for thousands of previously uncharacterized genes is *in vivo* study of gene expression at cellular and sub-cellular resolution. For example, having

information about the temporal and spatial expression of the genes that comprise a gene family could be used to guide the selection of double or triple mutants to be constructed and examined in functional studies. Moreover, the study of the spatial-temporal dynamics of a gene expression, and the corresponding protein localization are highly informative approaches crucial for the in-depth functional characterization of a gene. The systematic analysis of protein localization and on a proteome scale could assign putative functions for proteins without obvious phenotype in loss-of-function screens. In summary, high resolution spatial and temporal gene expression information will have a tremendous impact on:

1) Supporting the existence of bioinformatically predicted genes;

2) Discovering phenotypes in mutants of unknown functions by focusing the phenotypic search on the specific cells or developmental times where/when the gene is expressed;

3) Understanding the role of functionally redundant genes. More than two thirds of the *Arabidopsis* genes belong to multigenic families, limiting the power of the gene functional analysis in single mutants. Generation of double, triple, and higher-order mutant combinations is required in many cases. Detailed gene expression information will provide essential criteria for deciding which mutations in the family members should be combined;

4) Enhancing the analysis of gene function in cases where expression patterns and subcellular localization provide essential information to understand the function of a gene.

Therefore, it would be highly desirable to develop tools that will enable the plant community to easily obtain spatial-temporal expression data for any gene of interest. As indicated above, these tools will be valuable for finding conclusive evidences for the

existence of predicted genes, and, in the case of mutant analyses, identifying specific cell types and/or developmental stages where the mutant phenotype may be observable.

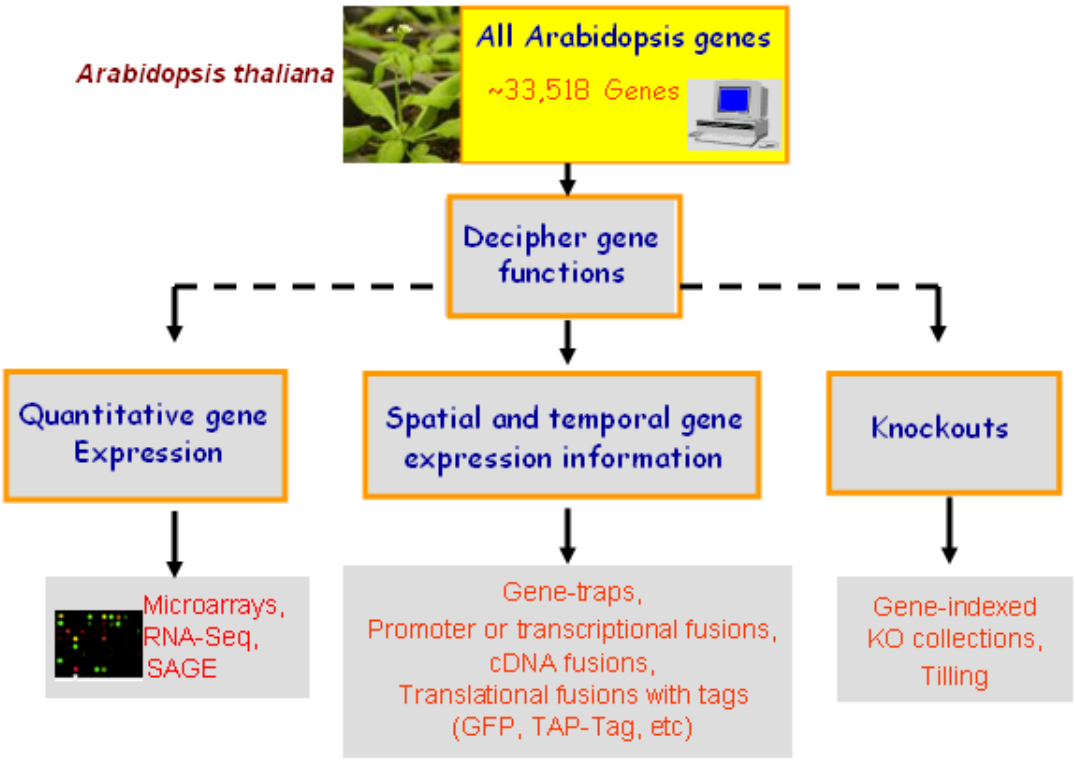


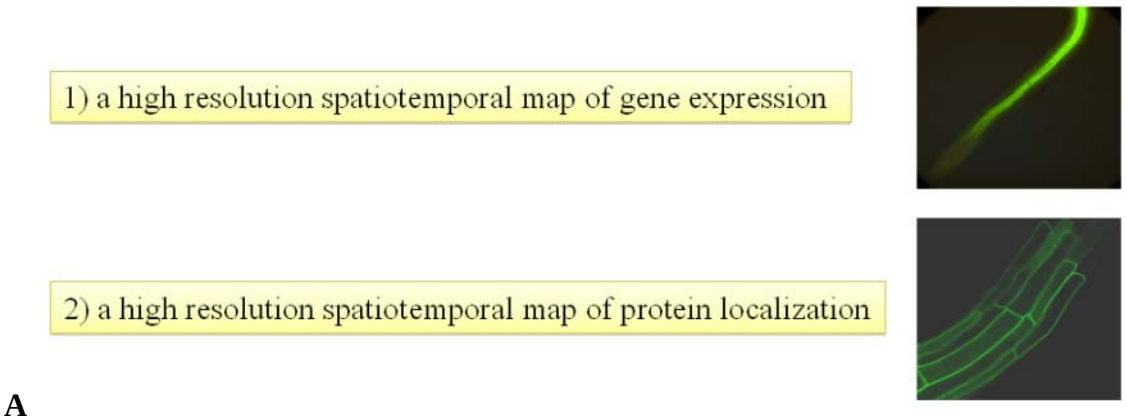
Figure 1. Systematic functional analysis: from genome to function. Extensive sequencing of large EST and full-length cDNA libraries (Seki et. al., 2002; Yamada et.al. 2003; Haas et. al., 2005), together with other genome-wide approaches such as RNA-seq, serial analysis of gene expression (SAGE; Fizames et. al., 2004; Robinson et. al., 2004) and genome expression arrays (Yamada et. al. 2003; Schwab et. al., 2005), have all been extremely useful in the gene discovery process, whereas phenotypic analysis of knockout mutants including gene-indexed knock out (KO) collection (Alonso et. al., 2003) and targeting induced local lesions in genomes (Tilling; Till et. al., 2003; Till et.al., 2006; Comai et. al., 2006), has proven to be tremendously powerful in gene functional analysis. In spite of the success of these strategies used to determine a gene's function, there are still a large number of predicted genes in the *Arabidopsis* genome without experimentally validated function. One of the main reasons for the lack of experimental support for predicted genes and for absence of functional information is the very limited temporal and spatial gene expression data. A variety of different experimental strategies can be used to determine the expression pattern of a gene (gene-traps, promoter or transcriptional fusions) or the subcellular localization of the corresponding gene products (cDNA fusions, translational fusions).

Although antibody-based approaches are widely used to establish protein localization,

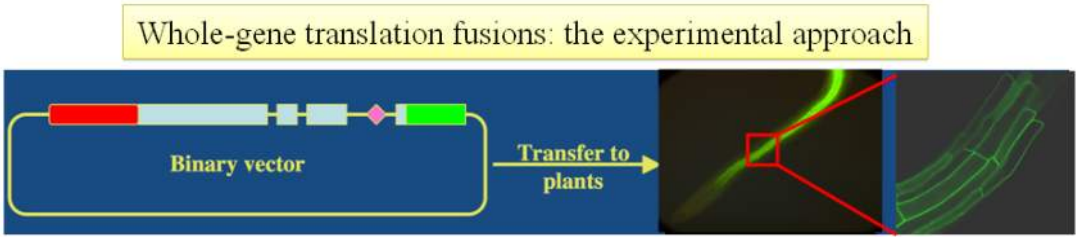
raising and testing antibodies specific for large sets of proteins could be extremely laborious and expensive. In contrast, protein tagging with fluorescent or affinity tags such as Green Fluorescent Protein (GFP, Chalfie et. al., 1994) and expressed under the control of proper regulatory elements can provide reliable and reproducible information and are much more suitable for genome-scale application (Figure 2). In addition to GFP, other tags can also be utilized to generate translations fusion. For example, protein fusions with β-glucuronidase (GUS; Jefferson et. al., 1987), although less suitable than GFP for studies at the sub-cellular level, have the advantage of their relative ease, high sensitivity, and wider accessibility of the detection methods (Taylor, 1997). Combinations of short tandemly fused epitopes (TAP-tags; Rigaut et. al., 1999) can be employed to rapidly purify tagged proteins under conditions that preserve *in vivo* protein-protein or protein-nucleic acid interactions. By combining these purification approaches with highly sensitive analytical methods such as MS-MS (Yates et. al., 2004) or microarray technologies (Brown et. al., 1999), large numbers of protein-protein as well as protein-nucleic acid interactions have been identified. Furthermore, "localization and affinity purification" (LAP) tag can be used for both localization and affinity purification therefore greatly facilitating the characterization of protein function in cellular and developmental contexts (Cheeseman et. al., 2005)**.**

Up to date, yeast has the most detailed localizome map (Kumar et. al., 2002; Huh et. al., 2003) since relatively small and simple constructs can contain entire yeast genes and are easily engineered with classical techniques dependent on the availability of appropriate restriction sites and the size of the fragments to be integrated. However, it is difficult to implement these copy-and-paste techniques in more complex genomes that have more sophisticated gene structure and regulatory elements. In general, multiple cloning steps

are necessary to generate complicated constructs.



A



B

Figure 2. Translational fusions in plant: reliable informers of gene expression and sub-cellular localization. Translational fusions are in-frame gene fusions between tags (such as GFP) and gene of interest. Tags can be inserted at any point in the open reading frame as long as protein function or topology is not disrupted. (A) Translational fusions can be used to examine both the expression pattern (cellular level) and subcellular protein localization. (B) The basic experimental approach of translational fusions in plant includes; in-frame clone of the tags and the gene of interest with all its regulatory sequences into a binary vector (cloning vectors replicate in both *E. coli* and *Agrobacterium*) and subsequent transformation into plant genome.

Currently, cDNA-based transgenes are commonly used for protein tagging and subsequent subcellular localization study in higher organisms. Random cDNA fragments fused to the 3' end of the *GFP* gene have been transformed "*en masse"* into *Arabidopsis* to localize plant gene products, discover subcellular structures and observe dynamic processes in plant cells (Cutler et. al., 2000). Although many useful subcellular markers were obtained, a high proportion of out-of-frame fusions to *gfp* were found and stable transformation used in that study is labor intensive and time consuming. Alternatively, Using the TMV-based vector *30B.GFPc3*, libraries of random cDNAs fused to both the 5'

and 3' ends of *gfp* were transiently expressed on inoculated leaves and screened for high-throughput localization studies (Escobar et. al., 2003). The viral system results in high-level expression of the fusion proteins and allows several hundred cDNA-*gfp* fusions to be screened daily. However, the result is compromised since high-level gene expression could also affect the native proteins subcellular localization. These cDNA-based methods are very useful for examining subcellular localization of the proteins, but not to determine the expression pattern of a gene. Also, transcriptional and post-transcriptional regulation can not be obtained using these approaches since ubiquitous or tissue-specific promoters are used to drive the expression of tagged genes and endogenous regulatory sequences such as introns or untranslated regions (UTRs) are typically missing. Moreover, comprehensive cDNA libraries are not available for most model organisms since sequence-verified, full-length cDNAs, especially large or rare transcripts are expensive and time consuming to generate.

Thus, accurate information on gene expression and subcellular protein localization can best be obtained by using genomic DNA clones which comprise entire coding and regulatory regions of the gene of interest, as well as all splice variants. Commonly used techniques of translational fusion with genomic DNA clones fall into two categories (Figure 3): classical molecular cloning and PCR-based methods (Tian et. al., 2004). Classical molecular genetic approaches are extremely labor-intense and involve digestion with restriction enzymes and *in vitro* manipulation of these large genomic DNA fragments. On the other hand, alternative PCR-based strategies are limited by the size of the genomic fragments that can be efficiently amplified and by the need of extensive re-sequencing of the obtained chimeras. In both cases, some regulatory elements may be inadvertently excluded which will effect the expression of the gene. These methods are

also difficult to scale up since different conditions and protocols are required for processing different genes. Hence, there is still the need for developing high-throughput methods to easily obtain accurate and reliable information of *in vivo* gene expression and protein localization for any given *Arabidopsis* gene.

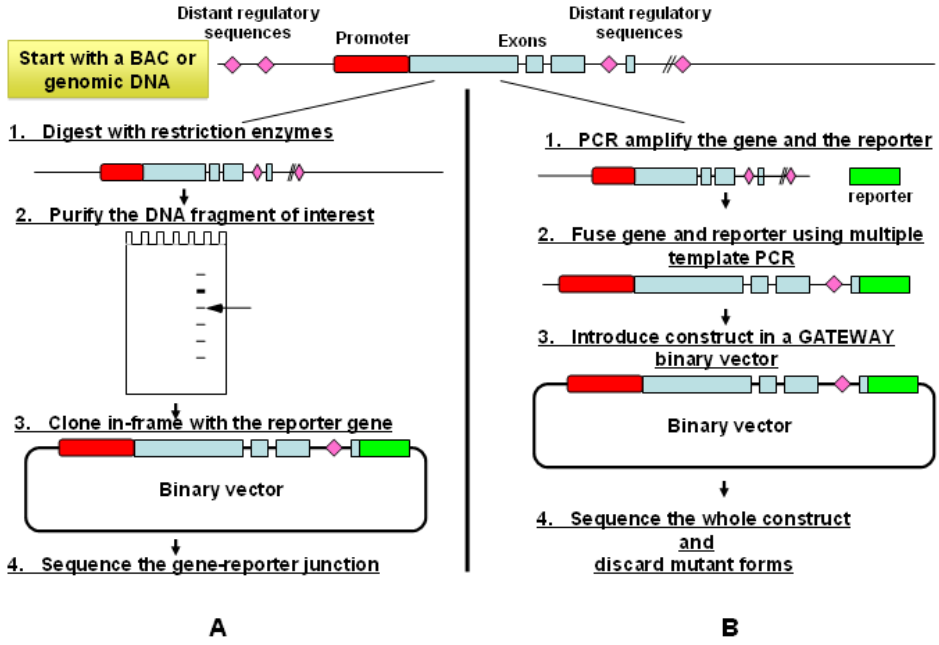


Figure 3. Typical methods used to generate whole-gene translational fusions. A: The Classical molecular genetic approach involves digestion with restriction enzymes, purification the DNA fragment from gel and then clone this DNA in-frame with the reporter gene. B: The alternative way, which could avoid to some extent the problems associated with the classical cloning, is the GATEWAY system. The idea is to PCR amplify the gene and the reporter separately and then use multiple-template PCR to fuse together the gene of interest and the desired reporter. These PCR-based strategies are limited by the size of the genomic fragments that can be efficiently amplified, and by the need of extensive re-sequencing of the obtained cassettes.

λ Red-mediated homologous recombination

As we saw above, typical methods used to manipulate DNA fragments *in vitro* involve several steps of cutting and ligating DNA which unnecessary modifications and mutations can happen. In contrast, *in vivo* strategies based on recombination are more efficient and greatly reduce the number of steps involved in manipulation of large DNA molecules.

One of the first *in vivo* cloning strategies added to the molecular biology tool box takes advantage of the extremely efficient DNA double-strand break and repair pathway in yeast *Saccharomyces cerevisiae* (Szostak et. al., 1983). Transformed linear DNA or short synthetic oligonucleotides can recombine with homologous DNA sequences in the yeast chromosome through very short (<50-bp) regions of DNA identity, thereby generating recombinant DNA *in vivo* without using restriction endonuclease and DNA ligase (Yamamoto et. al., 1992; Baudin et. al., 1993; Lafontaine et. al., 1996). However, homologous recombination in most organisms is such a rare event that a complex set of reactions and extensive homology is required.

Nevertheless, large fragments of genomic DNA can be cloned into yeast artificial chromosomes (YACs) thus recombinant DNA of other organisms can be generated through recombination in yeast and subsequent manipulation of recombinant DNA molecules. Yet, there are inherent limitations in using YACs: 1) The recombination activities in yeast are always active and are not controllable for genetic engineering; 2) Isolation of intact yeast artificial chromosomes (YACs) is tedious and difficult; 3) High degree of chimerism and colonial instability (40-50%) has been reported during transformation and propagation of YACs (Kouprina et. al., 1994; Peterson et. al., 1998).

As an alternative to YACs, Bacterial Artificial Chromosomes (BACs) are much easier to isolate and manipulate due to their smaller size, and therefore, higher transformation efficiency. Since BACs have been widely used for gemomic sequencing, transforming and cloning in bacteria (Shizuya et. al., 1992; Zhao, 2001), more BAC libraries are presently available from different species of animals, plants and microbes. Importantly, BACs have higher stability (Shizuya et. al. 1992; Ioannou et. al. 1994) and show a much

smaller percentage of rearrangements even after many generations of growth (Marra et. al. 1997; Kelley et. al. 1999).

For these reasons, a regulated homologous recombination system in *E. coli* that mimics the yeast double-strand break repair mechanism would be an invaluable tool. Recent advances in *E. coli*-based homologous recombination systems allow for the rapid and precise modification of BACs without the need for restriction enzymes and classical cloning techniques. Recombineering (recombination-mediated genetic engineering) is a powerful method based on homologous recombination systems to modify DNA sequences in *E.coli* or other bacteria (Ellis et. al., 2001). Thus far, strategies to achieve highly efficient chromosome engineering via recombination in *E.coli* with electroporated DNA fall into two main categories, *RecA*-dependent (Yang et. al., 1997; Imam et.al., 2000; Lalioti et. al., 2001; Misulovin et. al., 2001; Gong et. al.; 2002, 2003) and Red/ET homologous recombination systems (Zhang et. al., 1998, 2000; Muyrers et. al., 1999, 2000; Yu et. al., 2000; Datsenko et. al.; 2000; Zhang et. al., 2002).

It has been shown that the RecA protein plays a central role in the recombination pathway of *E.coli* (Roca et. al., 1990). Since BACs are commonly maintained in a *recA*$^-$ genetic background to prevent genomic rearrangement of cloned DNAs, *RecA*-dependent engineering typically requires prior construction of a targeting vector containing a conditional replication origin, homology sequences, the *RecA* gene and a selectable marker (Wang et. al., 2006). As an example of the *RecA* system, a temperature-sensitive shuttle vector expressing recA protein (pSV1.recA) was used to manipulate the sequence of a BAC clone (Yang et. al., 1997). An important characteristic of the *RecA*-dependent engineering is that it requires long homology regions (typically larger than 500 bp).

In comparison, phage-mediated recombination systems can generate recombinant

molecules using short sequence of homology (50 bp or even less). It has been shown that *RecE* and *RecT* genes (encoded by part of the cryptic RAC PROPHAGE) enable genomic DNA to be modified directly with a PCR-generated linear dsDNA targeting cassette (Zhang et. al., 1998). Since linear DNA is rapidly degraded by the bacterial RecBCD nuclease (Benzinger et. al., 1975), linear dsDNA mediated recombination was limited to *recBC*$^-$ strains that lack this nuclease activity. However, *recBC*$^-$ strains grow pooly and the efficiency of recombination is therefore compromised. To solve this problem, the ET system (pBAD-ETγ plasmid) was developed to allow homologous recombination in recBCD$^+$ strains (Zhang et. al., 1998). pBAD-ETγ contains arabinose-inducible *RecE* gene to provide high recombination activity, constitutively expressed *RecT* gene and bacteriophage-λ gam gene to inhibit RecBCD. The main disadvantages of the system are that constitutive gam expression can be toxic to *E.coli* and leaky expression of *RecE/T* from plasmid can cause unnecessary expression of recombination functions.

Functionally analogous to the RecET system is λ-Red homologous recombination system that includes the Gam, Exo, and Beta proteins required for linear dsDNA- mediated recombination. The current model is that Gam inhibits host RecBCD activities, thus protecting electroporated linear dsDNA from degradation so that Exo and Beta can gain access to DNA ends to promote recombination. Exo is a dsDNA-dependent exonuclease, degrading from each end in 5'-3' direction and generating overhangs with 3' single-stranded DNA tails serving as substrates for recombination. Beta, a ssDNA binding protein binds 3' overhangs to protect them, promote annealing of complementary single strands and stimulate the strand exchange.

λ Red genes can be expressed in *E.coli* in different manners. The first way is with

plasmid systems containing inducible promoters (Zhang et.al., 1998; Muyrers et.al. 1999). Construction of a series of Red expression plasmids was also reported to retain tight regulation of λ proteins but with a wider host range (Datta et. al., 2006). In addition, pSC101-BAD-gbaA system (expression of λ Red gene regulated by L-arabinose-induced promoter pBAD) incorporated with both λ-Red and RecA has been shown to help *E.coli* host to survive DNA transformation and enhance the total number of successful recombinatants (Wang et. al., 2006). The disadvantage of the plasmid-based system is the need of plasmid introduction and elimination.

The alternative way is a prophage-based system utilizing *E.coli* strains with a defective λ prophage inserted into the bacterial genome (Yu et. al., 2000). To make recombination system suitable for the manipulation of BACs, *E. coli* strain DY380 was generated by the introduction of the λ prophage into the typical BAC host strain, DH10B to provide transiently inducible recombination (Yu et. al., 2000; Lee et. al., 2001). In this strain, the recombination genes exo, beta, and gam are transcribed from the λ*P*L promoter under the control of a temperature-sensitive λ *cI*857 repressor. Expression of recombination proteins is repressed at 32°C. However, after a brief (15 minute) heat-shock at 42°C, these genes are switched on by inactivation of the repressor and a sufficient amount of recombination proteins can be produced. Also, new recombineering strains (SW102, SW105 and SW106) were constructed from DY380 to allow modification of BACs using *galK* positive/negative selection (Warming et. al., 2005). These strains contain a fully functional *gal* operon except the *galK* gene is missing. Insertion of a *galK* expression cassette into a BAC can restore the ability of these strains to grow on galactose minimal medium, and by selection against *galK* using 2-deoxy-galactose, the *galK* cassette can be

replaced by any dsDNA carrying a desired mutation. SW105 and SW106 in addition contain $_{L}$- arabinose -inducible Flippase (*Flp*) and Cre Recombinase (*Cre*) genes respectively.

The λ red-based recombination system has been used as a genetic tool to modify BACs with linear double-stranded DNA (dsDNA). It has been shown that by using optimal conditions, 0.1% of surviving cells from a standard electroporation can have recombinants (Yu et. al., 2000). Homologous recombination increases 10^4 fold from 20 bp to 40 bp of homology and 10 fold from 40 bp to l kb of homology.

It has also been shown that recombination can take place by using synthetic single-stranded oligonucleotides (SSO) as short as 30 bases long (Ellis et. al., 2001). A 70-base single-stranded oligonucleotide was used to correct the amber mutation $galK_{tyr145am}$. In this case very high recombination efficiencies were observed, ~ 2×10^5 Gal^+ recombinants were obtained per 10^8 viable cells when using 10 ng of oligo. In comparison, 60-, 50-, and 40-based oligos recombined less efficiently and a significant drop in efficiency was found from 40 to 30 bases (Ellis et. al., 2001). When synthetic single-stranded oligonucleotides were used to delete nadA::Tn10 insertion at *galK*, up to 6% of recombinants could be found among the treated cells (Ellis et. al., 2001). λ-Red based ssDNA recombination has also been used to create mutations within the murine Brca2 gene contained on a BAC and a frequency of one recombinant per 90-260 electroporated cells was observed (Swaminathan et. al., 2001). All these examples show λ-Red mediated ssDNA recombination is highly efficient and recombinants can be identified by specific PCR amplification of the mutated BAC from cultures of pooled bacterial cells even in the absence of any selectable marker, thus providing a rapid and highly efficient approach for BAC modification.

λ Beta is required for SSO-mediated recombination activity by promoting homologous recombination between the chromosome and a synthetic ssDNA donor (Ellis et. al., 2001; Yu et. al., 2003). The current model suggests Beta binds ssDNA donor and then triggers SSO to anneal to a transient single-stranded region of DNA (near the DNA replication fork) thus incorporating SSO into the chromosome to generate the recombinant (Ellis et. al., 2001; Li et. al., 2003). It has been demonstrated that oligos corresponding to either of the two complementary DNA strands can be used for recombinants. However, oligos corresponding to the lagging DNA strand recombines with higher efficiency than that targeting to the leading stand (Costantino et. al., 2003). Also, when ssDNA oligos are used to introduce a change near a DNA replication fork, methyl-directed mismatch repair (MMR) has been shown to be triggered therefore affecting ssDNA-mediated recombination by eliminating the recombinant allele and restoring the original sequence (Junop et. al., 2003; Matic et. al., 1996). It has been discovered that MMR can reduce recombination efficiency by >100-fold, however, the efficiency of MMR is determined by the mismatch encoded by the SSO with a pattern for the eight mismatch pairs G-T, A-C, A-A, G-G>T-T, T-C, A-G>C-C (Costantino et. al., 2003). Since C-C mismatch is a fairly weakly bound, and poorly corrected by MMR, it has been shown that oligos creating C-C mispairs can be efficiently recombined with 25% recombinants among the electroporation survivors (Costantino et. al., 2003).

These λ-Red-mediated recombination systems can be easily applied to replace a gene (Figure 4), disrupt any functional gene in BACs by introduction of a selectable marker (drug resistance gene or a prototrophic marker) and introduce point mutations or deletions. A nonselectable DNA fragment (a reporter gene, such as *GFP* or a TAP-Tag) can also be efficiently inserted into a genomic region typically using a two-step methods (Zhang et.

al., 1998; Muyrers et. al., 1999). The first step involves introducing a positive selection marker together with a counter-selectable marker at the desired point in the BAC. In the second step, a DNA fragment containing the desired mutation is targeted to the same region and recombinants are selected based on loss of the counter-selectable marker. Several selectable/counter-selectable strategies have been developed. One of the first strategies is the *sacB-neo* fusion gene that uses *neo* resistance as positive selection and *sacB,* which is lethal to cells in the presence of sucrose, as negative selection (Zhang et. al., 1998). Alternative methods use *rpsL* gene (Imam et. al., 2000) or a rare restriction enzyme such as I-SceI (Jamsai et. al., 2003) as counter-selection. The major drawback of all these methods is that spontaneous point mutations or deletions in the counter-selection component of the system can occur without affecting the growth of bacteria under positive selection, but significantly increase the background under negative selection. In 2005, a novel and highly efficient *galK*-based positive/negative selection system was reported for the manipulation of BACs (Warming et.al, 2005). Galactokinase, the product of *galK* gene catalyzes the phosphorylation of galactose to galactose-1-phosphate, but can also efficiently phosphorylate a galactose analog, 2-deoxy-galactose to 2-deoxy-galactose-1-phosphate which is toxic to the bacterial and cannot be further metabolized (Alper et. al., 1975). Compared with relatively large *sacB-neo* cassette (3 kb), the small size of the *galK* selection cassette (1231 bp) makes PCR amplification and subsequent transformation of PCR fragments into bacteria easier. Most importantly, since *galK* is used for both selection and contra-selection, mutations generated during either PCR amplification or the first recombination will be selected against during positive selection. Thus, this *galK* system achieves significant reduction of the background during negative selection. Overall, the efficiency of *galK*-based positive/negative selection

system is so high that little screening is required (Warming et.al, 2005). The problems of this methodology include long incubation time for selection on minimal growth media, relative expensive 2-deoxy-galactose and specialized wash protocols. Recently, *tolC* has also been reported as a selectable/counter-selectable marker for the facile modification of DNA in *E.coli* (Devito, 2007). Selection of *tolC* insertions/deletions is based on toxicity of the small molecule colicin E1 to bacteria expressing TolC (an outer membrane protein). Compared with *galK* selection scheme, it is faster (selection on LB medium) and more cost-effective (no need for 2-deoxy-galactose).

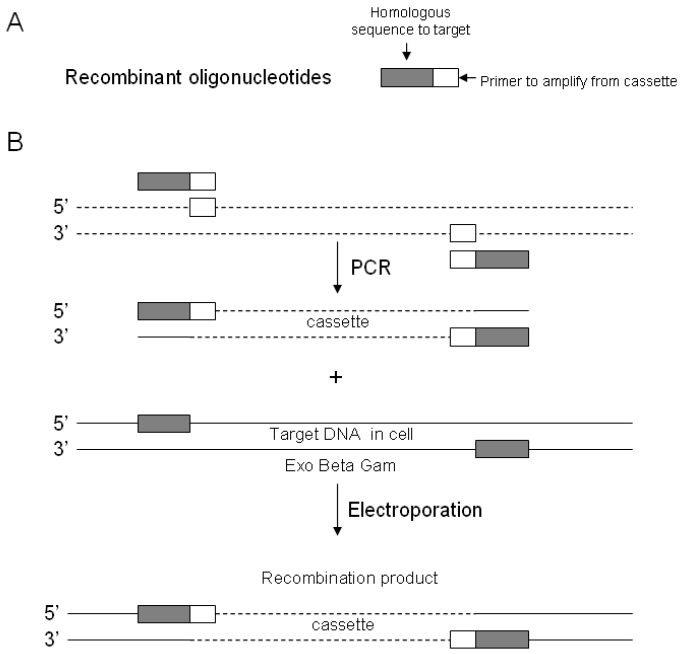


Figure 4. Strategy for gene replacement by homologous recombination. (A) chemically synthesized oligonucleotides including 5′ 50 nt homologous to sequences franking the target gene and the 3′ 20 nt identical to the ends of the selectable cassette to be amplified. (B) A recombineering cassette flanked by 50-bp homologies present at the target is generated by PCR. Exo, Beta, and Gam expression is induced in cells carrying the target DNA. Electrocompetent cells are prepared and then mixed with the amplified cassette. After electroporation, recombination occurs between the homologous sequences on the linear cassette and the target sequences, resulting in the integration of the recombineering cassette and the removal of target sequences.

Recombineering-based gene tagging

For a translational fusion to truly reflect the expression/localization of the gene/protein of interest, there are several prerequisites that need to be fulfilled:

1) The chimeric gene should include all of the regulatory sequences;
2) The protein should be localized in the proper sub-cellular compartment.
3) The added tag should not alter the three-dimensional structure of the tagged protein;

It is difficult to achieve these requirements due to lack of reliable computational tools. However, homologous recombination systems make it possible to keep the tagged-genes in a chromosomal context and maintain all known gene regulatory and protein localization signals required for proper expression/localization. This can be achieved by modifying the BACs containing large genomic regions flanking the gene of interest. In addition, unwanted conformational changes in the chimera can be reduced by using flexible poly-alanine and poly-glycine arms to isolate tags such as GFP from the tagged protein.

In recent years, a comprehensive picture of the core proteome in *Saccharomyces cerevisiae* has been accomplished by genome-wide studies of protein localization (Gavin et. al., 2002). The large-scale application of GFP localization is possible because of efficient intrinsic homologous recombination in yeast. Modifications of genes in their native chromosomal position with electroporated PCR products containing only 40 bp of homology can be achieved through homologous recombination, and tag-coding sequences can be introduced at any particular location of a gene in the genome. Since regulatory elements are effectively preserved in the TAC clones, modified genes can be expressed at near-physiological levels and proper protein complexes can be formed *in vivo.*

However, the efficiency of homologous recombination in many model organisms is several orders of magnitude lower. Hence, transgenes based on large genomic clones such as BACs have been generated by recombineering in bacteria and then used for protein localization studies in worm, flies, zebrafish, mice and human tissue-culture cells. Mapped BAC libraries are typically generated as part of genome sequencing projects (Mozo et. al. 1999; Hoskins et. al. 2000; Osoegawa et. al. 2000; McPherson et. al. 2001). Clones in these libraries harbor large DNA inserts containing complete transcription units together with extensive surrounding DNA likely to contain all their regulatory sequences. High-fidelity λ phage-mediated *in vivo* recombineering strategy in *E.coli* has enabled rapid and precise modification of these large constructs. One of the key advantages of recombineering is DNA context independence since efficient homologous recombination can be carried out between short, terminal homology regions on a PCR product and sequences on a BAC. Furthermore, a single BAC is likely to contain all the regulatory sequences required for proper temporal and spatial expression of a gene contained in that BAC, so that transgenes can be expressed at normal levels and specific cellular/subcellular localization. Recent advances of DNA recombineering have further demonstrated that multiple DNA engineering steps in *E.coli* can be carried out in liquid culture without sacrificing efficiency and fidelity, therefore a door for high-throughput application of recombineering has been opened up.

In *Arabidopsis*, where homologous recombination is highly inefficient, the gene modifications in a chromosomal context can be achieved by recombineering of transformation-ready BAC clones (TACs) (Figure 5). Using this system, tags can be fused to the gene of choice and subsequently reintroduced into the *Arabidopsis* genome by

Agrobacterium-mediated transformation. This method has many important advantages over classical cloning and PCR-based methods:

1) All known regulatory sequences are retained;

2) Very few undesired mutation are introduced so only junction and tag sequences need to be re-sequenced after recombination;

3) Tags can be introduced to any given position in the gene of interest;

4) The use of restriction enzymes is no longer needed and the same experimental protocol can be used for all of the genes being tagged.

One last advantage of the recombineering method is that, at least in theory, it should be possible to scale it up and use it as a high-throughput gene tagging system. Although there are also some potential disadvantages such as the introduction of additional copies of many different genes in the recipient genome, the advantages clearly overweight the problems.

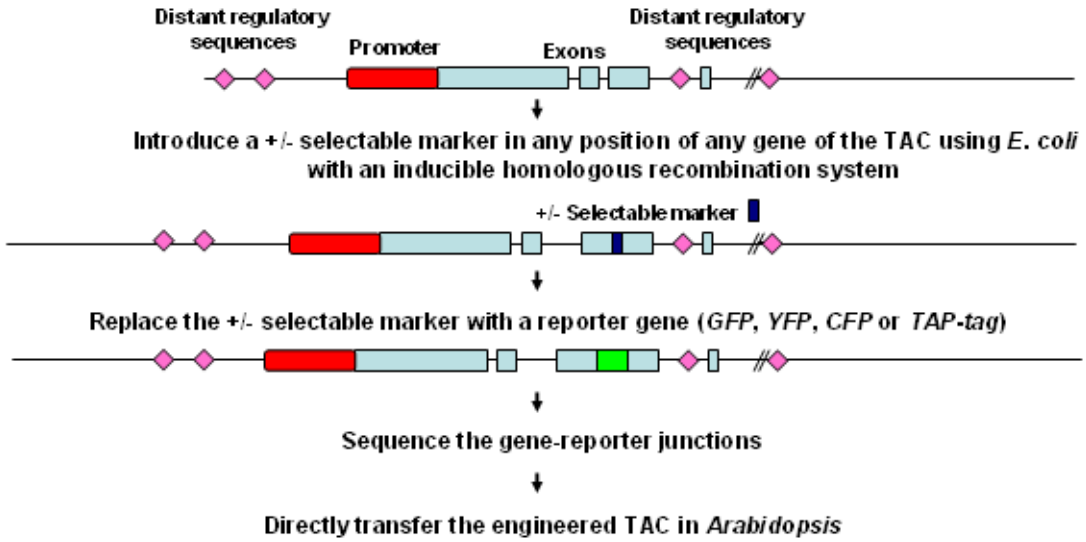


Figure 5. Recombineering-based gene tagging in *Arabidopsis*. Starting with a transformation-ready BAC clone (TAC), a selectable marker is introduced in the first recombination step and subsequently replaced by desired tag through second recombination. The engineered TAC can be reintroduced into *Arabidopsis* through *Agrobacterium*-mediated transformation after sequencing confirmation.

High-throughput recombineering pipelines

To systematically explore localizomes in higher organisms, it is necessary to have high-throughput generic pipelines that can simply, rapidly and inexpensively generate tagged transgenic sets for all the genes in the genome. However, two important criteria have to be satisfied. First, expression levels and patterns of tagged proteins must reliably and reproducibly match the endogenous counterparts. Second, every step of the process must be efficient, precise and scalable, and therefore, suitable for large-scale use. Recently, recombineering pipelines which support the large scale production of BAC transgenes have been developed and applied for functional genomics studies for higher organisms (Figure 6).

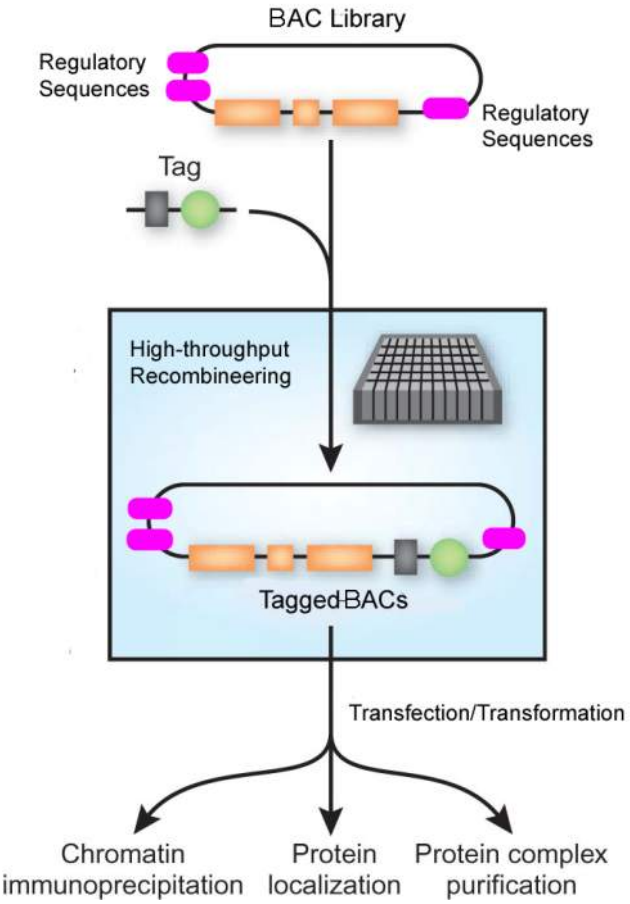


Figure 6. A schematic of the high-throughput pipeline to engineer BACs with recombineering and subsequently introduce modified BACs into higher organisms. A TAC library can be rapidly modified with a modular tag introduced (96-format at one time). After transfected into mammal cells or transformed into plant cells, protein function can be interrogated in high throughput.

C. elegans provides an excellent model for protein tagging and functional studies due to it is small size and invariant cell lineage. An efficient recombineering pipeline has been described to engineer *C. briggsae* genomic BAC clones to generate tagged transgenes (Sarov et. al., 2006). A *pRedFlp* expression plasmid including a rhamnose-inducible promoter driven Red operon and Flp recombinase controlled by tetracycline-inducible promoter is transformed into the BAC host cells containing a suitable BAC clone preselected for the gene of interest. *egfp-kan* cassette is electroporated into the recombination-competent host cells after induction of Red operon (red) with rhamnose.

Next, *kan* marker flanked with Flippase Recognition Target *(FRT)* sites is removed by anhydrotetracycline-induced expression of Flippase. In the final step, a subcloning cassette *pPUB* containing *unc-119* trasformation marker for *C. elegans*, the blasticidin resistance gene (*bsd*) for selection in *E.coli* and homology arms to subclone the genomic region is electroporated into host cell after induction of red operon. The recombineering pipeline can be efficiently carried out in the liquid culture and allows simultaneous processing of several genes. Nine out 11 genes were successfully tagged and subcloned in two independent experiments. Similar transformation efficiency as commonly used *unc-119* vectors (pAZ132, Maduro et. al., 1996) was achieved when generating transgenic worms using *EGFP*-tagged *C.briggsae* ortholog of *lin-59* by ballistic transformation and the same expression pattern was observed in all GFP-positive lines. This approach is simple, fast and all the steps can be processed in liquid culture without checking or cloning until the final product. Therefore, it can be feasibly scaled up for parallel processing of several genes at the time, and generate tagged transgene sets for all of protein-coding genes in *C. elegans* at a relatively low cost.

Similarly, a toolkit and robust pipeline has been described for the generation of reporter genes within *C. elegans* fosmid clones (Tursun et. al., 2009). The protocol makes use of a recombineering strain SW105 and a single selection marker, the *galK* gene, flanked by *FRT/Flp* sites to insert fluorescent proteins (GFP, YFP, CFP, VENUS, mCherry) or affinity tags at specific target sites.

Also, an efficient and versatile tagging technology paving the way to genome-wide gene tagging in *Drosophila* has been reported (Venken et. al., 2008). Novel N-tag- and C-tag-template vectors containing a tag (such as Cypet, a cyan fluorescent protein and YPet, a yellow fluorescent protein derivative) and a *kanamycin* (*Kan*) selection marker

were constructed. 70-bp primers including 20-bp homology to the tag-template vectors required for PCR amplification and 50-bp homology arms to the target region required for recombineering are used to generate tag-selection cassettes by PCR. Amplified cassettes are subsequently transformed in recombination competent bacteria that harbor *P*(*acman*) BAC vector (Venken et. al., 2006) containing the gene of interest. Recombination between the homology arms in the PCR fragment and the gene of interest occurs and recombination events are selected using *kanamycin* and further identified by PCR. The *Kan* cassette can be removed using Cre recombinase; and therefore, the final product is a gene fusion between the tag and gene of interest, with an in-frame *LoxP* site as linker. It has been demonstrated that this recombineering procedure is efficient to generate protein fusions at either terminus in an endogenous genomic context. Also, tagged constructs have been shown to express under the proper control of regulatory elements, localize at expected positions and function similarly to their endogenous counterparts in transgenic *Drosophila* after P-element-mediated transformation. The method allows for the efficient tagging of any gene of interest in *Drosophila* transformation vectors. Tagged proteins can maintain their functionality and detectability in live animals. The P vectors used in the pipeline contain an *attB* site which allows for ϕC31-mediated integration of larger pieces of DNA and facilitate high-throughput applications.

In addition, a fosmid vector (*pFlyFos*) containing *attB* sequence and the eye promoter (driving dominant selectable marker *dsRed*) was developed and used to construct two complementary genomic fosmid libraries for *Drosophila melanogaster* and *Drosophila pseudoobscura*. Then a high-throughput 96-well-format liquid culture recombineering pipeline was adapted to introduce fluorescent tags into Flyfos clones (Ejsmont et. al., 2009). It has been shown that the fosmid transgenes can recapitulate endogenous gene

expression pattern and allow imaging of gene products in living flies (Ejsmont et. al., 2009).

Recently, engineered BACs generated by recombineering have been successfully used to modify mice embryonic stem (ES) cells (Zhou et. al., 2004). And recent development of "ES cell complementation of tetraploid embryos" technology has made it possible to clone mice directly from ES cells. By injecting ES cells containing tagged genes into tetraploid blastocysts, animals that are born derive from injected cell and can be used for protein localization and interaction studies. Combining recombineering and ES cell complementation of tetraploid blastocysts, transgenic mice containing tagged gene can be produced in 3 months and the same protocol can be used to tag any mouse gene with high throughput. BAC TransgeneOmics, a fast and reliable pipeline to rapidly generate BAC transgenes using 96-well-format recombineering and study protein function in mammalian cells, has been reported (Poser et. al., 2008). A modified version of the TAP-tag containing enhanced GFP for protein localization and S-peptide for affinity purification has been generated. In the N-terminal cassette, a *neo* gene is placed within an artificial intron inside *EGFP* sequence and flanked by two *loxP* sites which can be removed by Cre recombinase. In the C-terminal cassette, tag sequence and *neo* resistance gene are separated by an internal ribosome entry site (IRES). The recombineering plasmid *pSC101gbaA* is introduced to bacterial cultures each containing a specific BAC clone. A shift in temperature from 30 °C to 37 °C and addition of $_{L}$-arabinose can be used to induce the expression of the proteins required for recombination. Tagging cassettes containing 50 nucleotides of PCR-introduced homology arms are electroporated to competent cells and successfully recombined clones are selected by kanamycin. The recombination plasmid containing temperature-sensitive *pSC101* origin of replication can

be easily removed by incubation at 37°C. 86% success rate was reported after high-throughput transfection of human cell culture and transgenes were expressed in 90% of transgenic cells and integrated in low or single copy number. It was demonstrated that isolation of single cell-derived clones is not necessary since transgene expression levels between transgenic lines generated from independent BAC clone and clone pools do not vary significantly. The pipeline performs very well for protein localization, tandem affinity purification and chromatin immunopurification (ChIP). Fifteen well-characterized genes were selected to test the reliability of the tagging system and known localization patterns were reproduced for 11 of these. Also, all 15 tagged-proteins could be purified by two-step affinity purification based on GFP and S-peptide. It has also been shown that the DNA-binding sites identified by CHIP with tagged BAC transgene produce comparable results to conventional antibody-based approach. Moreover, precise expression pattern of a *GFP*-tagged GOI (*PCNA*) was observed in BAC-transgenic ES cells and ES-derived transgenic mice. Thus, the BAC tagging approach can be applied to a multicellular context to explore protein function in developing mouse embryo or adult mice. In summary, BAC TransgeneOmics has a success rate of 80-90% at each step of the pipeline. Liquid handling, 96-well format and automation of most experimental steps enable the system to rapidly generate comprehensive tagged BAC transgenes with genome wide coverage. Furthermore, the authors suggest the same high-throughput approach is amenable for general application to other model systems that permit stable transgenesis.

High-throughput gene tagging in *Arabidopsis*: from genome to localizome

In 2000, the *Arabidopsis* genome became the first sequenced plant genome (The *Arabidopsis* Genome Initiative, 2000). With a relatively smallest plant genome size (Bennett et. al., 2003) and a relatively short life cycle, *Arabidopsis* is widely used as a

model organism and a popular tool in plant biology and genetics studies (Rensink et. al., 2004; Coelho et. al., 2007). The most up-to-date version of the *Arabidopsis thaliana* genome information is provided by *Arabidopsis* Information Resource (TAIR, http://www.arabidopsis.org/). The newest TAIR9 release contains 27,379 protein coding genes, 4827 pseudogenes or transposable elements and 1312 ncRNAs (33,518 genes in all, 39,640 gene models).

Plant cells are organized into a complex network of subcellular compartments that are specialized for a multitude of biological functions. Comprehensive knowledge of the gene expression patterns and localization of corresponding protein products within these cellular microenvironments is crucial to fully understand plant physiology, metabolism and development. Furthermore, the subcellular location of a protein can provide clues for the molecular functions and biological processes in which a protein may be involved (Carrari et. al., 2003; Huq et. al., 2003) since there is a tight correlation between protein localization and function (Kumar et. al., 2002; Huh et. al., 2003).

Unfortunately, our knowledge about subcellular locations of *Arabidopsis* is scarce. Only about 5% (1,300) of all predicted *Arabidopsis* proteins have had their subcellular locations determined empirically (GO Term Enrichment, The *Arabidopsis* Information Resource [TAIR], http://arabidopsis.org/tool/bulk/go/index.jsp). Hence, the functional characterization of the *Arabidopsis* genome demands high-throughput tools and technologies to analyze gene expression, protein localization, and other *in vivo* biochemical aspects of gene function. Now that it is clear: 1) reliable functional information can be best obtained by *in vivo* expression of tagged proteins using their native genomic context; 2) tags can be precisely and easily fused to a gene of choice present in a large artificial chromosome by highly efficient homologous recombination

systems in *E. coli* (Lee et. al., 2001), these type of technologies should also be implemented in plant systems.

The binary BIBAC vector can replicate as a single-copy plasmid in both *E.coli* and *Agrobacterium*. It has been shown that a 30-kb yeast genomic DNA fragment and a 150-kb human genomic DNA fragment can be inserted into the BIBAC vector and maintained in both *E. coli* and *Agrobacterium tumefaciens* (Hamilton et. al., 1997). Notably, BIBAC system has been demonstrated to be able to transfer more than 150 kb of DNA into plant chromosomes (Hamilton et. al., 1996). Including helper plasmids in addition to BIBAC has been shown to increase the efficiency of transfer of unusually large BIBAC T-DNAs (Hamilton et. al., 1997). These plasmids carry additional copies of *Agrobacterium tumefaciens* virulence genes *virG* and *virE.* pYLTAC7, a TAC vector was also developed to accept and maintain large genomic DNA fragments in both *E.Coli* and *Agrobacterium tumefaciens* (Liu et. al., 1999). It has been shown that large genomic DNA fragments of *Arabidopsis thaliana* can be successfully cloned into this vector and most of the DNA fragments can be stably maintained. Also, cis sequences contained in the vector make it suitable for *Agrobacterium*-mediated gene transfer of large T-DNA into plants without helper plasmids commonly used in the binary-BAC system. It has been demonstrated that TAC clones carrying 40- to 80-kb of genomic DNA fragments can be efficiently transferred back into *Arabidopsis* and faithfully inherited among the progeny. Furthermore, the ability to complement the gravitropic defects of a SGR1 locus homozygous mutant line by TAC clone carrying 80-kb inserts covering SGR1 region proved the practical utility of this vector system for positional cloning in *Arabidopsis*. Using the vector pYLTAC17, a derivative of the previously introduced pYLTAC7 (Liu et. al., 2002), a TAC library has been constructed at John Innes Centre (Norwich, UK) using

DNA extracted from purified nuclei from the leaves of *Arabidopsis thaliana* ecotype *Columbia* through partially digestion with HindIII and ligation into the HindIII site of the vector. It consists of 36,864 clones with average insertion size 80 kb and provides 14x coverage of the *Arabidopsis* genome Importantly, more than 8,000 clones covering 90% of the *Arabidopsis* genes have been end sequenced therefore the gene content of each one of these clones is known.

Additionally, a method to transform large DNA molecular by electroporation has been developed (Sheng et. al., 1995) therefore BACs can be efficiently transferred between different *E.coli* strains and *Agrobacterium tumefaciens*. In addition, developments of the *in planta Agrobacterium*-mediated transformation methods have made it possible to generate a large population of transgenic *Arabidopsis* (Bechtold et. al., 1998; Clough et. al., 1998).

Hence, the vectors and the libraries suitable for manipulation large DNA molecules with powerful DNA recombineering technology should be valuable resources for efficient and reliable gene function studies and open up the possibility to develop a recombineering pipeline for plant.

Here we propose an efficient, reliable and scalable approach for TAC-based tagging system in *Arabidopsis thaliana* (Figure 7A). The first step is the identification of a suitable TAC clone (TAC clone with smaller size yet larger DNA sequences flanking the gene of interest is recommended) where the gene of interest is situated approximately in the middle of the clone. This can be easily achieved using the ATIDB browser (http://atidb.org/cgi-perl/gbrowse/atibrowse/). Once the TAC clone has been identified, it is transferred to a recombineering *E. coli* strain (SW102 or SW105). Next, depending on the position in the gene where the tag is going to be inserted, a pair of primers is designed

(Figure 7B). Each one of these two primers will consist of two parts, the first part or 5' region of the forward primer, for example, consists of 50 nt identical to the 50 nt of the gene of interest just upstream of the point where the GFP will be inserted. The second part of the forward primer will consist of ~18 nt corresponding to the 5' of the recombineering cassette (i.e. the *galK* cassette or the GFP cassette). Similarly the reverse primer will also have 2 parts, but in this case they will correspond to the 50 nt just downstream of the point where the GFP is going to be inserted and the ~24 nt required for the amplification of the recombineering cassette. Using these primers the recombineering cassette is amplified (for example the *galK* cassette), resulting in a DNA product formed by the *galK* cassette flanked by 50 nt of sequence of the gene of interest at either side. This linear DNA is electroporated in recombineering competent cells containing the TAC with the gene of interest, and the recombination events selected using growth media where only bacteria that have incorporated the *galK* cassette can grow. Using a similar PCR strategy the same 50 nt of gene specific sequences can be incorporated to each side of the GFP cassette. This GFP cassette can then be used to replace the *galK* during a second recombination event resulting in a final product where the gene of interest contained in a large TAC clone is been fused to GFP. This modified TAC clone is then transferred to a *recA- Agrobacterium tumefacien* strain that is then used to mediate the incorporation of this "recombineered" TAC clone into the genome of the plant.

We describe construction of TAC transgenes using a robust recombineering procedure and establish protocols for efficient, stable transformation of these large constructs into *Arabidopsis* plants. We demonstrate the feasibility of this approach for the analysis of protein cellular and subcellular localization. We also optimize steps in the pipeline and prepare it for high-throughput application to tag all the genes in *Arabidopsis* genome. We

believe this robust and reliable recombineering-based *Arabisdopsis* gene tagging system along with optimizations for potential high-throughput applications would dramatically accelerate gene function studies in the postgenomic era and facilitate the generation of a reliable localizome map in the reference plant *Arabidopsis*.

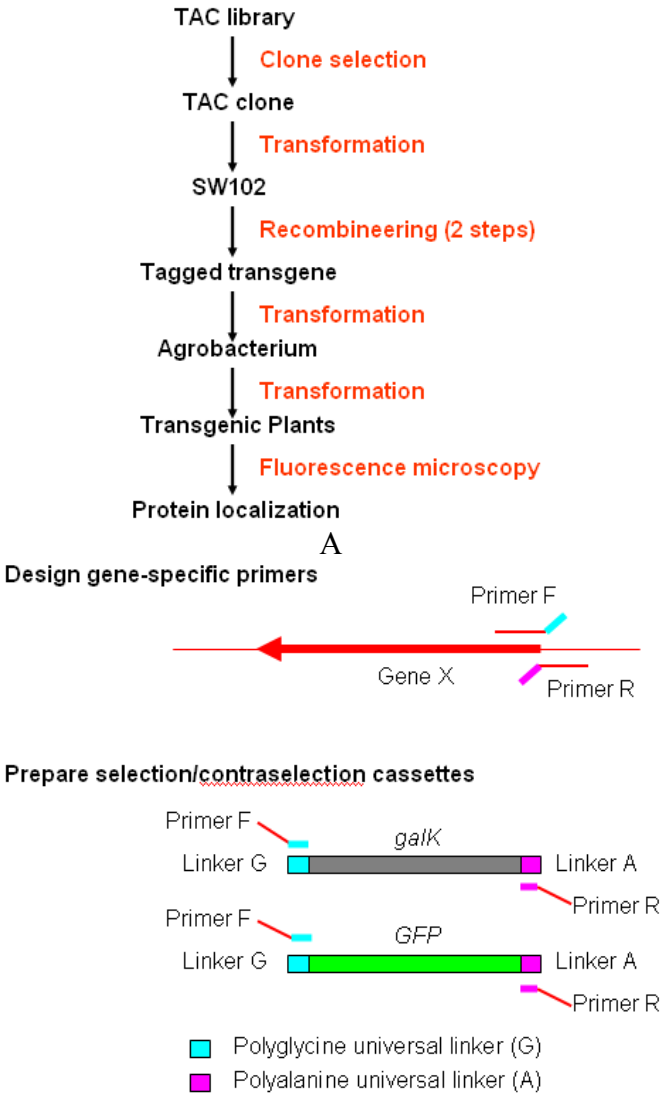


Figure 7. *Arabidopsis* tagging pipeline. (A) Flow chart of the pipeline. (B) Schematic representation of the two recombination events to create a precise GENE-GFP fusion. In the first step, gene of interest is tagged by homologous recombination creating a in-frame *galK* fusion. In the second step, *galK* is replaced by GFP. Tag is flanked by the the upstream Linker G (GGA GGT GGA GGT GGA GCT) (Gly-Gly-Gly-Gly-Gly-Ala) and downstream Linker A (GCT GGT GCT GCT GCG GCC GCT GGG GCC) (Ala-Gly-Ala-Ala-Ala-Ala-Ala-Gly-Ala) to provide structural flexibility to the junction sites.

CHAPTER 1

Recombineering-mediated Tagging of *Arabidopsis* Genomic Constructs for *in vivo* Localization

SUMMARY

Inducible homologous recombination in bacteria has been shown to be very efficient in the precise manipulation of DNA sequences contained in large BAC clones, and therefore could be used to generate gene-reporter fusions in a pseudochromosome context. On the other hand, TAC (*T*ransformable bacterial *A*rtificial *C*hromosomes) clones containing the gene of interest with all its regulatory sequences have been developed to facilitate complementation experiments and allow transformation of the genomic DNAs contained in these clones back into the plant genome via *Agrobacterium*-mediated transformation. We combined recombineering-based gene tagging using the λ-Red system and the *JAtY* TAC library to develop a pipeline to generate whole-gene *GFP* fusions required in detailed spatial-temporal gene expression studies. However, before recommending this system as a general tagging tool in *Arabidopsis*, it seemed desirable to experimentally evaluate the qualities of this approach. Here we examined the efficiency and fidelity of each step in the recombineering procedure as well as some of the factors that may affects such parameters. We also describe a simple procedure to obtain better transformation efficiency using *JAtY* clones. Our results indicate that the combination of the λ-Red recombineering system and the end-sequenced *JAtY* library can be adopted as a general

strategy for the generation of whole-gene reporter fusion constructs to study gene function in *Arabidopsis*.

Table 1. List of recombination constructs generated. ID: short name for each of GOI processed in the project and 8 constructs, P1-P8 were generated by Dr. Jose Alonso. AtY: name of the TAC clone. Size: length in bp of the *Arabidopsis* genomic DNA present in a particular *JAtY* clone. This estimate is based on the end sequences of the *JAtY* clone. Tag: G for *GFP*, S for single copy of codon-optimized *Ypet* (*Sypet*) and T for three copies of codon-optimized *Ypet* (*Typet*). Details for these two improved versions of fluorescent proteins are described in chapter 2. E1: efficiency of the second recombination (replacement of the *galK* by the *GFP* sequences) as percentage of PCR confirmed recombinants out of tested colonies*)*. E2: efficiency of plant transformation (percentage of resistant lines geminated from 8,000 T1 seeds on basta plates). Sequence: number of clones that needed to be sequenced to find one with no mutations. Position: N for N-terminal tagging and C for C-terminal tagging. nd: not determined. na: not apply.

Table 1 (continued)

Gene	Name	ID	JAtY	Size	Tag	E1	E2	Sequence	Position
At5g53980	*ATHB52*	*P1*	*49E13*	74Kb	G	nd	nd	1	N
At1g70560	*TAA1*	*P2*	*50P13*	82Kb	G	75%	nd	1	N
At1g70560	*TAA1*	*P3*	*73L21*	30Kb	G	80%	nd	1	N
At1g23320	*TAR1*	*P4*	*70I13*	89Kb	G	nd	nd	1	N
At4g24670	*TAR2*	*P5*	*73P23*	104Kb	G	nd	nd	1	N
At4g24670	*TAR2*	*P6*	*76L12*	62Kb	G	100%	nd	1	N
At1g34040	*TAR3*	*P7*	*62I21*	64Kb	G	37%	nd	2	N
At1g34060	*TAR4*	*P8*	*62I21*	64Kb	G	100%	nd	2	N
At4g14560	*IAA1*	*A1*	*68P07*	68Kb	G	29%	0.18%	1	C
At3g23030	*IAA2*	*A2*	*78G07*	78Kb	G	61%	nd	1	C
At1g04240	*IAA3/SHY2*	*A3*	*53F09*	76Kb	G	35%	0.11%	1	C
At5g43700	*IAA4*	*A4*	*57M08*	66Kb	G	87%	0.4%	1	C
At1g15580	*IAA5*	*A5*	*61G08*	27Kb	G	96%	0.74%	1	C
At1g52830	*IAA6/SHY1*	*A6*	*74P09*	38Kb	G	87%	0.21%	1	C
At3g23050	*IAA7/AXR2*	*A7*	*78G07*	78Kb	G	30%	0.13%	1	C
At2g22670	*IAA8*	*A8*	*68I18*	69Kb	G	37%	nd	1	C
At5g65670	*IAA9*	*A9*	*70O17*	62Kb	G	54%	0.2%	1	C
At1g04100	*IAA10*	*A10*	*74L24*	65Kb	G	58%	nd	1	C
At4g28640	*IAA11*	*A11*	*52A17*	64Kb	G	41%	0.35%	1	C
At1g04550	*IAA12/BDL*	*A12*	*58D24*	37Kb	G	89%	0.74%	1	C
At2g33310	*IAA13*	*A13*	*80H09*	67Kb	G	33%	0.53%	1	C
At4g14550	*IAA14/SLR*	*A14*	*68P07*	68Kb	G	96%	0.3%	1	C
At1g80390	*IAA15*	*A15*	*80H05*	69Kb	G	42%	0.8%	1	C
At3g04730	*IAA16*	*A16*	*53E01*	74Kb	G	58%	0.18%	1	C
At1g04250	*IAA17/AXR3*	*A17*	*53F09*	76Kb	G	87%	0.11%	1	C
At3g15540	*IAA19/MSG2*	*A19*	*52P18*	79Kb	G	83%	0.27%	1	C
At2g46990	*IAA20*	*A20*	*60K12*	75 Kb	G	13%	0.9%	1	C
At3g16500	*IAA26/PAP1*	*A21*	*66K03*	47Kb	G	91%	0.44%	1	C
At4g29080	*IAA27/PAP2*	*A27*	*57L03*	67Kb	G	42%	nd	1	C
At5g25890	*IAA28*	*A28*	*71M07*	76Kb	G	62%	0.13%	2	C
At4g32280	*IAA29*	*A29*	*62E08*	83 Kb	G	13%	0.29%	1	C
At3g62100	*IAA30*	*A30*	*68P09*	60 Kb	G	17%	1.16%	1	C
At3g17600	*IAA31*	*A31*	*74G10*	38Kb	G	37%	2.64%	1	C
At2g01200	*IAA32*	*A32*	*67N06*	57Kb	G	17%	0.32%	1	C
At5g57420	*IAA33*	*A33*	*54H01*	54Kb	G	79%	0.9%	2	C
At1g15050	*IAA34*	*A34*	*64F13*	64 Kb	G	4%	0.33%	1	C
At1g22780	*PFL*	*H1*	*50E23*	27Kb	G	nd	nd	1	C
At1g66240	*ATX1*	*H2*	*53J22*	44Kb	G	nd	nd	1	C
At3g56240	*CCH*	*H4*	*49O07*	59Kb	G	nd	nd	1	C
At3g53420	*PIP2A*	*H5*	*69J10*	70Kb	G	nd	nd	1	C
At3g18280	*LTP*	*H6*	*68L19*	82Kb	G	nd	nd	1	C
At1g22780	*PFL*	*H1*	*50E23*	27Kb	S	nd	nd	1	C
At1g66240	*ATX1*	*H2*	*53J22*	44Kb	S	nd	nd	1	C
At3g56240	*CCH*	*H4*	*49O07*	59Kb	S	nd	nd	2	C
At3g53420	*PIP2A*	*H5*	*69J10*	70Kb	S	nd	nd	1	C
At3g18280	*LTP*	*H6*	*68L19*	82Kb	S	nd	nd	1	C
At1g22780	*PFL*	*H1*	*50E23*	27Kb	T	nd	nd	1	C
At1g66240	*ATX1*	*H2*	*53J22*	44Kb	T	nd	nd	1	C
At3g56240	*CCH*	*H4*	*49O07*	59Kb	T	nd	nd	1	C
At3g53420	*PIP2A*	*H5*	*69J10*	70Kb	T	nd	nd	2	C
At3g18280	*LTP*	*H6*	*68L19*	82Kb	T	nd	nd	1	C

RESULTS AND DISCUSSION

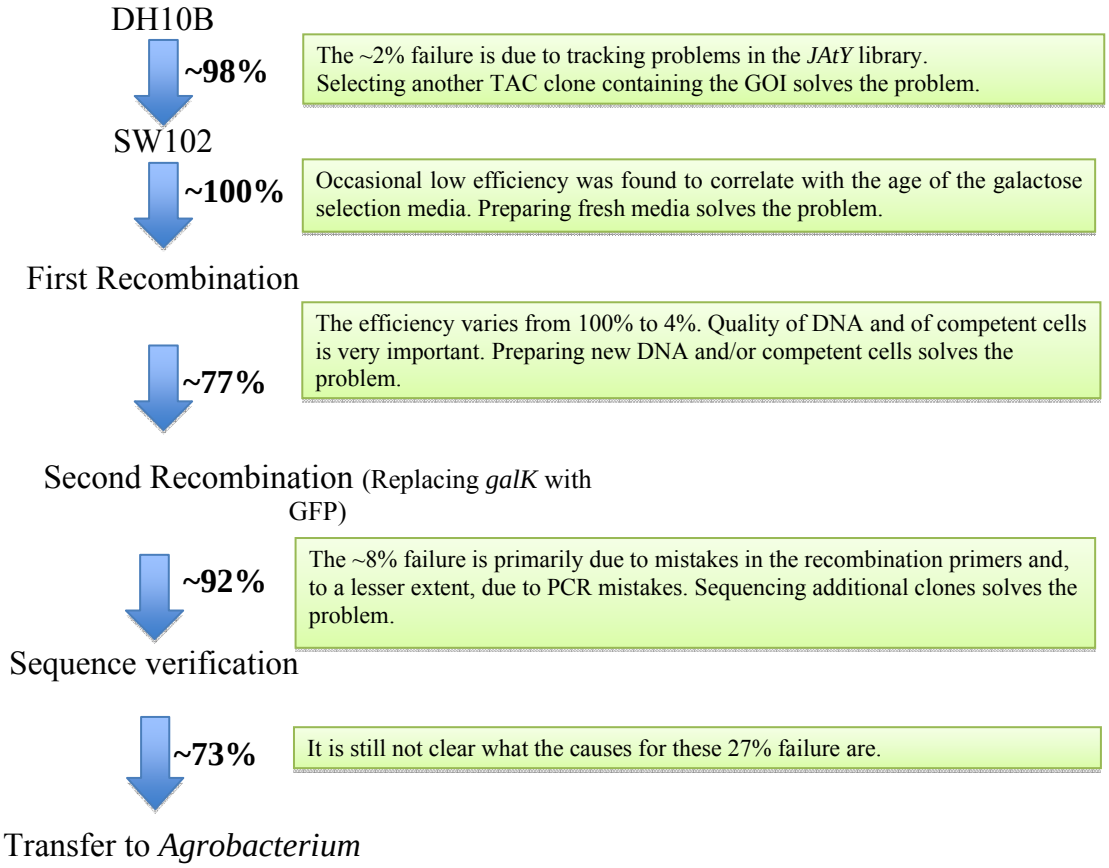
Efficiency of the process

Although the effectiveness of the λ-Red system to modify large DNA clones has been thoroughly examined using mouse BAC clones (Poser et. al., 2008), as well as *C. elegan* fosmids clones (Tursun et. al., 2009), it is prudent to re-examine the efficiency, fidelity and integrity of the system using *Arabidopsis JAtY* TAC clones. Hence, to determine the suitability of recombineering as a general tool for gene function studies in *Arabidopsis*, the efficiency and fidelity of each one of the seven steps involved (Figure 7A) was examined using *JAtY* clones ranging in insert size from 27kb to 104kb (Table 1) and the average efficiency of each step was estimated (Figure 8A). Among these *JAtY* clones, 6 of them with insert size from 27kb to 79kb (with approximately 10 kb increments between them) were carefully evaluated to explore the possible correlation between TAC size and efficiency of each step (Table2).

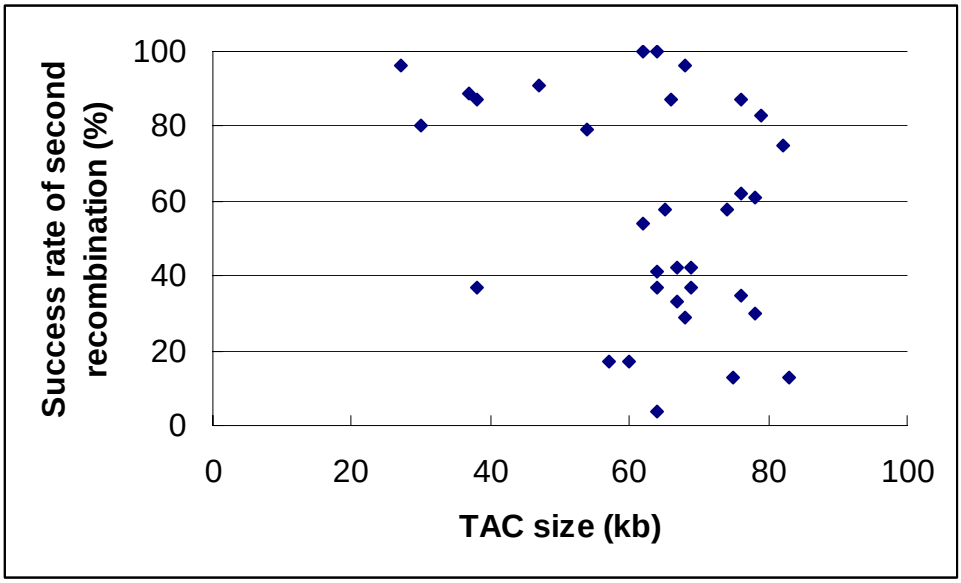
The first step of the process consists on identifying a suitable TAC clone for each gene of interest (GOI). A TAC clone containing the gene of interest among the end-sequenced *JAtY* clones could be found for 39 out of 40 GOI (no TAC containing the gene *At1g51950* (*IAA18*) was found). After this, the desired TAC clones were transferred from DH10B to SW102 *E. coli* strains. Using the procedure described in the methods section, every clone (ranging in size from 27 to 104 Kb) was easily transferred from DH10B to SW102 (~98% success rate) and latter into *Agrobacterium* after recombineering (~73% success rate). The apparent ~2% failure of transformation of TAC clone from DH10B to SW102 was not due to a problem in the trasformation per se, but to tracking problems of the *JAtY* library. This problem can be solved by selecting another TAC clone containing the GOI. In comparison, transformation from SW102 to *Agrobacterium* is more difficult and approximately 27%

colonies can survive from selection without harboring a TAC clone. The reason for these false positive colonies to escape the selection is unclear at this point. Nevertheless, it is clear that size of TAC is not related to the efficiency of transformation. Next the efficiency of inserting *galK* in the carboxyl terminus of each of the 34 genes in Table 1 was examined. Recombination efficiency was estimated by selecting between 16 to 32 $galK^+$ colonies per recombination and testing the presence of the *galK* gene by PCR analysis. In most experiments no false positives were found and almost 100% of the $galK^+$ colonies had the *galK* inserted in the desired location. The high stringency of the selection in the first recombination also suggest the possibility of using pools instead of single *JAtY* clones during the first recombination step. Only in some experiments low efficiencies in this first recombination step were found. This drop in efficiency was found to correlate with the age of the galactose selection media and the efficiencies went back up to a 100% as soon as new galactose selection media was used. Additionally, agar should also be tested to determine if $galK^-$ strains can grow in galactose supplemented minimum media.

A similar approach was employed to test the efficiency of replacing the *galK* gene by *GFP*. As it can be seen in Table 1 and Figure 8B, the efficiency of this step was much more variable, ranging from ~4% to 100%. Usually, low efficiencies were associated with suboptimal competent cells or poor quality of the recombination cassette DNA. No obvious correlation was observed between TAC size and the efficiency of second recombination.



A



B

Figure 8. Efficiency of recombineering process. (A) Average efficiency, the possible cause of failure and feasible solutions associated with each step are shown. (B) The correlation between the TAC size and the efficiency of second recombination.

In summary, by examining over 1200 recombination events using 41 different *JAtY* clones, we concluded that the experimental procedure describe here is very robust and provide an overall recombination efficiencies between 4% to 100%, with an average of ~55%. Moreover, as expected in a counter-selection approach, some of the false positives (*galK*$^-$) colonies correspond to deletions of the *galK* gene instead of its replacement by *GFP*. Although we did not sequence the *galK* gene, we speculate that the rest of the false positives with no obvious deletion would likely correspond to point mutations.

Fidelity of the process

In addition to the efficiency and robustness of the procedure, we also wanted to test its fidelity. In other words, whether or not the multiple transformation and recombination events will result in undesirable alterations in these large DNA clones, and whether the *GFP* had been precisely inserted in the desired location. To test the fidelity of the system, two different approaches were used. First, presence of local alteration at the recombination site was examined by sequencing the junctions between the *GFP* and the gene of interest, as well as the *GFP* itself. The sequence of the junction between the gene of interest and *GFP* indicates that the reporter gene can be directed to the TAC clone with nucleotide precision (Figure 9). Among 51 constructs made for 29 GOI, precise and flawless *GFP* insertions were confirmed in the first sequenced clone for 45 constructs (Table 1). In the 6 clones with mutations, 5 of them harbored a single nucleotide change in the region corresponding to the recombination primers, and were probably the result of mistakes incorporated during the synthesis of this long ~70 oligonucleotides. It is important to mention that although we found that “desalted” non-purified primers work quite well, it is clear that the quality of these long oligomers is critical and results can significantly vary depending on the manufacturer and the purification degree of the oligos.

Only one mutation was found in the *GFP* itself and was probably due to a mistake introduced by the proofreading polymerase during the *GFP* amplification. In all 6 cases where mutations were found, sequencing of an additional independent clone was sufficient to obtain a mistake free construct. This indicates that even when dealing with large number of genes it should be relatively easy to find precise modification events by simply sequencing one or two additional clones.

Second, to test for the presence of alterations in other parts of these large TAC clones, six *JAtY* clones with a wide distribution of insert size were selected (Table 2). DNA from several independent clones of each of the selected genes was fingerprinted after transformation from DH10B to SW102 (Figure 10). As it can be observed, no detectable alterations were observed in any of the 30 independent clones examined. Also, the same fingerprinting process was carried out after replacement of *galK* by *GFP* via the second recombination (Figure 11). In all 27 independent clones examined, no major DNA rearrangements associated with the recombineering process could be detected.

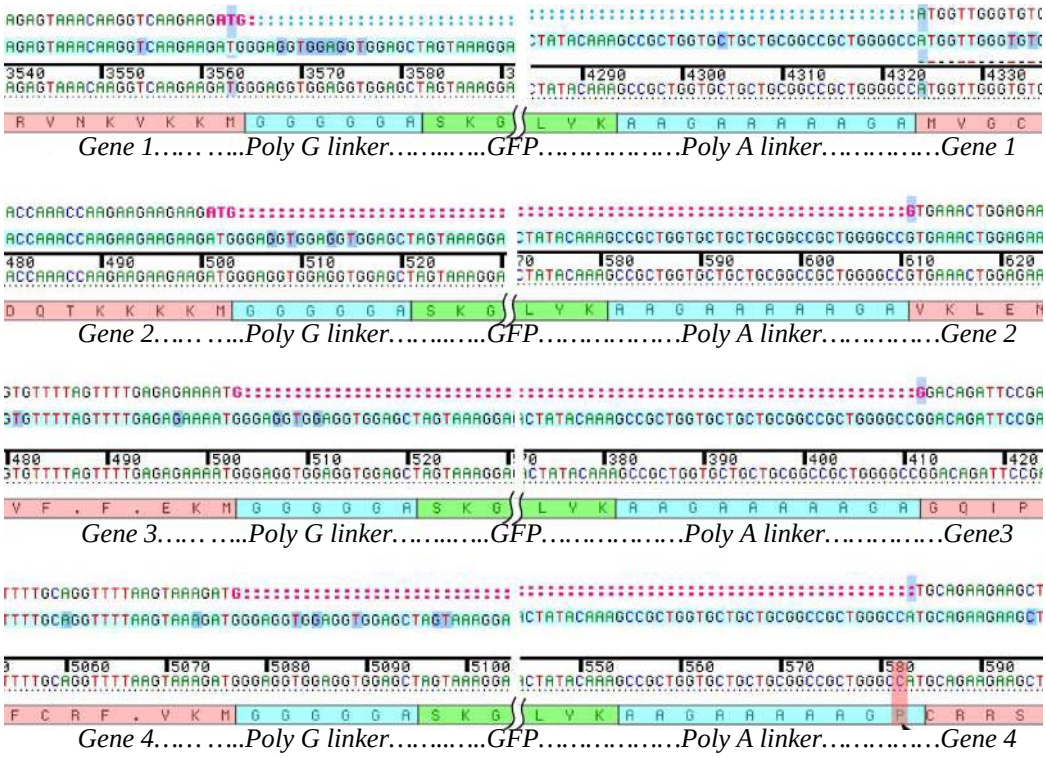


Figure 9. DNA sequencing results of representative recombination events. In all 4 cases, *GFP* was precisely inserted after start codon to produce a scarless chimeric gene.

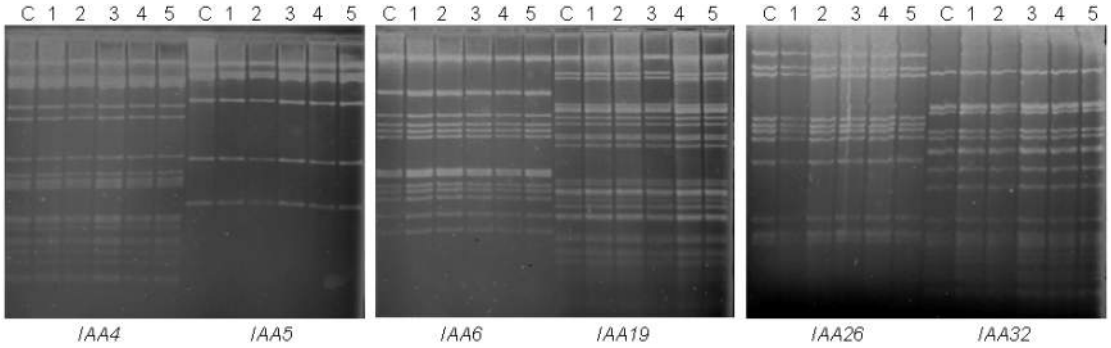


Figure 10. Integrity of TACs after transformation from DH10B to SW102. For each construct, DNA from 5 independent clones after transformation were digested with *HindIII* and examined by 1% agarose gel electrophoresis. C: The control lane shows the corresponding DH10B clones digested with *HindIII*.

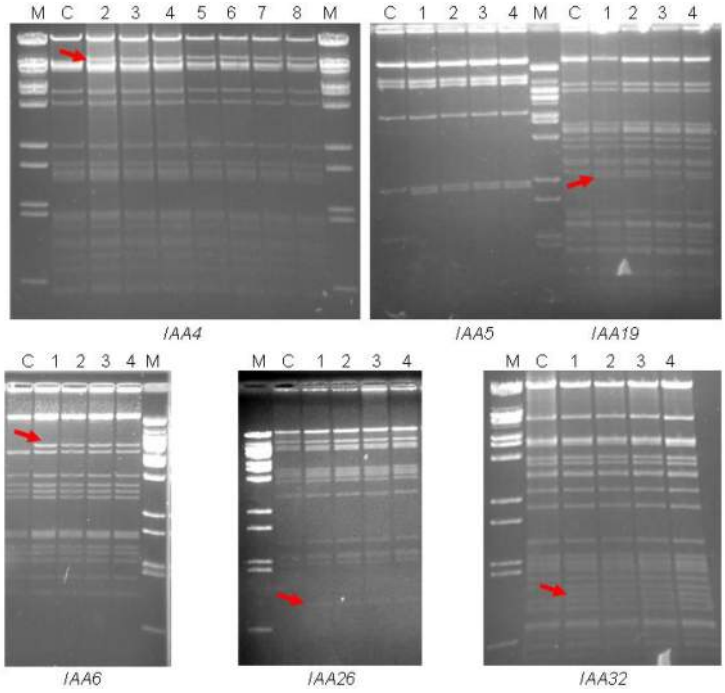


Figure 11. Integrity of the recombinant *JAtY* clones. DNA from 4 or 7 independent *GFP*-tagged clones for 6 GOI after second recombination were digested with *HindIII* and examined by 1% agarose gel electrophoresis. M: DNA ladder. C: the control lane corresponds to the untagged clone digested with *HindIII*. Arrows indicate the position of the bands that shifted due to insertion of *GFP*.

Stability of large *JAtY* clones in *Agrobacterium*

One important factor that could compromise the usefulness of the system is the stability of the *JAtY* clones in *Agrobacterium*. In a comprehensive study using potato BIBACs and TACs (Song et. al., 2003) the authors concluded that clones larger than 100 kb suffer frequent DNA losses after transfer to *Agrobacterium*. This "instability" of large clones was seen even when *rec*A$^-$ *Agrobacterium* strains were used. In a study more closely related to ours (Liu et. al., 1999), only one out of 35 randomly picked *Arabidopsis* TACs suffered a deletion when transferred from DH10B to *Agrobacterium*. It is important to point out that the size of the clones in this last study was not reported, and therefore, it was not possible to determine whether or not a direct correlation between size and instability in *Agrobacterium* exists. Since these studies reach different conclusions, it is necessary to further examine

the possible alterations of the *JAtY* clones during their transfer from *E. coli* into *Agrobacterium*. Six clones ranging in size from 27 to 79 kb were initially examined (Table 2). As it is shown in Figure 12, DNA alterations were not observed in any of these clones. To further explore the long term stability of large *JAtY* clones in *Agrobacterium*, three *JAtY* clones with relative larger size (104 Kb-127kb) were propagated for approximately 48 generations (~12 days). PCR using primers specific to GOI and *SacB* was conducted every three days to check the integrity of TAC (Figure 13) and after the last sampling, 3 independent clones per construct were fingerprinted (Figure 14). Notably, only in one case a single band change was observed in one of the clones (*JAtY50P23*) (Figure 14). It can be speculated that this was due to the incomplete digestion, or point mutation on restriction enzyme recognition site. These results suggest that clones containing over 100Kb TAC could be propagated in *Agrobacterium* for many generations with minimal detectable alteration. However, it cannot be concluded that all *JAtY* clones are stable in *Agrobacterium*. In fact, some *JAtY* clones were found to be difficult to transfer into *Agrobacterium*. Although we do not know what the reason is for this, it does not seem to be related to the size of the clone. Finally, it is important to mention that, in most cases, once these clones are transferred to *Agrobacterium* they can be maintained in these cells without suffering any apparent alterations (data not shown).

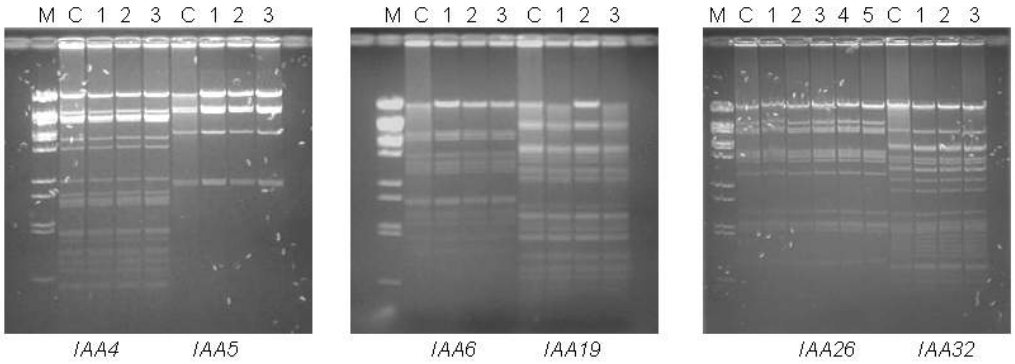


Figure 12. Integrity of TAC after transformation from SW102 to *Agrobacterium* strain C58 derivative GV3101 (pMP90). For each construct, TACs from 3 or 5 independent *Agrobacterium* transformants were purified and transferred back to *E. coli*. TACs were re-isolated after transformation, digested with *HindIII* and examined by 1% agarose gel electrophoresis. C: The control lane corresponds to the SW102 clones before transformation to *Agrobacterium*.

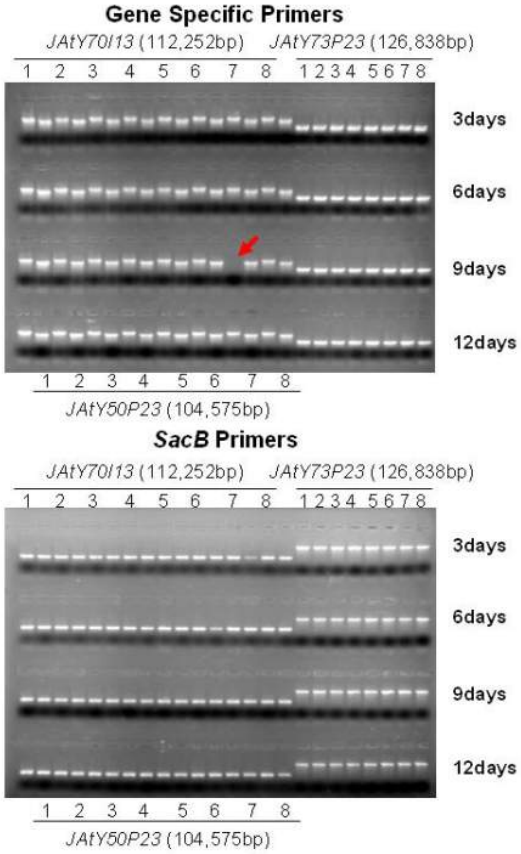


Figure 13. Stability of large TACs in *Agrobacterium*. *Agrobacterium* Strains containing *JAtY70I13* (112,252bp), *JAtY73P23* (126,838bp) and *JAtY50P23* (104,575bp) were inoculated in LB media for 12 days (cell culture were transferred to fresh media every 3 days, and an aliquot of cells streaked into LB Kan plates. For each construct, 8 independent clones were tested by PCR using gene specific primers (top part) and primers for *SacB* (bottom part). Arrow shows a missing band due to failure of the initial PCR amplification. The corresponding clone was retested and confirmed to contain the GOI.

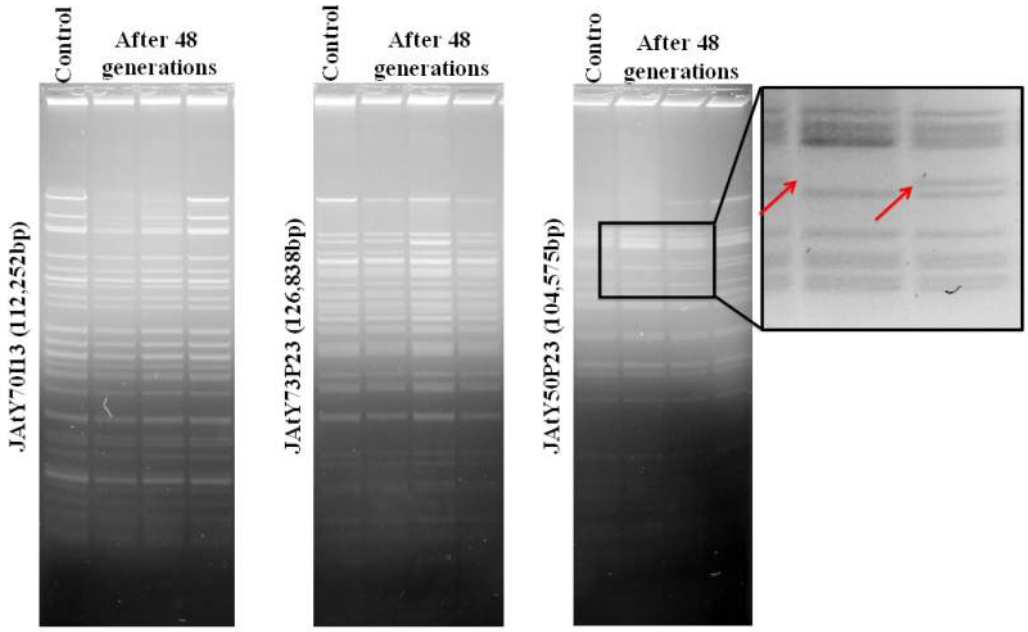


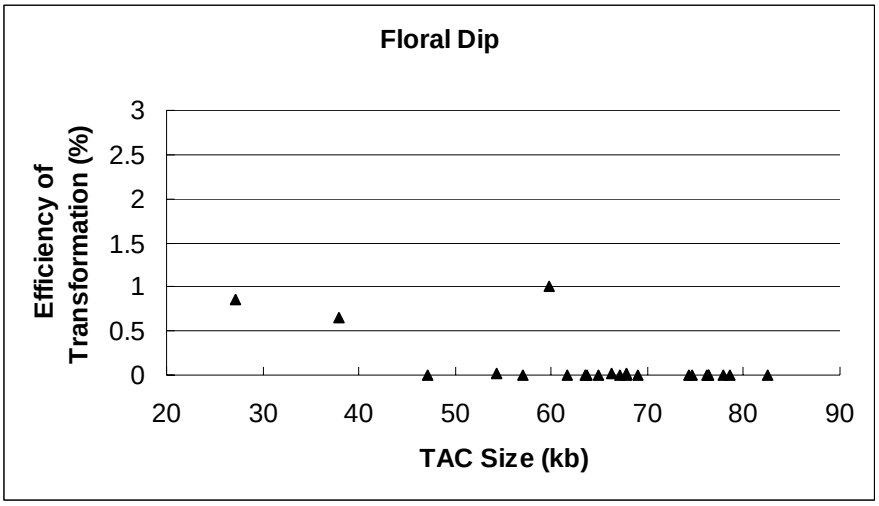
Figure 14. Long-term stability of TAC clones in *Agrobacterium*. For each construct, TACs from 3 independent *Agrobacterium* strains grown for 12 days were purified and transferred back to *E. coli*. TACs were re-isolated from transformants, digested with *HindIII* and examined by 1% agarose gel electrophoresis. Control: the *Agrobacterium* clones before the inoculation. Arrow shows the single band missing in one of the *JAtY50P23* clones.

Arabidopsis transformation with *JAtY* clones

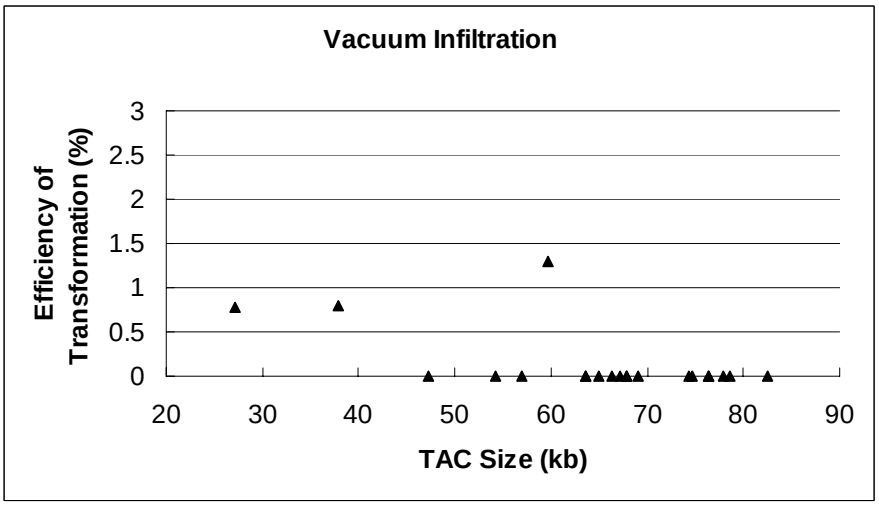
Transfer of the genomic DNA from the *JAtY* clones into the *Arabidopsis* genome is obviously a critical step in the process. Initially, the transformation efficiency of 25 *JAtY* clones was examined using standard *Agrobacterium*-mediated floral dip transformation (Clough et. al., 1998). As it can be seen in Figure 15A, the efficiency varies quite substantially from clone to clone. These low, and very variable transformation efficiencies were quite surprising since consistently high efficiencies of transformations using a very similar TAC library has been previously reported (Liu et. al., 1999). Two of the possible reasons for this discrepancy were, the vector used (*pYLTAC17* vs. *pYLTAC7*), and the transformation method (floral dip vs. vacuum). However, neither the use of vacuum infiltration (Figure 15B), nor the use of *pYLTAC7* derived clones such as *K2I5* resulted in

any significant improvement (data now shown). It is important to mention that these results were very reproducible in three independent transformation experiments, and thus, clones with low transformation efficiency will always produce poor results (data now shown), while clones with higher efficiencies produce consistently better results. Since the only difference between different *JAtY* clones is the genomic DNA that they carry, it was plausible that some property of this DNA was responsible for the observed variability. The obvious candidate, the size of the clone, was unlikely to be the culprit since we did not observe a correlation between efficiency of transformation and clone size (Figure 15A-B). Looking into the vector used to generate the *JAtY* library, we realized that in this library, the *Arabidopsis* genomic DNA fragments would be immediately adjacent to the *SacB* coding region present in the vector *pYLTAC17*. Thus, it was possible that, some of these *Arabidopsis* sequences could activate the expression of *SacB* in *Agrobacterium*. Activation of *SacB* would be very detrimental for an *Agrobacterium* cell in the transformation media that contains high sucrose concentrations. To test this possibility, we examine the effects of 5% sucrose (the concentration typically use in the transformation media (Clough et. al., 1998) on the growth of *Agrobacterium* harboring either clones that we had successfully transferred to plants or those for which we had failed. As it can be seen in Figure 16, there was a perfect correlation between the efficiency of transformation and the ability of the *Agrobacterium* to grow in the presence of 5% sucrose. Since it has been shown that sucrose can be substituted by glucose in the infiltration media without any adverse effects on the transformation efficiency (Clough et. al., 1998), and glucose is not a substrate for *SacB*, we decide to test the plant transformation efficiency when 5% glucose was used in the transformation media. As it can be seen in Table 1, Table 2 and Figure 17A-B, all clones could be readily transformed using glucose. Although we did not test it, it is possible that

similar improvement could be obtained by reducing the amount of sucrose in the infiltration media. In fact, Liu et. al. used 1% sucrose instead of the typical 5% giving a possible explanation to the discrepancy between theirs and our initial results. However, using 5% glucose is highly recommended since it is possible that even 1% of sucrose could have detrimental effects on *Agrobacterium* clones expressing high levels of *SacB*. Furthermore, from the results shown in Figure 18, we can also conclude that the transformation efficiency of these large clones is not greatly affected by the *Arabidopsis* ecotype (Columbia or Wassilewskija, WS), transformation methods (dip floral or vacuum infiltration) and *Agrobacterium* growth (on a plate or in liquid media).



A



B

Figure 15. Efficiency of plant transformation with TAC clones following traditional transformation protocol (5% sucrose infiltration media). A: floral dip. B: vacuum infiltration. Efficiency was calculated as percentage of basta resistant lines in 8,000 T1 seeds tested.

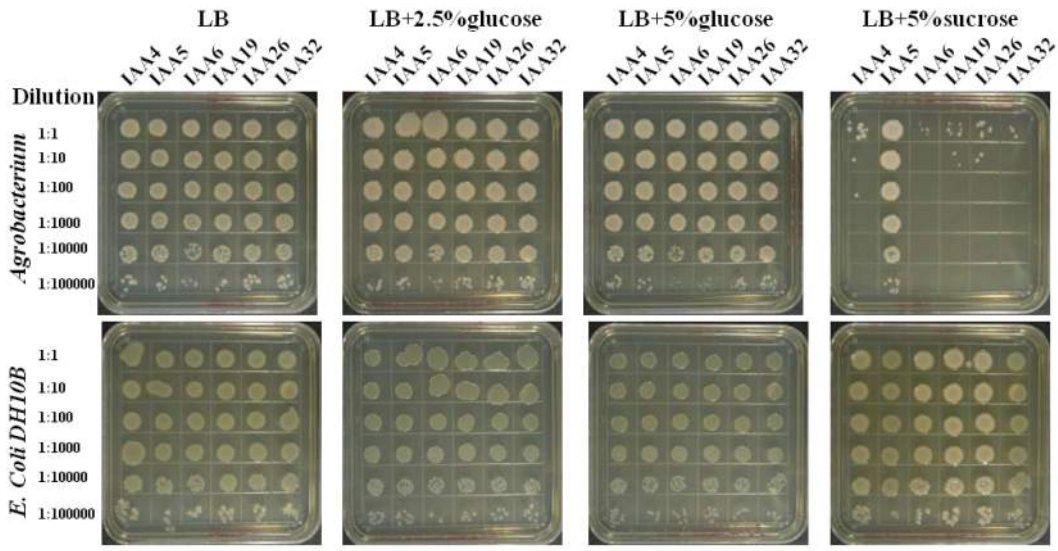
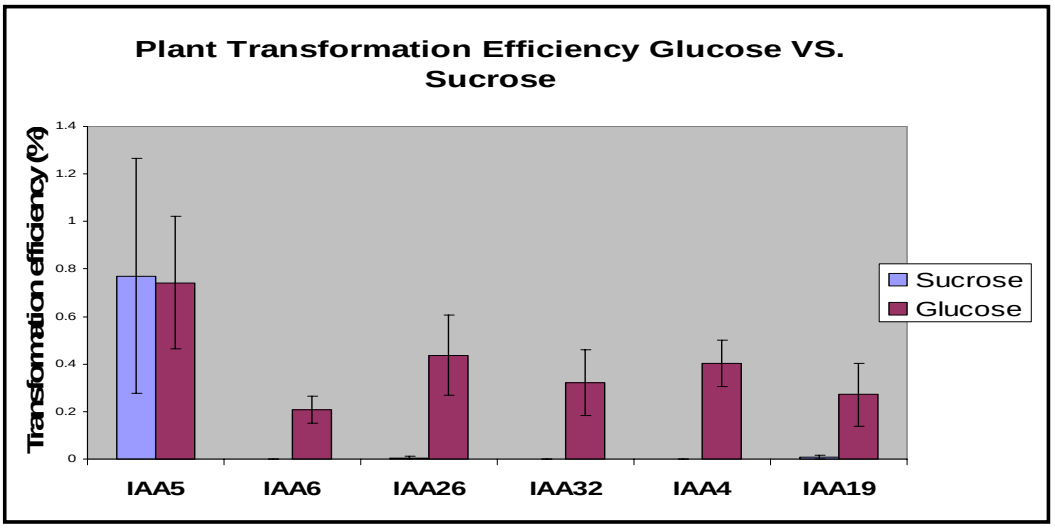
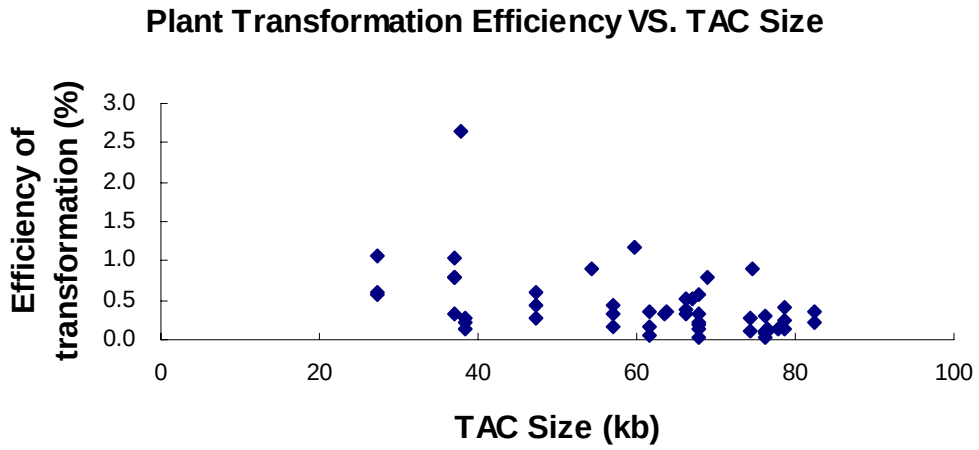


Figure 16. Effects of 5% sucrose on the growth of *Agrobacterium*. *Agrobacterium* strains (top part) and corresponding *E. coli* DH10B strains (bottom part) corresponding to 6 TAC clones (Table 2) were grown in LB plates, LB plates supplemented with 2.5% glucose, 5% glucose and 5% sucrose. The results from 10 fold serial dilutions are shown.

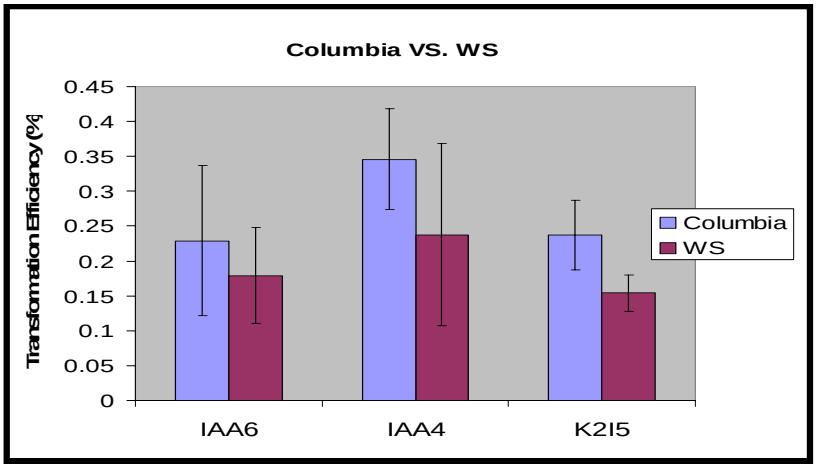


A

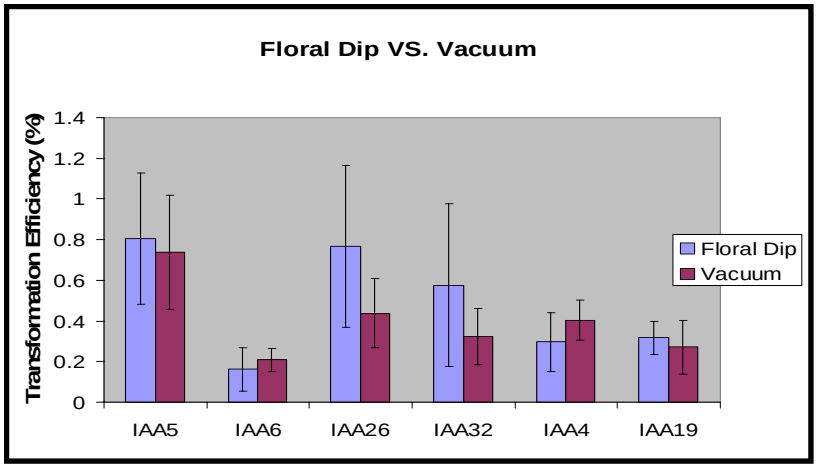


B

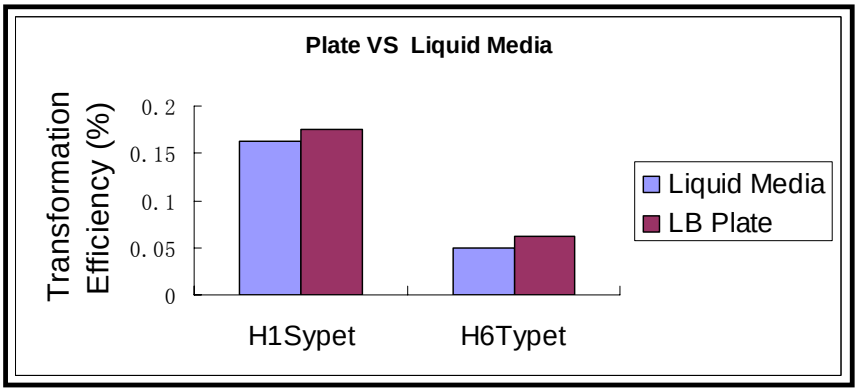
Figure 17. Efficiency of plant transformation with TAC clones following modified floral dip transformation protocol (5% glucose supplied in infiltration media). A: comparison of transformation efficiency of 6 TAC clones using sucrose or glucose. Each transformation was repeated 3 times and graph displays the average and standard error of the efficiency. The name of TAC clones are displayed on the horizontal axes of the bottom panels. B: correlation between transformation efficiency using glucose in the infiltration media and TAC size. Each data point represents a single transformation experiment.



A



B



C

Figure 18. Factors affecting transformation efficiency. A: Effect of *Arabidopsis* ecotype (Columbia or WS), B: transformation methods (floral dip or vacuum) and C: *Agrobacterium* growth media (LB plates or in liquid LB media) on plant transformation efficiency. Efficiency was calculated as percentage of basta resistant lines for each 8,000 T1 seeds tested. The name of TAC clones are displayed on the horizontal axes of the bottom panels. *K2I5: a pYLTAC7* derived clone with 87 Kb *Arabidopsis* genomic DNA insertion. Each transformation in A and B was repeated 3 times and graphs display the average and standard error of the efficiency. Transformation in C was performed once.

Integrity of the T-DNA during plant transformation

Although the percentage of herbicide-resistant plants is usually taken as a direct measurement of the rates of successful transformation events (i.e., whether the whole T-DNA is inserted in the plant genome), it may not be applicable when using large clones. In fact, previous studies have shown that large deletions of the T-DNA can occur during the plant transformation process (Liu et. al., 1999 and Hamilton et. al., 1996). To characterize this phenomena in more detail, and to determine the frequencies of deletions and the possible effect of the clone size, over 500 T1 (or T2 pools) from 6 *JAtY* clones ranging in size from 27 to 79 Kb were examined (Table 2, Figure 19). In each case the presence of the *SacB* and the tagged gene were tested by PCR. Results clearly indicate that deletions of the T-DNA do occur during the process of integration of the T-DNA into the plant genome (Figure 19A). We differentiated three types of deletions, deletions where only the *SacB* has been lost but the tagged gene is still present (category B), deletions in which both the *SacB* and the tagged gene have been lost (category D), and category C where the tagged gene could not be detected, but still retain the *SacB* (Figure 19A, category B, C and D). Average integrity of the T-DNA was maintained in 70% T1 lines selected from Basta plates (Figure 19B). The proportion of T1 lines containing intact T-DNA after plant transformation varied from clone to clone (Figure 19C). Importantly, we observed a direct correlation between the percentage of deletions and clone size of the T-DNA (Figure 19 D). In one case, however, a relatively small clone (IAA6) showed a lot more frequent deletions than what would have been expected for a clone of that size. Interestingly, most of the deletions in this clone belong to the category D where both *SacB* and the tagged genes were missing. One possible explanation for this is that IAA6 gene is relatively close to the end of the TAC clone (Figure 19E), and thus even small deletions could affect this gene. This,

however, does not explain the large number of deletions observed for this clone. Thus, although we could not identify any obvious sequence peculiarity in this clone such as tandem repeats, it is plausible that some specific sequences may favor deletions of the T-DNA during the transformation process. Nevertheless, and from a practical point of view, the overall results from these experiments indicate that identification of 10-20 basta resistant T1s should be sufficient to obtain a few (5 to 10) transgenic lines containing an "intact" T-DNA.

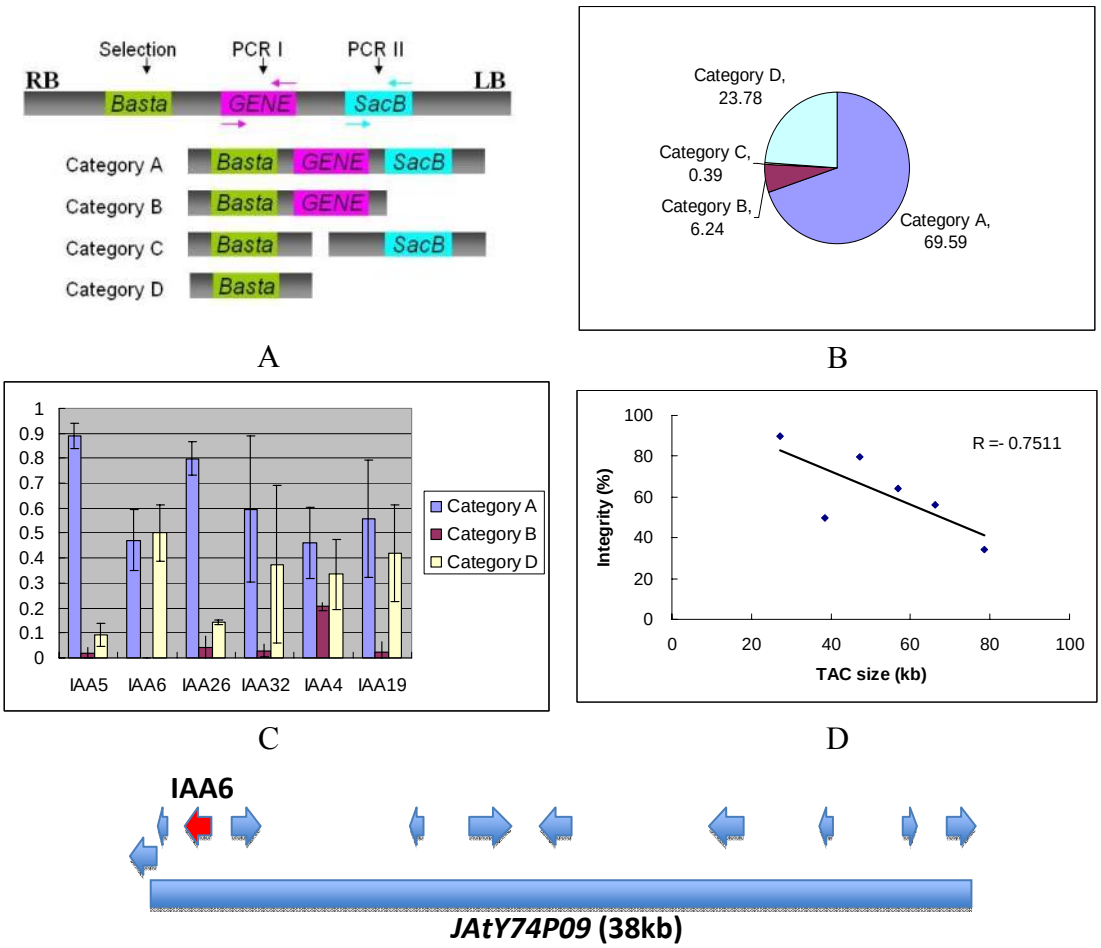


Figure 19. Integrity of T-DNA after transformation of plants. (A) Categorization of transgenic lines based on basta selection and 2 sets of PCR (targeting GOI and *SacB* respectively). Category A: T-DNA is intact. Category B: the T-DNA is truncated near the left border and *SacB* gene is missing. Category C: T-DNA has internal truncation with GOI missing. Category D: most part of T-DNA is missing except the selectable marker. (B) Pie chart of the distribution of all categories. (C) Proportion of category A, B, D in transgenic lines transformed with each one of 6 TAC clones. Transgenic lines from each of 3 independent transformations were classified in one of the four categories described above. The graph displays the average and standard error from the three experiments. (D) Correlation between the TAC size and integrity of plant transformation with TAC clones (percentage of Category A). (E) Structure of *JAtY* clone *JAtY74P09* containing gene *IAA6.* Arrows indicate the position and direction of transcription of all the *Arabidopsis* genes contained in this clone.

Table 2. Detailed information on efficiency and fidelity of each recombineering step of 6 well-characterised *JAtY* clones. Success rate: percentage of PCR confirmed independent colonies in all tested colonies. Verified clones: number of clones checked by fingerprint.

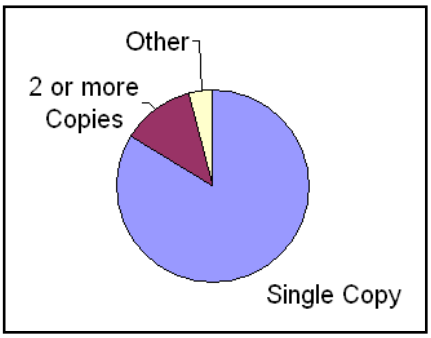
Name	BAC Size (bp)	SW102 Transformation		*galK* Insertion	*GFP* Insertion		
		Success Rate	Verified Clones	Success Rate	Success Rate	Verified Clones	Sequenced Clones
IAA5	27204	100	5	100	95.65	4	1
IAA6	38393	100	5	100	87.50	4	1
IAA26	47225	100	5	100	91.30	4	1
IAA4	66258	100	5	100	86.96	7	1
IAA19	78622	100	5	100	82.61	4	1
IAA32	56967	100	5	100	16.67	4	1

Name	*Agrobacterium* Transformation		Plant Transformation Efficiency		Plant Transformation Integrity		
	Success Rate	Verified Clones	Sucrose	Glucose	Category A	Category B	Category D
IAA5	50.00	3	0.77%	0.74%	90%	2%	8%
IAA6	50.00	3	0	0.21%	51%	0	49%
IAA26	83.33	5	0	0.44%	80%	3%	16%
IAA4	87.50	3	0	0.32%	56%	25%	18%
IAA19	100.00	3	0	0.40%	64%	3%	33%
IAA32	66.70	3	0.01%	0.27%	34%	2%	64%

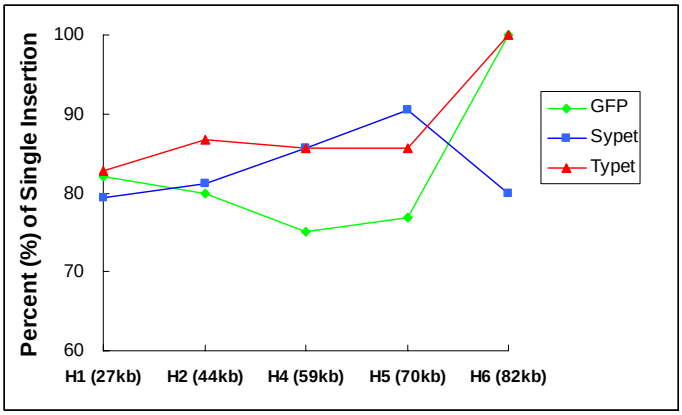
Although this PCR based strategy has been previously used to test the integrity of the T-DNA during the plant transformation process (Liu et. al., 1999), it has some potential problems. For example, a T-DNA insertion could be classified as intact, when in reality it corresponds to two complementary truncations of a T-DNA. To address this potential problem, we examined the copy number of three genes positioned in the left, center and right of the T-DNA. At the same time, we also determined the number of unlinked insertions per line using the segregation of the resistance to Basta.

First, the segregation of the basta resistance among the progenies of over 230 T1 plants generated was examined using 5 different *JAtY* clones ranging in size from 27 to 82 kb

(Table 1, H1-H2 and H4-H6). According to this data (Figure 20A), ~78% of the lines show single insertion site, while ~13% had more than one insertion site and ~8% of the lines show segregations <1:3 that would be consistent with a reduced transmission of the T-DNA. This initial analysis indicates that the TAC size is not correlated with the proportion of single site insertions (Figure 20B). Notably, the percentage of single insertion lines reported here is much higher than that present in the SalK T-DNA collection, where only 50% of the lines show single insertion. These differences may be due to a lower frequency of insertions when using large TAC clones compared instead of the relatively small T-DNA of pROK2 used in the generation of the Salk collection. In any case, the predominance of single locus insertions may be an additional advantage of using the *JAtY* clones.



A



B

Figure 20. Copy number of T-DNA insertion determined by segregation. Copy number for 244 independent T2 lines from 15 constructs were determined by calculate segregation rate. (A) Pie chart showing distribution of categories of transgenic lines based on segregation. Other: transgenic lines showing segregations <1:3. (B) Correlation between the percentage of single copy T-DNA lines and the size of TAC clones. Graph displays the percentage of transgenic T1 lines with single-copy T-DNA insertion for each of the 15 constructs (data point connected by lines). The GOI targeted by each construct and corresponding size of TAC clone are displayed on the horizontal axes of the bottom panels.

The segregation of the basta resistance provide very useful information on the number of unlinked insertion sites, however it does not tell us about the total number of T-DNA copies inserted in the genome. To determine the T-DNA copy number and also get a better idea of the integrity of these T-DNAs, the copy number of the T-DNA was further examined by quantitative real-time PCR. Ten independent lines of the construct *H4-GFP* were examined. In *H4-GFP*, the *GFP* was inserted just before the stop codon of the gene At3g56240 contained in the 59 kb TAC clone *JAtY49O07*. In addition to the tagged gene (At3g56240) we also examined the copy number of the first (At3g56140) and last

(At3g56310) gene in this TAC clone. Consistent with previous observations (Bubner et. al., 2004), the variability between replicates was larger than that typically observed in quantitative RT-PCR experiments, preventing us from discriminating differences in copy number smaller that 2 fold. In spite of these limitations, the results obtained (Table 3) strongly indicate that transformation with these large TAC clones can result in multiple linked T-DNA insertions in one or several loci as illustrated by lines *H4GFP-15* and *H4GFP-7*. However, not in all cases these linked T-DNA are complete and large deletions can be observed as illustrated in lines *H4GFP-5* and *H4GFP-10*. Real-time PCR data also largely back up the results of regular PCR used to determine the integrity of T-DNA after plant transformation. For example, for the line *H4GFP-14* classified into B category, real-time PCR confirmed that T-DNA is truncated close to the left border since there was no amplification of At3g56140 gene. Finally, it important to point out that in all cases the copy number of the genes close to the right border (RB) was equal or higher than that of the gene close to the left border (LB). This strongly suggests that the deletions predominantly happen at the left border side of the T-DNA as it has been previously suggested (Forsbach et. al., 2003). This has important practical implications since the presence of the *SacB* gene close to the LB will be indicative of the presence of at least one whole copy of the T-DNA in the plant genome. In fact, we have tested more than 500 independent T1 plants (or pools of the corresponding T2) and found that in all but two cases, the presence of the *SacB* was predictive for the presence of the *GFP*-tagged gene located in the middle of the TAC clone.

Table 3. Copy number determined by real-time PCR. DNA extracted from 9 independent *H4-GFP* T2 pools was amplified with primers specific for gene L (*At3g56140*, close to left border), G (*At3g56240*, GOI) and R (*At3g56310*, close to right border). A gene known to have two copies in *Arabidopsis* genome was used as control to normalize the transgenic lines and wildtype Columbia line. Genotype: indicates the integrity of the T-DNA where A indicates that the T-DNA is intact, and B that the *SacB* gene is mising. Fluorescence: fluorescent intensity was classified as L: low; H, High; M, medium; NF: no fluorescence observed. Segregation: copy number of T-DNA determined by checking the Basta resistance segregation rate in the T1 progenies. The real-time PCR experiments involve was performed using 2 biological (Exp1 and Exp2) and 3 technical replicates (average is shown in the table).

Line	**Genotype**	**Fluorescence**	**Segregation**	**Copy Number Based on Real-time PCR**					
				L	**G**	**R**	**L**	**G**	**R**
				Exp1	Exp1	Exp1	Exp2	Exp2	Exp2
H4GFP-5	A	L	**1**	1.2	1	4.8	1.2	1.3	5.7
H4GFP-6	A	L	**2**	0.9	1.6	1.5	1.2	1.9	2.4
H4GFP-7	A	M	**2**	4.3	5.5	3.6	4.3	6.7	4.8
H4GFP-8	A	NF	**1**	0.9	0.9	1.9	1.0	1.0	2.4
H4GFP-10	A	NF	**1**	1	1	3.3	1.2	1.3	3.1
H4GFP-13	A	L	**1**	1.7	0.3	1.5	2.2	1.9	2.1
H4GFP-14	B	L	**1**	-0.4	1.8	2.5	0.1	2.0	3.2
H4GFP-15	A	H	**1**	6	7.6	7.5	8.0	10.2	11.2
H4GFP-19	A	M	**1**	0.5	1.3	0.7	0.9	1.4	0.9

Variability of the expression patterns among independent transgenic lines

Although the presence of detectable fluorescence will obviously depend on the expression levels of the tagged gene, even when a gene is highly expressed not all the lines are expected to show fluorescence. To estimate the percentage of lines that are expected to be informative on a given gene, the GFP fluorescence pattern of over 1500 T2 seeding from >350 independent T1 lines generated using 5 different TAC clones ranging in size from 27 to 82 kb (Table 1, H1-H2, H4-H6) was examined by microscropy. As it was shown in Figure 21, when we exclude the lines corresponding to the *H2-GFP* construct for which we could not detect any signal, fluorescence can be observed for more than 70% of independent PCR confirmed transgenic T2 families (each family correspond to an individual T1 line). Based on microarray data available (Stepanova et. al., 2005), the only

undetectable construct *H2-GFP* correspond to a gene with very low expression level. Nevertheless, fluorescence for this gene could be observed from *H2-Sypet* and *H2-Typet* when improved versions of Ypet were used (more details in Chapter 2).

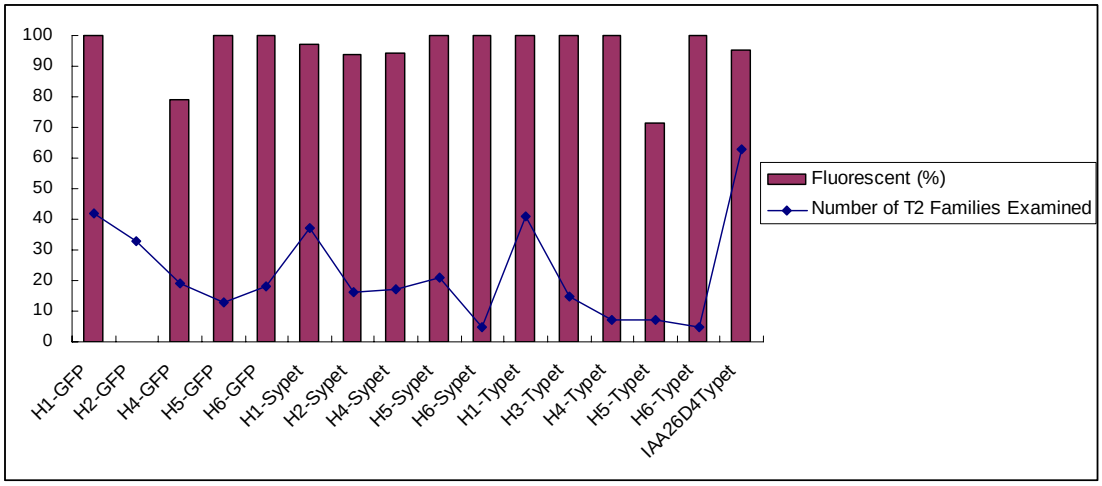


Figure 21. Percentage of fluorescent T2 families. The fluorescence of T2 families was determined by examining 10 individual T2 seedlings of each family using fluorescent microscopy. Only families from categories A (intact) or B (only the *SacB* gene is missing) were used in this analysis. Dots indicate the number of T2 families examined for each of 16 constructs generated with recombineering pipeline (359 total). . The names of constructs are displayed on the horizontal axes of the bottom panels.

Another typical concern when generating translational fusions is the reproducibility of the observed expression patterns among independent transgenic lines, or what is typically called "the positional effect". This variability is thought to result from the presence of strong transcriptional enhancers in the vicinity of the T-DNA insertion site that can effect the expression of the transgene. Obviously, one additional advantage of using large TACs is that the effects of these transcriptional enhancers over the expression of the tagged gene should be buffered by tens of thousands of base pairs flanking the gene of interest.

As it is shown in Figure 22, each one of the five tagged genes has a distinctive expression pattern. Importantly, no alterations in the expression pattern or sub-cellular localization were observed among different versions of fluorescence proteins, even when three copies

of Ypet, an improved YFP (reducing ability to move from cell to cell), were used. Moreover, the expression pattern of each gene was consistently observed in all fluorescent T1 lines obtained for that particular gene. These results have obvious practical implications, as few lines would be needed to establish the expression pattern of the tagged gene.

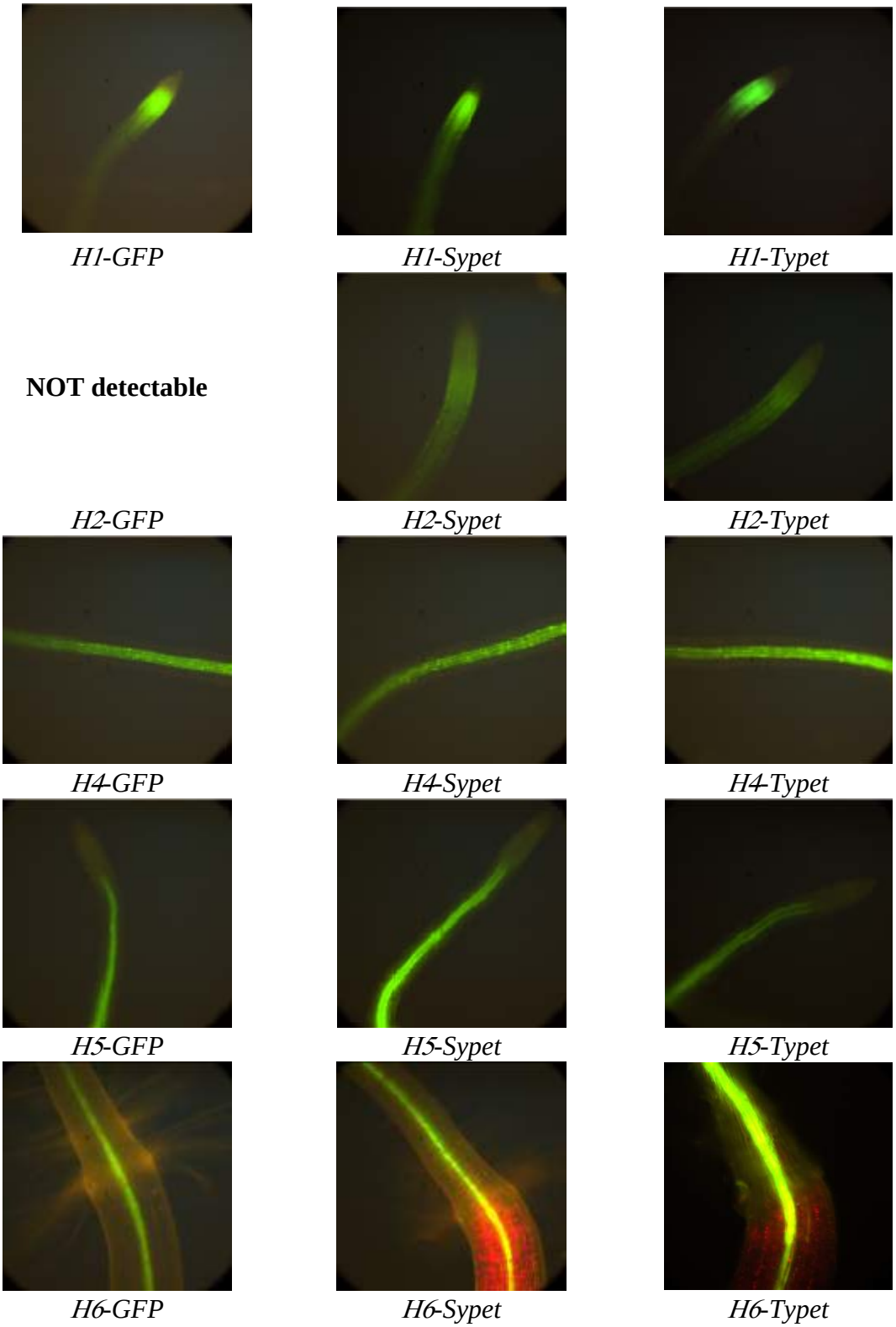


Figure 22. Expression pattern of tagged gene in *Arabidopsis* root. 5 representative genes (H1, H2, H4, H5 and H6, see also Table 1) were tagged with 3 fluorescence proteins (GFP, Sypet and Typet) through recombineering pipeline. Fluorescent microscopy imagines were taken from representative T2 seedlings for each construct.

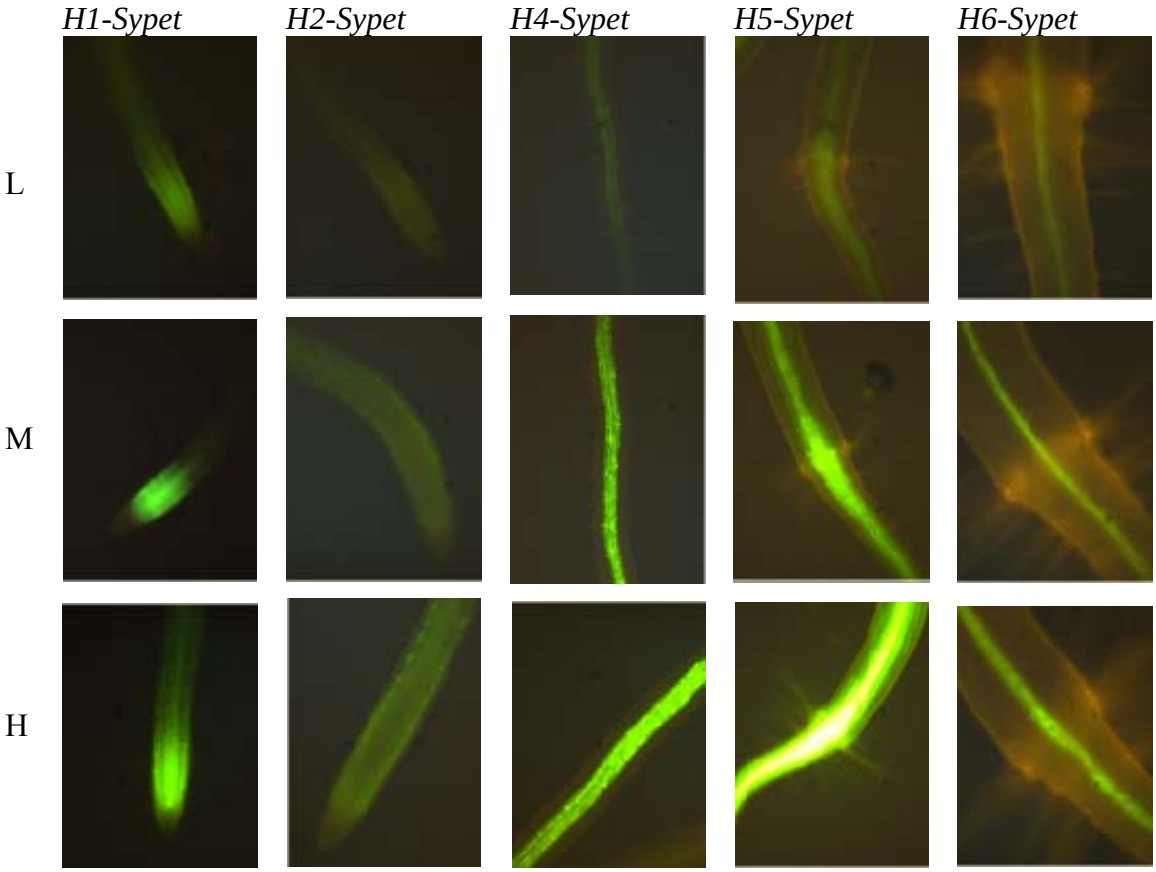


Figure 23. Fluorescent microscopy images showing variable fluorescence intensity. For each of five *Sypet*-tagged constructs (H1, H2, H4, H5 and H6), independent T2 lines were categorized based on average intensity of observed fluorescence from 10 seedlings. L: low fluorescent intensity; M: medium intensity; H: high intensity. Representative images for each category of all constructs are displayed. Images were taken using exactly the same settings.

Variable fluorescent strength was observed across independent T2 lines for each construct (Figure 23 and Figure 24). The observed inter-line variability could be caused by the copy number of the T-DNA or by the positional effects which may includes chromatin structural or the presence of enhancer elements in the vicinity of the insertion site. As a first approximation to test these possibilities, we examine the possible correlation between fluorescence intensity and copy number. As shown in Figure 25, there is no clear correlation between the number of T-DNAs and the fluorescence intensity. However, these results are based on segregation analysis, and it has been shown before that some

lines that seem to have a single insertion based on segregation of Basta resistance, have actually several T-DNA copies as indicated by real-time PCR. Using the lines tested by real-time PCR (Table 3), lines with high copy number tend to show also high levels of fluorescence. Hence, this data suggest that the variability on the fluorescence intensity may by, at least in part, due to the T-DNA copy number.

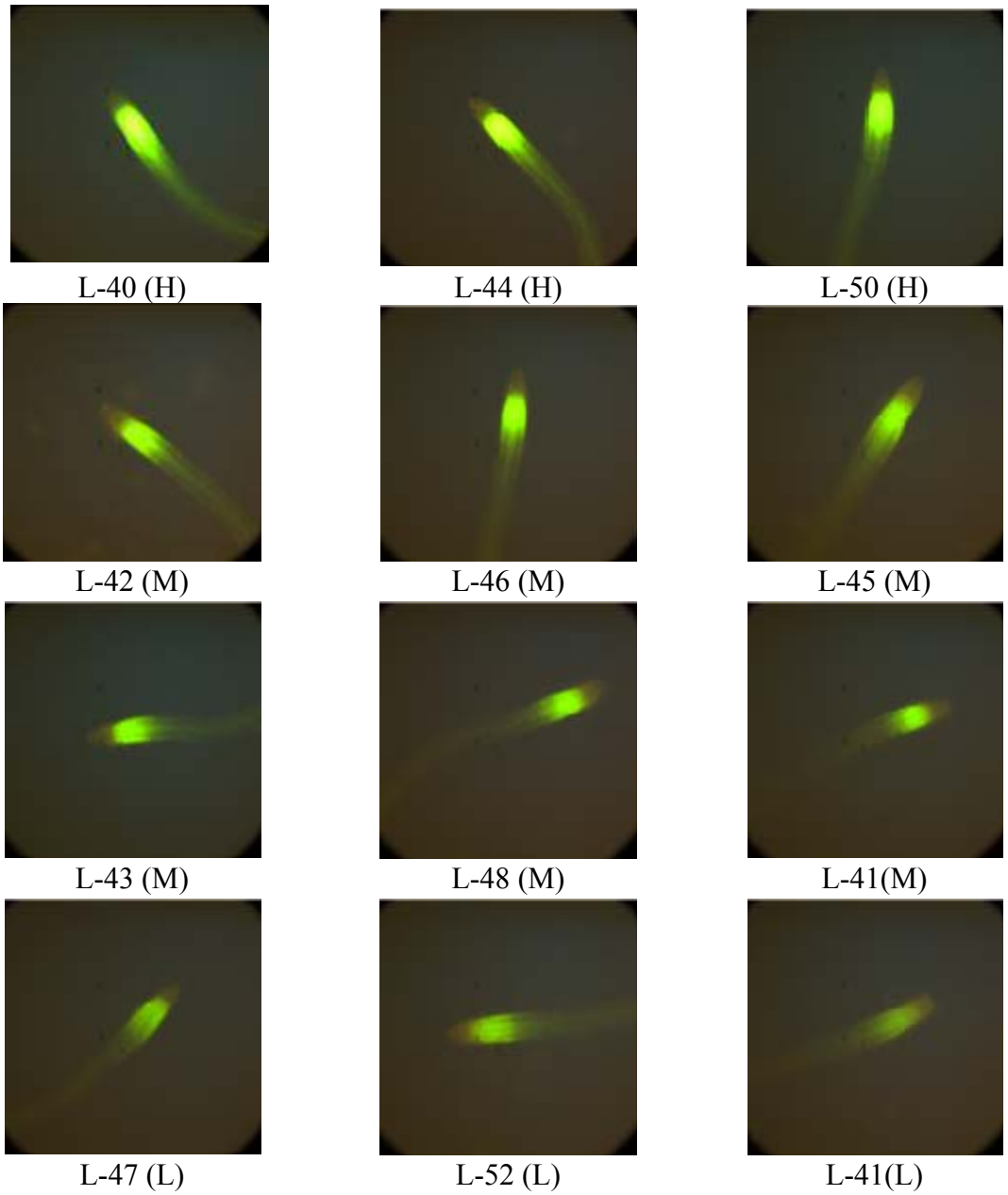


Figure 24. Fluorescent microscopy images of 12 independent T2 lines from *H1-Sypet* construct. Independent T2 lines were categorized based on average intensity of observed fluorescence from 10 seedlings. Representative images for each T2 line are displayed. The ID of each line and corresponding fluorescent strength are displayed on the bottom panels. H: high intensity. L: low intensity. M: medium intensity.

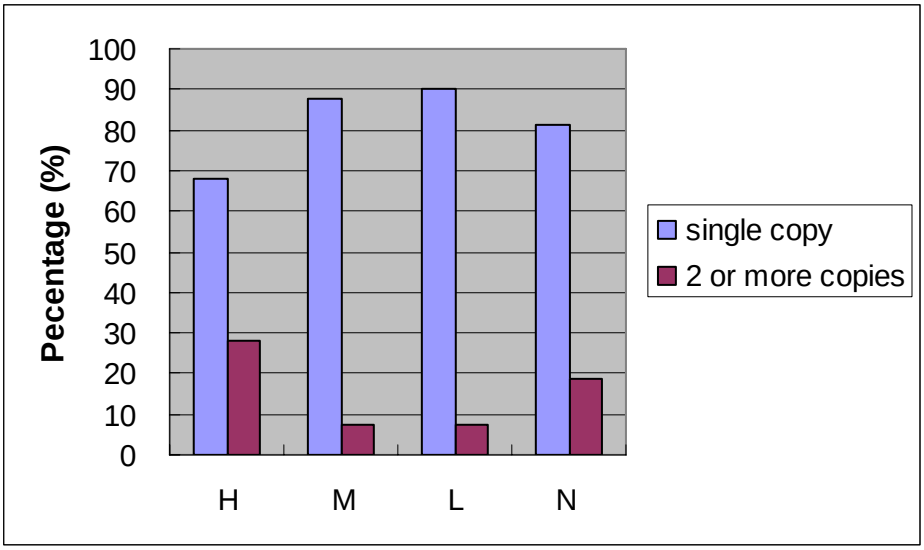


Figure 25. Percentage of T-DNA copy number across different categories of fluorescent strength. Copy number of each independent T2 line (230 lines, all from category A and B based on integrity) was determined by segregation rate. H: high intensity; M: medium intensity; L: low intensity; N: no detectable fluorescence.

Sub-cellular localization of tagged proteins

To further evaluate the feasibility and accuracy of the system to determine sub-celluar localization of tagged fusion proteins, representative lines were imaged using confocal microscopy at both lower (20×, Figure 26) and higher magnification (40×, Figure 27). Notably, distinct and reproducible subcellular localization of each tagged-protein was observed regardless of tag used.

Gene AT1G22780 (H1) is predicted to be a structural constituent of the ribosome involved in the binding of f-Met tRNA during initiation of mRNA translation (Lijsebettens, et. al., 1994). Our results clearly show that this gene is expressed in the root meristem and its protein localized in the cytoplasm as predicted (Chang, et. al. 2005). At3g53420 (H5), corresponds to the plasma membrane intrinsic protein 2A (*PIP2A*), a channel protein (Cutler et. al., 2000) shown to localize to the plasma membrane (Tian et. al., 2004; Laxmi et. al., 2008). Our results are consistent with these previous studies

(Figure X). At1g66240 (H2) is annotated as a heavy metal transporter and detoxification protein (Schmid et. al., 2005) predicted to localize in endomembrane systems. Our obervations are consitent with this sub-cellular localization.

At3g18280 (H6), a plant lipid transfer protein, is predicted to localize to endomembrane systems (Schmid et. al., 2005). In our tagged lines, this gene was expressed in the root vasculature. Due to the very internal position of these cells, their small size, and the relatively low expression levels (low fluorescence intensity) we were unable to determine its subcellular localization. At3g56240 (H4) is a small copper binding protein involved in copper trafficking (Himelblau, et. al., 1998). The localization of this protein is controversial. Both chloroplast localization (Zybailov, et. al., 2008) and cytoplasm (Mira et. al., 2001) have been reported. Our results however, suggest CCH protein can be found in both the cytoplasm and the nucleus.

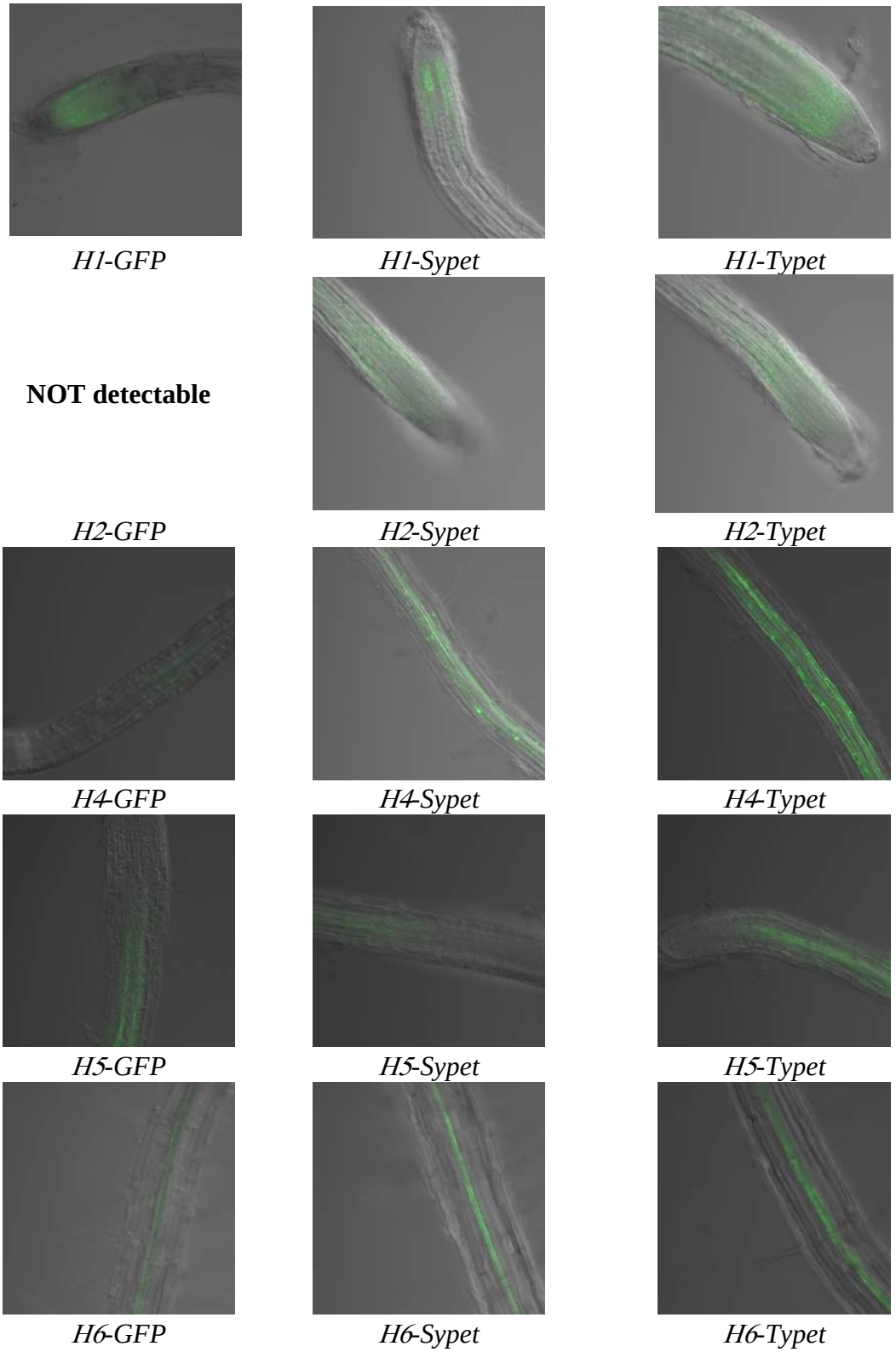
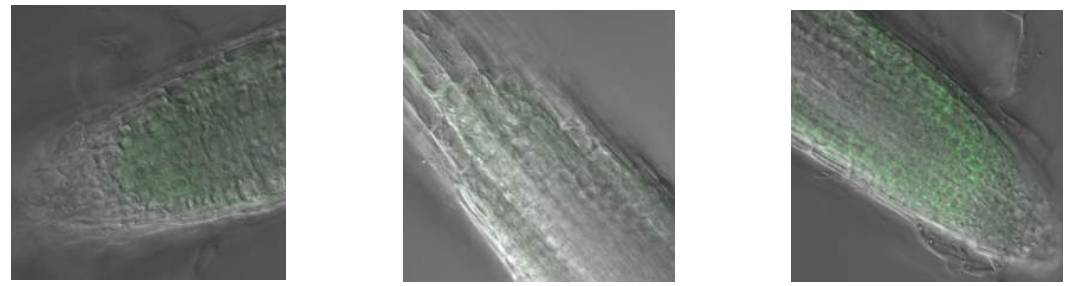
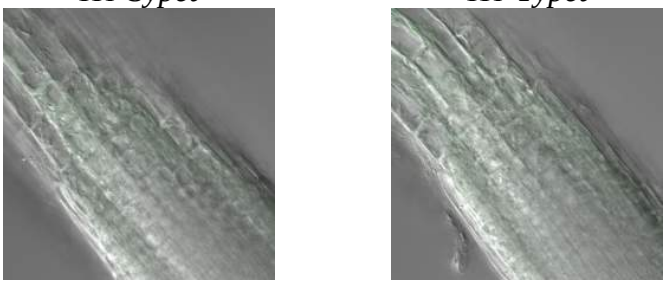


Figure 26. Confocal images with lower magnitude. For each construct, image was taken from one representative T2 line with higher intensity. The names of constructs are displayed on the bottom panels. Overlay images of fluorescence and interference contrast microscopy (DIC) are shown.

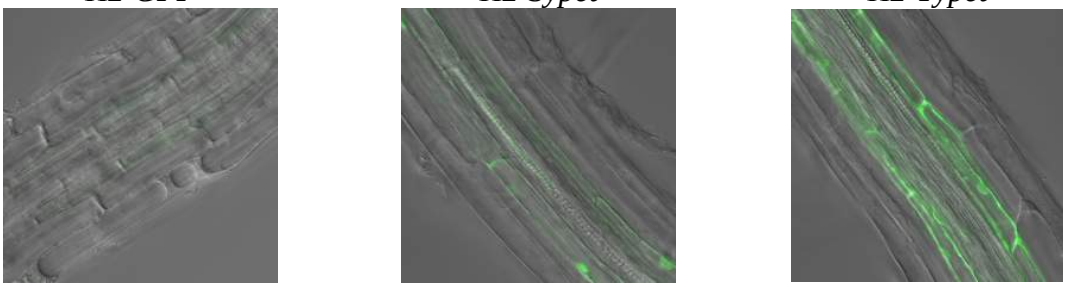


H1-GFP *H1-Sypet* *H1-Typet*

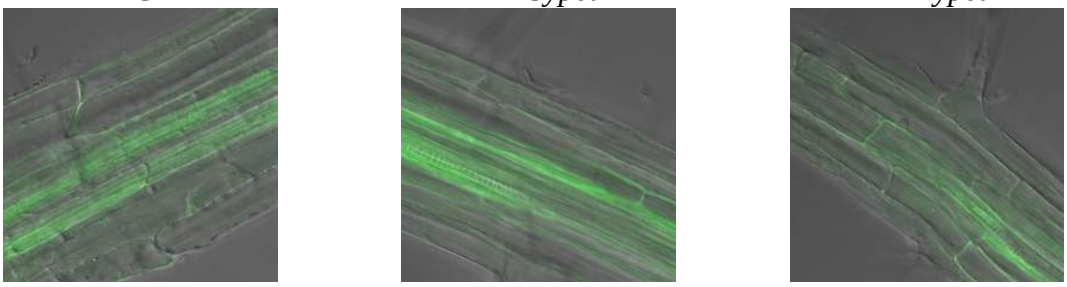
NOT detectable



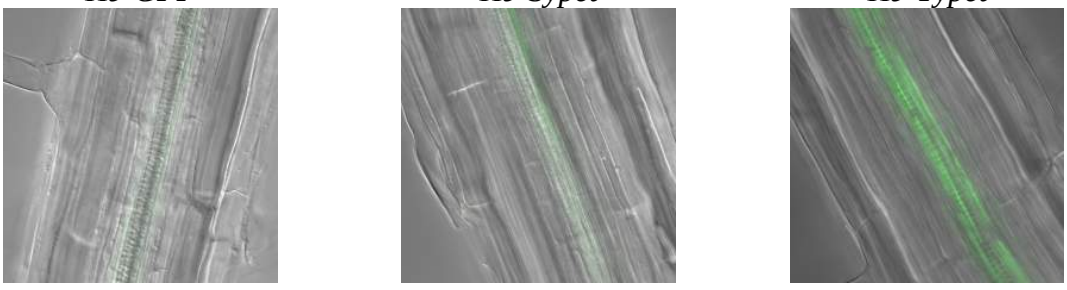
H2-GFP *H2-Sypet* *H2-Typet*



H4-GFP *H4-Sypet* *H4-Typet*



H5-GFP *H5-Sypet* *H5-Typet*



H6-GFP *H6-Sypet* *H6-Typet*

Figure 27. Confocal images with higher magnitude. For each construct, image was taken from one representative T2 line with higher intensity. The names of constructs are displayed on the bottom panels. Overlay images of fluorescence and interference contrast microscopy (DIC) are shown.

In summary, we have examined all the steps involved in the generation of gene translation fusion utilizing TAC clones and the λ-Red recombineering system. The results presented here indicate that this system is reliable and efficient. Importantly, it provides several advantages to the alternative existing methods. 1) By tagging the gene of interest in the pseudochromosomal context of a bacterial artificial chromosome the presence of all the regulatory sequences of a gene, such as potential miRNA targets sites, introns, 5' and 3' UTRs, alternative splicing variants, and nearby as well as distant promoter sequences, is ensured. 2) By using homologous recombination, the need for most standard *in vitro* cloning techniques and extensive re-sequencing of the obtained chimeras is circumvented. Precise modifications in large DNA molecules can be generated without constraints imposed by availability of appropriate restriction sites or the size of the genomic fragments that can be efficiently amplified with PCR-based strategies. 3) By combining the advantages of recombineering with the availability of end-sequenced TAC clones and the ease of *Arabidopsis* transformation, the whole procedure can be easily scaled up to a genome-wide level.

MATERIAL and METHODS

Reporter gene constructs

Prepare of competent cell

SW102/SW105 or *Agrobacterium C58* cells were grown in a water bath with constant shaking overnight at 32°C in 3 ml LB supplemented with tetracycline (Sigma-Aldrich®, MO, 12.5 μg/ml final concentration) or rifampicin (Sigma-Aldrich®, MO, 100 μg/ml final concentration) respectively. For SW102/SW105 cells containing targeting TAC, 25 μg/ml kanamycin (Fisher Scientific, PA) was also added to the media. Next day, 500 μl of the overnight culture was diluted in 50 ml LB within a 125 ml baffled conical flask and

incubated at 32°C in a water bath with constant shaking to an OD600 of approx. 0.6 (0.55-0.6). The samples were briefly cooled in an ice/waterbath slurry (~5 mins), transferred to one 38 ml centrifuge tubes (NALGENE® Labware, NY) and then pelleted by centrifugation at 7000 RPM and 0°C for 5-10 min (Sorvall RC 5B Refrigerated Superspeed Centrifuge). Supernatant was poured off and the pellet was resuspended in 1 ml ice-cold 10% glycerol by gently swirling the tubes in an ice/waterbath slurry. When resuspended, ~37 ml ice-cold 10% glycerol was added to tubes and the samples were pelleted again. The centrifuge and resuspend step were repeated for another two times, and then all supernatant was removed by aspiration using a 9″ Pasteur Pipets, (Fisher Scientific, PA) and the pellet was kept on ice.

Mini-prep preparation

TAC DNA mini-preps were prepared from 3 ml overnight cultures. The cultures were transferred to 1.5 ml eppendorf tubes and pelleted at 14,000 RPM (Eppendorf Centrifuge 5417C) at room temperature for 30 sec. Supernatant was removed by aspiration. An alkaline lysis procedure was used to prepare BAC DNA using solutions I, II and III (Sheng et. al., 1995). Cell pellets were resuspended in 150 µl chilled solution I and placed on ice. To this, 200 µl freshly prepared solution II were added, the tubes were inverted 7-9 times to mix and placed again on ice. After adding 150 µl solution III, the tubes were again inverted 7-9 times and immediately spun (Eppendorf Centrifuge 5417C) for 6 min at full speed at room temperature. The supematant was transferred into a new tube and the nucleic acid precipitated by addition of 1 ml of ethanol at room temperature. After mixing by inversion, samples were immediately spun for 6-10 min at full speed and room temperature. DNA pellets were washed with 1 ml 70% ethanol, air dried and resuspended in 30 µl TE (10mMTris, 1 mMEDTA, pH 8.0).

Transfer of TACs from DH10B to SW102 or from SW102 to *Agrobacterium*

About 6 μl of the DNA from a mini-prep were added to 40 μl ice-cold, freshly-prepared electrocompetent cells. Electroporation was carried out in cold 0.1 cm electroporation cuvette on Eppendorf Electroporator 2510 at 1.8 kV. After electroporation of the TACs, the bacteria were recovered in 1 ml LB in a 15 ml culture tube (Fisherbrand®) for 1 hour at 32 C in an incubator with constant shaking (~190 RPM). After the recovery period the bacteria was pelleted in a 1.5 ml eppendorf tube at 14,000 RPM for 30 sec and the supernatant was removed with a pipette. The pellet is resuspended in 50 μl LB and plated onto LB plates supplemented with kanamycin (25 μg/ml final concentration) and incubated 2-3 days. Primers specific for GOI were used to check the colonies by PCR.

Recombineering using *galK* positive/negative-selection cassette

pBluescript containing *galK* or *GFP* cassettes were constructed in our lab by Dr. Jose Alonso and Dr. Larrisa Benavente. *Sypet* and *Typet* cassette were synthesized by GenScrip, NJ and cloned into vector pUC57. These plasmids were purified from host strains and used as PCR DNA template for the generation of the different recombineering cassettes. Primers used for generating recombineering cassettes were designed with 50 bp of homology to the sequences flanking the desired insertion site followed by 18 bp (forward primer) or 24 bp (reverse primer) of homology to the linker sequences of recombineering cassette (Figure 7B). These long oligos (REC_F and REC_R oligos in Table 4) were synthesized by Integrated DNA Technologies, IA. The *galK, GFP* cassettes were amplified by PCR (PTC-200 Thermo Cycler, MJ Reseach, MA) with the designed primers and a proof-reading Taq-mix (Pfx 50TM DNA Polymerase, Invitrogen, CA) using the following PCR conditions: 95°C 3 min., then 94°C 30 sec. -> 54°C 30 sec. -> 72°C 2 min. for 35 cycles. The *Sypet and Typet* cassettes were amplified using

Accuzyme™ DNA Polymerase (Bioline, MA) with the following PCR conditions: 94°C 10 min., 94°C 30 sec. -> 64°C 30 sec. -> 68°C 3 min., for 8 cycles then 94°C 30 sec. -> 56°C 30 sec. -> 68°C 3 min., for 25 cycles. After amplication, 2 μl DpnI were directly added to the PCR product and incubated at 37°C for 1-2 hour to cut methylated plasmid template. Chloroform extraction and ethanol precipitation were used to purify the DpnI-digested PCR product. DNA was resuspended in 20 μl ddH2O (or TE buffer, ph 8.0).

Before making competent cells, the SW102/SW105 culture was transferred to another prewarmed waterbath (42°C) and heat-shocked for 15 min with constant shaking. 4 μl *galK* cassette or *GFP* cassette were added to 40 μl electroporation in a pre-chilled 0.1 cm cuvette and immediately electroporated on Eppendorf Electroporator 2510 at 1.8 kV. After electroporation of the *galK* cassette, the bacteria were recovered in 1 ml LB (15 ml culture tube, Fisherbrand®) for 1 hour in at 32°C with constant shaking. After the recovery period the bacteria was washed twice in 1xM9 salts and plated onto M63 minimal media plates supplemented with 0.2% galactose (Warming et. al., 2005). To replace *galK* for *GFP*, the procedure was exactly the same except that bacteria containing the TAC clones tagged with *galK* were used, and after the electroporation of the *GFP* assette, the cells were recovered for 3 hours, washed twice with 1xM9 and plated on M63 minimal media plates supplemented with 0.1% 2-deoxygalactose and 0.1% glycerol (Warming et. al., 2005). After incubation for 4-5 days at 32°C in a "150 Science Teaching Incubator" (Lab-Line Instruments, Inc., IL), colonies were tested by PCR with forward and reverse primers flanking the insertion site (IGFPT-F and IGFPT-R primers listed in Table 4). Colonies with right insertion size (*galK* or *GFP*) were picked out for further

processes. Putative *galK*$^{+}$*/GFP*$^{+}$ colonies were streaked in LB plates supplemented with kanamycin and retested. PCR products using IGFPT-F and IGFPT-R primers from *GFP*-tagged clones were purified by ethanol precipitation and sequenced (Genewiz, NJ) using the forward primer: GFP-seq-F (5'-ACAAGACACGTGCTGAAGTC-3') and reverse primer: GFP-seq-R (5'-TATTCCAACTTGTGGCCGAG-3'). Sequencing results were analyzed with Sequencher® (Gene Codes Corporation, MI).

Plant transformation

The desired TAC containing *Agrobacterium* strain was inoculated in 50 ml of LB supplemented with kanamycin (25 ml /ml final concentration), grown at 28C and constant shaking overnight, and then used for starting a large culture of 50 ml of LB in a 1L Erlenmeyer flask. Cells were cultured at 28ºC for about 24 hours till near saturation with constant shaking. *Agrobacterium* cells were pelleted by centrifugation at room temperature and then resuspended in 250 ml of infiltration media (5% glucose or sucrose in water, plus 200 µl Silwet-77). Alternatively, TAC containing *Agrobacterium* stains were grown on plates (150mm×15mm, 2 plates for each transformation) supplemented with kanamycin at 28ºC for 2-4days. The cells formed a uniform lawn that was harvested from the plates and resuspended in 250 ml of infiltration media .

Plants were grown at 22°C under 16-h-light/8-h-dark cycles in a controlled climate growth chamber for approximately 4 weeks. Floral dip (Clough et. al, 1998) or vacuum infiltration (Bechtold et. al., 1998) was conducted. Pots with transformed plants were placed horizontally in a clean tray and immediatelly covered with a transparent plastic dome to maintain high humidity. The trays were moved back to the growth chamber. After ~24h, plastic domes were shifted ~2-3 cm sideways to facilitate air exchange (to gradually reduce humidity) for another ~24h. Plastic domes were then completely

removed, and the pots were positioned vertically. Transformed plants were watered as needed until they finish flowering.

Seeds were collected from fully dried plants and approximately 8000 T1 seeds (~0.2g dry weight) were surface sterilized for 10-15 minutes in 50 ml tubes using 30-40ml of 50% bleach spiked with 3-4 drops of Triton-100 per liter. After that the seeds were suspended in di H_2O, Basta (Phosphinothricin, glufosinate ammonoium) (20 µl/ml) and Timentin (300 µl/ml) were supplemented. Seeds were cold-treated at 4ºC for 2-3 days to equalize germination. To plate the seeds, pre-cooled 0.7% top agarose was added to the seeds and using a 10ml pipettes were uniformly distributed on the top of of AT plates (1x MS salts, 1% sucrose, pH 6.0; 0.6-0.8% agar) supplemented with 20 µl/ml Basta. Plates were incubated in the light for 1-2 hours at room temperature and then placed in the dark incubator at 22ºC. After 3 days in the dark, plates were moved to constant light for 2-5 days. Green seedlings were selected and transferred to soil for propagation and further analysis.

Genotyping of transformants

Tissues of basta-resistant transgenic lines (leaf tissue (10-20 mg) from ~14-day-old of T1 plants or 25-50 3-day-old T2 seedlings) were harvested into 96 well plates and DNA was extracted following the CTAB DNA extraction protocol (Doyle et. al., 1987). 3 sets of PCR were conducted to genotype transgenic plants: 1) IGFPT-F and IGFPT-R primers (Table 4) were used to amplify both the transgene and the endogenous gene (also used as a control for quality of DNA), 2) IGFPT-F and GFP-seq-R primers were used to confirm the presence of *GFP*-tagged GOI, and 3) SacB-F (5'-AATAAAGATTCTTCGCCTTG-3') and SacB-R (5'-TGTAAAACAAGCCACAGTTC-3') primers to examine the integrity of left border of TDNA.

Copy number of T-DNA insertion

Segregation of the basta resistance

Approximately 40 T2 seeds per T1 line were planted on AT plate and at same time, about 100 T2 seeds per T1 line were screened for resistance on basta plates. Seedlings germinated on AT plates were counted to calculate the germination rate of each T1 line examined. On basta plates, seedlings developing green color upon light exposure were counted to calculate the ratio of basta resistant vs sensitive plants. The resistant ratio was calculated correcting for the germination rate. Single copy lines were identified if the corrected resistant ratio falls between 0.63-0.86 (P=0.05, Pearson's chi-square test). Lines with revised corrected ratio of >0.86 were considered to have 2 or more insertions.

Real-time PCR

Genomic DNA was extracted from pooled T2 seedlings following CTAB DNA extraction protocol (Doyle et. al., 1987). DNA pellets were resuspended in 50µl di H_2O and 2 µl of each further diluted sample (40×) were PCR amplified using an ABI 7900 System (Applied-Biosystems, CA) in a 10-µl reaction, containing 5 µl 2×SYBR Green Master Mix, 1 µl of 10 µM forward primer, 1 µl of 10 µM reverse primer, and 1 µl of water. Reactions were incubated at 50 ºC for 2 min, 95 °C for 10min then 95 °C 15 s -> 60 °C 30 s -> 72 °C 30 s for 40 cycles. Dissociation curve analysis was performed subsequently. Real-time PCR results were analyzed using the SDS 2.2.3 (Applied-Biosystems, CA) package. The quality of the amplification was tested by examining the bimodal dissociation curve and amplification plot. All reactions were repeated 3 times (technical replicates) and the median of the 3 Ct values was used for further analysis. Using Columbia as control, the copy number of a gene was calculated based on ratio between transgenic lines and wildtype line. An *Arabidopsis* gene (*AT2G01830*) known to have two

copies in *Arabidopsis* genome was used as endogenous control (EC). Amplification efficiency for each primer pair was calculated using serial DNA dilutions. For this purpose, equal volumes of all DNA samples were pooled and 1x, 2x, 4x, 8x, and 16x dilutions were run in triplicates for each primer pair. Efficiency of primers was then determined the formula: E=2^[-1/slope] (Pfaffl et. al., 2001). The fold change between transgenic line and control line was calculated using the equation: Fold change=Egene^($Ct_{col\text{-}gene}$-$Ct_{sample\text{-}gene}$)/E_{EC}^($Ct_{col\text{-}EC}$-$Ct_{sample\text{-}EC}$) (Pfaffl et. al., 2001). The number of insertion was then calculated as 2×(Fold change-1).

The following primer pairs were used for the Real-Time PCR experiments: At3g56310 forward primer 5'-CCCACAGCATTCTTCCAAGT, and reverse Prime 5'-GTTGGTCATCG- TGTGC; At3g56140 forward primer 5'-GCGCCACCACTTCTAGTCTC, and reverse primer 5'-TTCCTTATTCACGGGTGGAG; AT3G56240 forward primer 5'-TGAG-AAGCAGTTTTCC, and reverse primer 5'-CGGTTTCGACCTTAGGATCA; EC forward primer 5'-TGGATGGTTACGTCTCCAAAC, and reverse primer 5'-GCAT-CAATCAGCAGAATCG.

Fluorescence microscopy and confocal laser scanning microscope

3 day old dark grown T2 seedlings were mounted in water on a microscope slide (Fisher Scientific, PA) and covered with a coverslip (Corning, MA). GFP fluorescence was observed with an Zeiss Axioplan (Carl Zeiss MicroImaging Inc, Germany) under 10 × or 20× magnitude and the images were acquired with a SPOT Insight™ (SPOT™ Imaging Solutions, MI) camera with the SPOT advanced software (SPOT™ Imaging Solutions, MI). To facilitate the comparisons between independent T2 lines of a construct, images were acquired using the exact same settings. Lines with strong fluorescence were selected

for confocal microscope imaging. A LSM 710 confocal workstation (Carl Zeiss MicroImaging Inc, Germany) was used. Fluorescence and differential interference contrast microscopy (DIC) images were acquired at both 20 × and 40× magnification. The obtained images were analyzed using the Zen 2009 software (Carl Zeiss MicroImaging Inc, Germany). All images were further processed with Adobe Photoshop (Adobe Systems, CA).

Table 4. Primers used to generate the constructs described in Chapter 1. IGFPT-F and IGFPT-R: forward and reverse primers used to test recombinants by PCR. REC_F and REC_R: primers used to amplify *galK*, *GFP*, *Sypet* and *Typet* cassettes.

NAME	IGFPT-F	IGFPT-R	REC_F	REC_R
IAA1	AAGATGGAGATTGGATGTTG	CGTATTTGTAACCCTATTGC	CTTGTCAAAAACTCAGAATCATGAAAGGATCCGAAGCTCCTACTGCCTTAGGAGGTGGAGGTGGAGCT	GTAGAATTTGTTTTTGCCTCGACCAAAAGGTGTTTTGAGACAATGGATCAGGCCCCAGCGGCCGCAGCAGCACC
IAA10	GTTCCTTGGCAGTAAGTGAC	TATTCACAAACGACCACAAG	TGACAAGACTGAGAATCATGAAGACATCAATTGGAGCTGGAGTAGGTAAGGGAGGTGGAGGTGGAGCT	AGTTACAGACCAAAGAGATGTAAACTGTTTTTGCTATATAGGACTCTCTAGGCCCCAGCGGCCGCAGCAGCACC
IAA11	GGATAATGAAAACATCGGAG	AAGAAACTGTAAAGGGGAGA	TCGTTGTGACCGTGTTTGTGTGATCCTTGTTTACAGCTCAGATGATATTAGGAGGTGGAGGTGGAGCT	TATTACTGCTTCAAACATAAAGGGTTCTTCGTAGATTATTATTTGCTTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA12	CTTCTGGTTTCTTGATTTCTCG	GCTTCTGCATGCTTTTATCC	CAGCTCCAAGACGTCAAGAGCAGAAGGATAGACAAAGAAACAACCCTGTTGGAGGTGGAGGTGGAGCT	TGCAAACCTCAAACAATACATAAACAATGCCAGCTTTGGAAGGGAAGCTAGGCCCCAGCGGCCGCAGCAGCACC
IAA13	AAGCTAATGGACTCGGTATG	GCAACCGAAAATTAAAAGAC	CAGCTGCACGAAATCAAGAACCAAACGAGAGACAGCGAAAGCAGCCGGTTGGAGGTGGAGGTGGAGCT	AAACCCCAAAATATAAAACCTGTAACACCGTAACGTCGAAAAGAGATCTAGGCCCCAGCGGCCGCAGCAGCACC
IAA14	TGTTCGTTTCTATAAGTGGC	AGGAAAAGAATAATAAAGTAGG	GTGTTTTTTTTTCAGCTCCAAGAGCAATGGAGAAGTTCAAGAACAGATCAGGAGGTGGAGGTGGAGCT	ATACCAAAAAAAAAAAAATCAATGCATATTGTCCTCTTTTTTTTTGTTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA15	GGACGGTGACTTGATGTTAG	ATTAAAACTTTGAGCATCCC	AATCTTGCAAACGTATGAGGTTAATGAAAACCGGAGATGCTATTGGATTAGGAGGTGGAGGTGGAGCT	CTTTCGAAGATCCGATGCGTTTACGTTGTGGGAAGCAGAATAGAGACCTAGGCCCCAGCGGCCGCAGCAGCACC
IAA16	AAGCAATCGGACTTGGTTAG	ATAGTTTCGAGCCGGTTTAG	TTCTATATATTGCAGCTCCAAGGGCATTAGAAAAGTGCAAGAACAGAAGTGGAGGTGGAGGTGGAGCT	AGGCTTTCTTCCTTTTTTAGGTAAGAACACGAAATGTCGTCGAGAACTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA17	TTTTAACAGAATAAATACCGTG	CCAAATCCAGATCAAAACAC	GTAAAAATATTTCAGCTCCGAGGGCGATGGAGAAGTGCAAGAGCAGAGCTGGAGGTGGAGGTGGAGCT	TGTTAATCAAATATATAATCGATACCACTTATCCTTTTAATTTGACTTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA19	GGGGTATGTTCTATCTTTTGAC	TCAACACTCAAGAAACAAGTAG	AAAGATCGGATGCTACCGGGTTTGGGCTGCAGCCTAGAGGAGTAGACGAGGGAGGTGGAGGTGGAGCT	AAGATTAAATTAATGAACCAGCTCCTTGCTTCTTGTTCAAGTCATCATCAGGCCCCAGCGGCCGCAGCAGCACC
IAA2	ACGAAGATAAAGATGGTGACTG	TTTCGTCTTTCACGGTCAAC	GACTCAGAATCATGAAGGGATCCGACGCTCCTGCTCTAGACTCTTCCTTAGGAGGTGGAGGTGGAGCT	ATAGTCATCATCTCCTTTTTAGACAAAAGATTCTCAGCTTCTCTGGATCAGGCCCCAGCGGCCGCAGCAGCACC
IAA20	ACGCAGATAAGGAAGGTGAC	ACTACCGAGAATAATCAGAG	TCTTGTCTACTGTGAGAAGACTGAAGATTTCAAGAGCTAATTACCACTACGGAGGTGGAGGTGGAGCT	CTTGGCATTACTTCGATTTAAACACAACAAGTACGTAGACTTATATATCAGGCCCCAGCGGCCGCAGCAGCACC
IAA26	AGTTGGATGCAGTAAGCAAG	AATCTCATCGTCTAGTGTCTTG	TCCTCTGTATATTTGCAGTTGGATGCAGTAAGCAAGAGAAGATGATGCACGGAGGTGGAGGTGGAGCT	AAGCTTTGAGACTCTCTGTCTGAAATAAAAGTTTCTAAAAAGACTCTTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA27	TACTGGAGAAACTAAAACGATG	GGAAGAGGCCTTATTAGAGC	TTTCTTTTTTTGCAGCTCCAAGGGTGATGGAGAAGTGCAGAAGCAGGAACGGAGGTGGAGGTGGAGCT	AAAACATATTTTTTCAAATCATATATTATCACATAAAAGGGCTTTTCCTAGGCCCCAGCGGCCGCAGCAGCACC
IAA28	ATGTTTTAAAGACCTCCCAC	AGTAAATCAAAGCCAAACCC	TAACTTATATATTTATTTATTTTTTCACAGCTAGAAAACATGGCAAGGAAGGAGGTGGAGGTGGAGCT	TACATTAAAAACAAACCATCGAACTGATGATTTTGGCCAACCTCTCTCTAGGCCCCAGCGGCCGCAGCAGCACC
IAA29	CTGGCTACTTCGAGGGGATG	AACCATTGGATTTGATACTAAG	GGATATCAATAATTAGAGATCGACCGTGTGCATATACAAGATGTTTGTTTGGAGGTGGAGGTGGAGCT	TTTTAATTACAAAATAAAAAATAATGGTCCGATTTGAACGCCTATCCTTAGGCCCCAGCGGCCGCAGCAGCACC
IAA3	CAAAGGTTTAGGCTGTGGTG	CCGATGAGGTAAATAAAGGC	GACTAAGGATCATGAAAGGATCAGAAGCCAAAGGTTTAGGCTGTGGTGTAGGAGGTGGAGGTGGAGCT	TTCTTCAAGTTCTTTTGTGTCTCTGTTAGATTTCTTGAAGATATATCTCAGGCCCCAGCGGCCGCAGCAGCACC

Table 4 (Continued)

IAA30	ACGCAGACAAA GAAGGTGAC	AAAGGATCTGCA TTAGCTAC	TGTTCTTGTCTAGCGTGAGAAGACTAAAGATCTCAA GAGCTTATCACTACGGAGGTGGAGGTGGAGCT	GACTGTATTGGTCAATCCTAGATTCTTCCAAGTTTGCTT TTAGAAAATCAGGCCCCAGCGGCCGCAGCAGCACC
IAA31	CATTGGTATAA TTTCCTCGG	ATCAATCAATTCT CCAACTTC	TGTTTCTTGAAACCGTGAGAAGACTAAAGATCACGA GACCGGAGAGGTATGGAGGTGGAGGTGGAGCT	TTACACGTCTCGTAACTGCGCAATCACAGCCTTGACCG ATCCAAGTTTTAGGCCCCAGCGGCCGCAGCAGCACC
IAA32	TGTCTGAAACC TGTAGGGAG	CAATAGAACCAC AAGTGAGC	GCGTGGATCGGATGCGAATCGCAAGAAGAAACGAT GCTCTTCTTCCCTTTGGAGGTGGAGGTGGAGCT	TCTTAGATAATCAGTGAACGCCCCAGAAGAAGCTTGG ACTTAAAAGCTTAGGCCCCAGCGGCCGCAGCAGCACC
IAA33	GGTTTTCGTTC AATGTTTAG	GGCTTATCTTATT TTGTTCTTG	TTCGAATCTTGCCGGTCAAGGGAAACACAAGACAA GTTAAAAGAAACGAGGGAGGTGGAGGTGGAGCT	TTACAAAACATCGTATGACATCGTTGCAAAAACGGTCT TTCTCCTCCTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA34	AACTCTTGCTC ATCAACTCG	GTAAGACGAGAA TCCAGACG	GCGTGGAGCGGCTGAGAATCACAAGAAGAAACGAT GCTGTACTTCCCTTTGGAGGTGGAGGTGGAGCT	CTGACTTGGAAATTATAACAAAGAGTGAAGAAAGAGA GAAGCTAATATTAGGCCCCAGCGGCCGCAGCAGCACC
IAA4	TGCAGGATGTT TGTTTCGTC	AACCCGAGCAGT TTCAGAGC	TAAGGATCATGAAAGGATCTGAAGTTAAAGGTTTAG GTTGTGGTGGTCTTGGAGGTGGAGGTGGAGCT	CTGTTTTTGAATCTTTCTCTTTGATCTTTATCTCTCAATT AGGTTTGTTAGGCCCCAGCGGCCGCAGCAGCACC
IAA5	AAGACAAAGAT GGAGATTGG	AGGCAATGTAAA CAGATTCG	TTGGATCGTGCAAGAGGTTAAGGATTATGAAGAGAT CATGTAACAGAGGAGGAGGTGGAGGTGGAGCT	GATATCTGTAAGGCTCACTCACATTCACATGTGACTAT AAATATCATTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA6	TAAGTTTATGT TTAGCTTGC	GTATGATACGAC CAGTAAGG	TGAGGATCGTGAAGAGATCAGATGCAACTGGTTTTG GTCTCCAGCAAGATGGAGGTGGAGGTGGAGCT	TTCAAACTTTATGGTTAATTAAAGATGATCAACACTAG CTAGTATGATTAGGCCCCAGCGGCCGCAGCAGCACC
IAA7	GCAGTTGGACT TGGTACATC	ATCATCAAACCG ACAGACAC	TCTTGGTTATAGCTCCGAGAGCAATGGAGAAGTACT GCAAGAACAGATCTGGAGGTGGAGGTGGAGCT	CCCGTCATATTGTTGATCATTGATGCTTCTTTTTTTTTA TTATTAATTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA8	CTGAATACAAA CTCAATGGC	AATCAAGATTTTG CGTAGAG	CTGTGGCAGCTCCAGGTGCAGTGGAGAAATCGAAG AACAAAGAGCGGGTTGGAGGTGGAGGTGGAGCT	ATTGGAACTGTGCATAGTGTTTTTTTTTTGTCTTGAGCA TCTTCGTTTCAGGCCCCAGCGGCCGCAGCAGCACC
IAA9	GCTATTGGGTT AGGTAAAGTC	TCATTTCCTTACA TTCGTAGTC	ATATTGCAGCTGCAGCTCCGAGAGCAATGGAGAAA TCGAAGATGAGAGCTGGAGGTGGAGGTGGAGCT	ATCGCACATATTTTGTATAGTTAGAAAGCTTTGTCATG TTCCATCTTTTAGGCCCCAGCGGCCGCAGCAGCACC
PFL	TCGGCAACAAA ACCTAATCTG	ATTCTTGCGTATG GAGGAGTG	CCAAAACCACCGGACGCCGTGGAAAGACTGTTGGT GTTTCCAAGAAGCGTGGAGGTGGAGGTGGAGCT	ATACAAAAAAAACAAAACAAAGCCATTGAACCAAGCT CAAACTCTTTTTAGGCCCCAGCGGCCGCAGCAGCACC
ATX1	GAAAAGATTGT CGATATCGCG	TGCCCCAGTGCAT ATTTATTC	CCGGAAAGAAAACGGCTTTTTGGGAAGCTGAAGGT GAAACTGCTAAGGCTGGAGGTGGAGGTGGAGCT	ATAGATTTTATTAAAACAGAGAATCACATAAAATGAT CTGCTTCTCCTTAGGCCCCAGCGGCCGCAGCAGCACC
CCH	GCTGATCCTAA GGTCGAAACC	AAAGAAAGTCGG ACCAACATG	CTGATGTTGAACCAAAAGCCGCAGAAGCCGAGACT AAGCCATCACAAGTTGGAGGTGGAGGTGGAGCT	GAAGCATTGTAAAGGTAGATTAGCCTTCCTTTTCATCG TTAGTATACTTAGGCCCCAGCGGCCGCAGCAGCACC
PIP2A	CATGGGATGAC CACGTAAGAC	CCAATCACTCGTT TAAAAGCG	GAGCTTCAGGTTCTAAGTCTCTTGGATCATTCAGAA GTGCTGCCAACGTCGGAGGTGGAGGTGGAGCT	GAGAGATCACAACTTCATTTATTCTTGAGAAATCTGGT TTGATGTGTTTAGGCCCCAGCGGCCGCAGCAGCACC
LTP	CTCCTAATGAT TCTAGCCCCG	TGCATACAAAGC ATCACCAAG	CCAACGCTCGTAAAGTCTCTAAAAGTTGCAAACTCC CCATCCCAAGGTGTGGAGGTGGAGGTGGAGCT	CAAACATATATAGGATGATCATAAACAGCAAACACTT TTCTGACTTTTTAGGCCCCAGCGGCCGCAGCAGCACC

CHAPTER 2

Recombineering-based Tagging of *Arabidopsis* Aux/IAA Gene Family

SUMMARY

As described in the previous chapter, a recombineering-mediated tagging system can be efficiently and reliably used to tag *Arabidopsis* genes. However, to ensure the system can be used as a general strategy to study gene expression and protein localization, it is desirable to evaluate the feasibility of examining genes with low expression levels and short protein harl-lives. *Arabidopsis* Aux/IAA family contains 29 genes with 4 conserved domains referred as domains I, II, III and IV (Hagen et. al., 2002; Liscum et. al., 2002). These early auxin response genes are expressed at relatively low levels and their corresponding proteins are rapidly turned over, which provides us a very good model to test our recombineering-based tagging system. Seventy-four constructs have been made to study the expression pattern and subcellular localization of 28 Aux/IAA. By mutating the core residues of the stabilization domain II, and replacing a conserved sequence of dimerization domain IV by an improved fluorescent protein (3 copies of codon-optimized Ypet) we were able to visualize the expression of these genes. Tagged Aux/IAA proteins were found to be expressed in specific parts of the *Arabidopsis* seedlings and to be nuclear localized. Our results demonstrate that recombineering-based gene tagging can be used, not only to insert fluorescent tags in a gene of interest, but also to introduce point mutations and replace one sequence by another. These types of sequence modification were shown to be essential to obtain gene expression information and protein subcellular

localization of these low abundant and short-lived proteins. The detailed analysis of these genes' expression will provide new insights on the regulation auxin-mediated processes in *Arabidopsis*.

Table 5. Constructs made for determining the expression pattern and subcellular localization of Aux/IAA gene family. Δ: deletion. C: C-terminal tagging. →: amino acids exchange.

GENE	CONSTRUCT	Domain II	Domain IV	TAG	TAGGING POSITION
IAA1	*IAA1-GFP*	-	-	GFP	C
	IAA1M2--GFP	PP→PS	-	GFP	C
	IAA1M2R4-Sypet	PP→PS	ΔGFVPTYEDKDGDWMLVGDVPWD	Sypet	Domain IV
IAA10	*IAA10-GFP*	-	-	GFP	C
	IAA10M2R4-Sypet	PP→PS	ΔSEYIITYQDKDGDWMLVGDVPW	Sypet	Domain IV
IAA11	*IAA1-GFP*	-	-	GFP	C
	IAA11M2R4-Sypet	PP→PS	ΔSGLVLTYEDKEGDWMLVGDVPW	Sypet	Domain IV
IAA12	*IAA12-GFP*	-	-	GFP	C
	IAA12M2--GFP	PP→SP	-	GFP	C
	IAA12M2R4-Sypet	PP→SP	ΔSDFVLTYEDKEGDWMLVGDVPW	Sypet	Domain IV
IAA13	*IAA1-GFP*	-	-	GFP	C
	IAA13M2R4-Sypet	PP→PS	ΔSEFVLTYEDKEGDWMLVGDVPW	Sypet	Domain IV
IAA14	*IAA14-GFP*	-	-	GFP	C
	IAA14M2--GFP	PP→PS	-	GFP	C
	IAA14M2R4-Sypet	PP→PS	ΔSEYVPSYEDKDGDWMLVGDVPW	Sypet	Domain IV
IAA15	*IAA15-GFP*	-	-	GFP	C
	IAA15M2R4-Sypet	PP→PS	ΔGEFVATYEDKDGDLMLVGDVPW	Sypet	Domain IV
IAA16	*IAA16-GFP*	-	-	GFP	C
	IAA16M2R4-Sypet	PP→PS	ΔSDYVPTYEDKDGDWMLVGDVPW	Sypet	Domain IV
IAA17	*IAA17-GFP*	-	-	GFP	C
	IAA17M2R4-Sypet	PP→PS	ΔWDYVPSYEDKDGDWMLVGDVPW	Sypet	Domain IV
IAA19	*IAA19-GFP*	-	-	GFP	C
	IAA19M2--GFP	PP→PS	-	GFP	C
	IAA19-GFPGAL4		-	GFP-IRES^{cp}-GAL4	C
	IAA19-Typet		-	Typet	C
	IAA19M2R4-Typet	PP→PS	ΔCEYVTIYEDKDGDWMLAGDVPW	Typet	Domain IV
IAA2	*IAA2-GFP*	-	-	GFP	C
	IAA2M2--GFP	PP→PS	-	GFP	C
	IAA2M2R4-Sypet	PP→PS	ΔGFVPTYEDKDGDWMLVGDVPWD	Sypet	Domain IV
IAA20	*IAA20-GFP*	-	-	GFP	C
IAA26	*IAA26-GFP*	-	-	GFP	C
	IAA26M2--GFP	PP→PS	-	GFP	C
	IAA26--GFPGAL4	-	-	GFP-IRES^{cp}-GAL4	C
	IAA26-Typet	-	-	Typet	C

Table 5 (Continued)

	IAA26M2R4-Typet	PP→PS	ΔGEFTLTYEDNEGDKMLVGDVPW	Typet	Domain IV
IAA27	*IAA27-GFP*	-	-	GFP	C
	IAA27M2R4-Sypet	PP→PS	ΔSEYVVTYEDKDSDWMLVGDVPW	Sypet	Domain IV
IAA28	*IAA28-GFP*	-	-	GFP	C
	IAA28M2R4-Sypet	PP→PS	ΔRQYTLVYEDTEGDKVLVGDVPW	Sypet	Domain IV
IAA29	*IAA29-GFP*	-	-	GFP	C
	IAA29M2R4-Sypet	PP→PS	ΔTNYTFTFQGKEGDWLLRGDVTW	Sypet	Domain IV
IAA3	*IAA3-GFP*	-	-	GFP	C
	IAA3M2--GFP	PP→PS	-	GFP	C
	IAA3M2R4-Sypet	PP→PS	ΔSDFVPTYEDKDGDWMLIGDVPW	Sypet	Domain IV
IAA30	*IAA30-GFP*	-	-	GFP	C
IAA31	*IAA31-GFP*	-	-	GFP	C
	IAA31M2R4-Sypet	PP→PS	ΔKHHVLTYEDKDGDWMMVGDIPW	Sypet	Domain IV
IAA32	*IAA32-GFP*	-	-	GFP	C
	IAA32--GFPGAL4	-	-	GFP-IRES^{cp}-GAL4	C
	IAA32-Typet	-	-	Typet	C
	IAA32M2R4-Sypet	ΔLIDWSQ	ΔSEFSLVYRDREGIWRNVGDVPW	Sypet	Domain IV
IAA33	*IAA33-GFP*	-	-	GFP	C
IAA34	*IAA34-GFP*	-	-	GFP	C
IAA4	*IAA4-GFP*	-	-	GFP	C
	IAA4M2--GFP	PP→PS	-	GFP	C
	IAA4--GFPGAL4	-	-	GFP-IRES^{cp}-GAL4	C
	IAA4-Typet	-	-	Typet	C
	IAA4M2R4-Typet	PP→PS	ΔSDFVPTYEDKDGDWMLVGDVPWE	Typet	Domain IV
IAA5	*IAA5-GFP*	-	-	GFP	C
	IAA5M2--GFP	PP→PS	-	GFP	C
	IAA5--GFPGAL4	-	-	GFP-IRES^{cp}-GAL4	C
	IAA5-Typet	-	-	Typet	C
	IAA5M2R4-Typet	PP→PS	ΔSECVPIYEDKDGDWMLAGDVPWE	Typet	Domain IV
IAA6	*IAA6-GFP*	-	-	GFP	C
	IAA6M2--GFP	PP→PS	-	GFP	C
	IAA6--GFPGAL4	-	-	GFP-IRES^{cp}-GAL4	C
	IAA-6Typet	-	-	Typet	C
	IAA6M2R4-Sypet	PP→PS	ΔCEYIIIYEDKDRDWMLVGDVPW	Sypet	Domain IV
IAA7	*IAA7-GFP*	-	-	GFP	C
	IAA7M2R4-Sypet	PP→PS	ΔSEYVPSYEDKDGDWMLVGDVPW	Sypet	Domain IV
IAA8	*IAA8-GFP*	-	-	GFP	C
	IAA8M2R4-Sypet	PP→PS	ΔSEFVLTYEDKDGDWMLVGDVPW	Sypet	Domain IV
IAA9	*IAA9-GFP*	-	-	GFP	C
	IAA9M2R4-Sypet	PP→PS	ΔKDYVLTYEDKDGDWMLVGDVPW	Sypet	Domain IV

INTRODUCTION

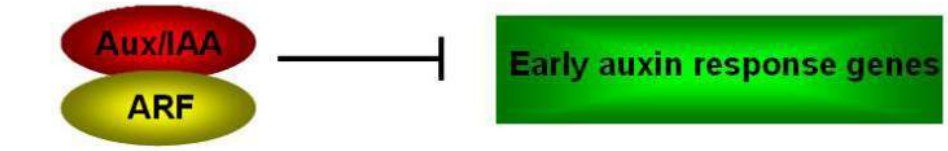
Function and structure of Aux/IAA family genes

The plant hormone auxin regulates diverse cellular and developmental responses in plants including, but not limited to, patterning in embryogenesis, apical dominance and tropic responses (Jenik et. al., 2005; Leyser, 2005; Muday et. al., 2001). Through the study of mutants in *Arabidopsis thaliana*, dramatic advances have been made in the understanding of the mechanisms of auxin signaling and response. Auxin-induced gene expression is regulated by interaction of two related families of proteins, auxin/indole-3-acetic acid (Aux/IAA) proteins and auxin response factors (ARFs) (Guilfoyle et. al., 2007). Aux/IAA genes were first identified as members of a family of early auxin response genes (Theologis et. al. 1985) and the *Arabidopsis* genome contains 29 Aux/IAA genes (Liscum et. al., 2002). Transcriptional factors from ARF family (23 in *Arabidopsis* genome) regulate the expression of genes containing auxin-response elements (ARE) in their promoter (Remington et. al., 2004). The current model (Figure 28 A-B) suggests that under low-auxin conditions, the activities of ARFs are repressed by Aux/IAA proteins. At high concentration of auxin, degradation of Aux/IAA proteins occurs and therefore, the inhibition of ARF activity is released and auxin-mediated transcriptional changes take place (Ulmasov et. al., 1999; Tiwari et. al., 2004).

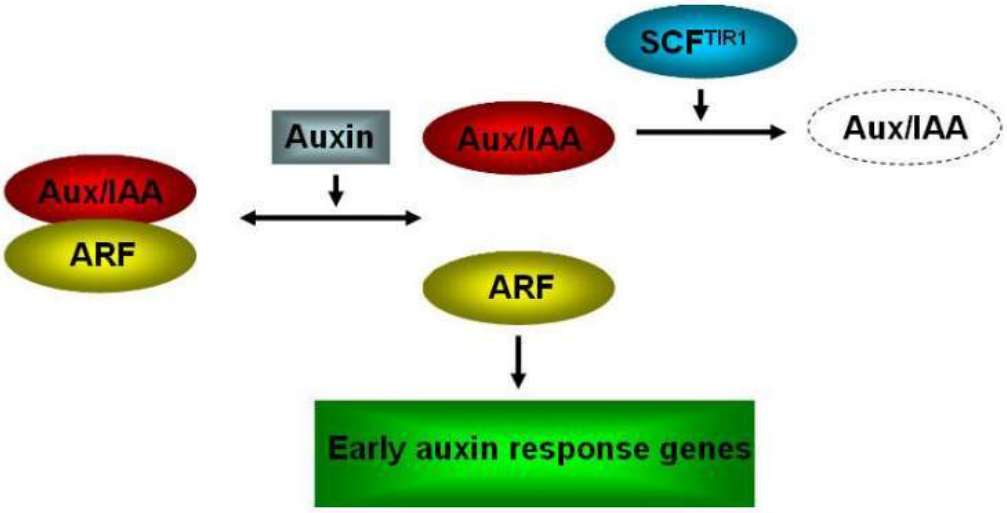
The majority of Aux/IAA genes are comprised of four domains (I-IV) (Liscum et. al., 2002) and share conserved domain architecture (Figure 28C). Domain I is the least strictly conserved and has the smallest size. Specific activity of domain I is still unclear, however it has been shown that domain I is not responsible for the stability of Aux/IAA proteins (Dreher et. al., 2006) and may play a role in Aux/IAA protein homodimerization (Ouellet et. al., 2001). Highly conserved domain II is the target for ubiquitination and is

responsible for mediating a quick turnover of Aux/IAA proteins (Colon-Carmona et. al., 2000). This quick protein turnover of the AUX/IAAs provides the mechanistic basis for the fast and efficient plant response to endogenous or external signals (Worley et. al., 2000). Domain III has been shown to be required for dimerization as mutant versions of domain III compromise the ability of AXR3/IAA17 to dimerize (Ouellet et. al., 2001). Domain IV, together with Domain III, mediate homo- and heterodimerization between Aux/IAA proteins and enable the AUX/IAAs to heterodimerize with ARF proteins (Overvoorde et. al., 2005). Most family members of ARF proteins also share the same structural features (Figure 28C) with an N-terminal DNA-binding domain, a divergent middle region and C-terminal dimerization domains homologous to domains III and IV of Aux/IAA proteins (Guilfoyle et. al., 2007). In the Aux/IAAs, divergent sequences outside the conserved domains are thought to enable these domains to adopt the correct tertiary structure (Reed, 2001).

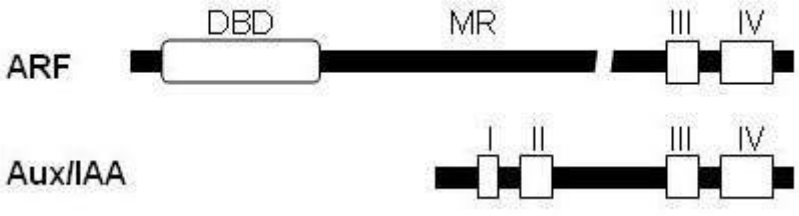
Aux/IAA genes have tissue-specific expression patterns and different Aux/IAA genes respond to auxin signaling with varying degrees and kinetics (Abel et. al., 1995). In fact, although the majority of AUX/IAAs are induced by auxin, the expression of some of them is not auxin-regulated, and in same other cases it can even be repressed by auxin (Rogg et al, 2001). Importantly, the fact that, in general, Aux/IAAs are up-regulated by auxin but act by inhibiting auxin responses suggests the existence of an Aux/IAA-mediated negative feedback loop. The probable role of such a regulatory loop is to provide very precise control of the auxin responses.



A



B



C

Figure 28. Function and structure of Aux/IAA family in *Arabidopsis*. (A-B) Current model of auxin-mediated signaling. Aux/IAA proteins repress the activity of ARFs under basal conditions, whereas increased auxin concentration accelerates degradation of Aux/IAA proteins by interacting with the SCF^{TIR1} ubiquitin-ligase complex. Degradation of the Aux/IAAs releases the ARFs and therefore activates auxin-mediated gene expression changes. It has been shown that auxin regulates degradation of the Aux/IAA proteins by binding directly to an auxin receptor, TIR1 (Dharmasiri et. al., 2005). (C) Conserved domains present in most Aux/IAA or ARF proteins. Sequences between the conserved domains vary in size and the case of the ARF middle region (MR) this variability is represented by a break line. ARFs also have a conserved DNA-binding domain (DBD) that recognizes auxin response elements (AuxREs) present in promoters of auxin-response genes (Remington et. al., 2004). Some Aux/IAA and ARF proteins lack one or more of the conserved domains.

Most of the functional information on the Aux/IAA genes comes from the detail phenotypic and molecular analysis of gain-of-function *aux/iaa* mutations. It is important to emphasize that all these gain-of-function mutants are caused by the stabilization of the protein due to very specific missense mutation in domain II. These phenotypic studies suggest distinct and overlapping roles of Aux/IAA family in development. *shy2-2/iaa3* (Tian et. al., 1999), *shy1-1/iaa6* (Kim et. al., 1996), *bdl/iaa12* (Hamann et. al., 1999), *axr3-1/iaa17* (Rouse et. al., 1998), and *iaa18-1* (Ploense et. al., 2009), all cause leaves to curl up, while *axr2-1/iaa7* (Nagpal et. al., 2000), *slr-1/iaa14* (Fukaki el. al., 2005), and *axr3-1/iaa17* all cause short root hairs and root agravitropism. Figure 29 shows the phylogeny of the *Arabidopsis* Aux/IAA protein sequences based on conserved regions. Members of Aux/IAA family fall into distinct clades and the existence of many closely related sister pairs suggests these genes arose from segmental duplications of the genome (Vision et. al., 2000, Remington et. al., 2004). The overlapping phenotypes shared by some members of the family suggest the existence of mutual compensation between Aux/IAA family proteins. This is further supported by the absence of phenotype in the loss-of-function (LOF) mutants. Thus far, no LOF phenotypes have been reported for Aux/IAA genes other than *shy2/iaa3* (Tian et. al., 1999). However, contrasting effects of different *aux/iaa* mutants on the same auxin response suggest different Aux/IAA genes may have opposing functions. *shy2-2/iaa3*, *shy1-1/iaa6*, *axr2-1/iaa7*, *axr3-1/iaa17* cause shorter hypocotyls than wildtype, but *iaa18-1* mutant has longer hypocotyl than wildtype. Also, *axr3-1/iaa17* mutant has more lateral roots than wildtype, whereas *shy2-2/iaa3*, *slr-1/iaa14*, *iaa28-1* (Rogg et. al., 2001) plants have fewer lateral roots.

Recent progress toward discovering the functions of Aux/IAA genes suggests the developmental specificity of an auxin response is generated by transcriptional

regulatation of optimized interacting Aux/IAA-ARF proteins pairs (Weijers et. al., 2005). Specific Aux/IAA-ARF protein pairs are thought to play key roles on concrete morphological and developmental processes by reading auxin concentration gradients and then translating into corresponding gene expression (Liscum et. al., 2002). However, the number of potential binding partners coupled with complex spatial expression and induction profiles suggest a complicated network of interactions that remains to be elucidated. We are now in a position to determine the expression pattern and subcellular localization of all of the Aux/IAA and ARF proteins (not included in this thesis). These *in vivo* localization studies will be essential to understand where and when auxin signals are turned on, and how different Aux/IAA-ARF pairs specify different auxin responses. This type of detailed expression study, could also guide the selection of those family members that may have potentially redundant functions in specific cell or tissue and particular developmental processes.

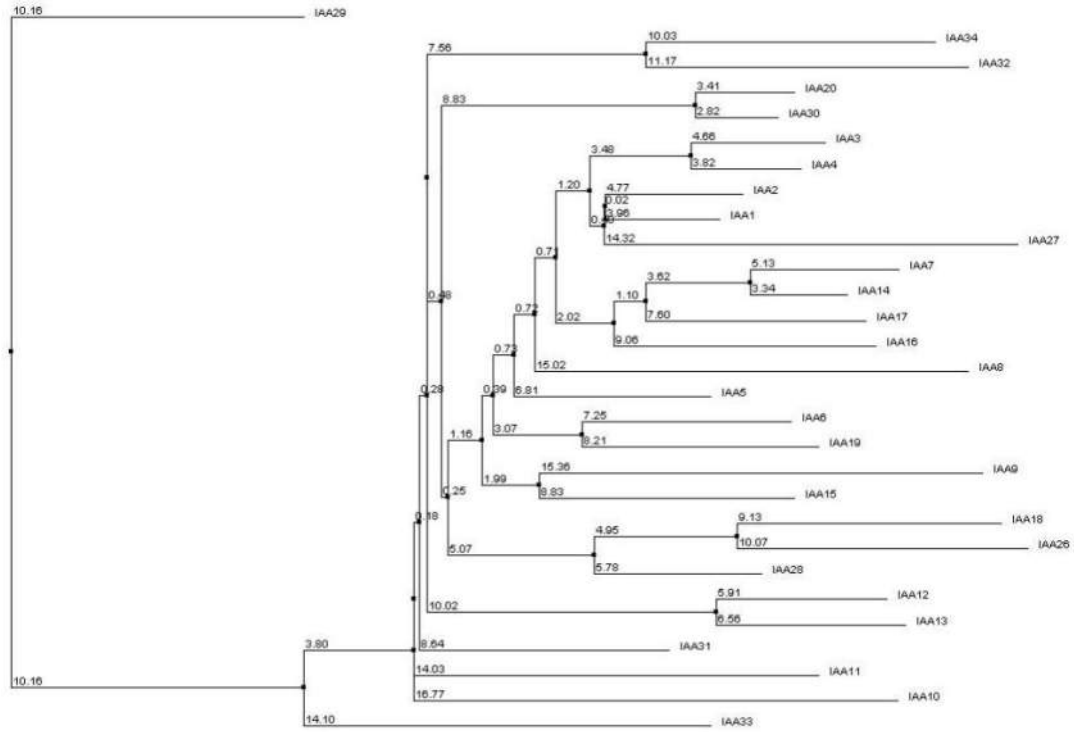


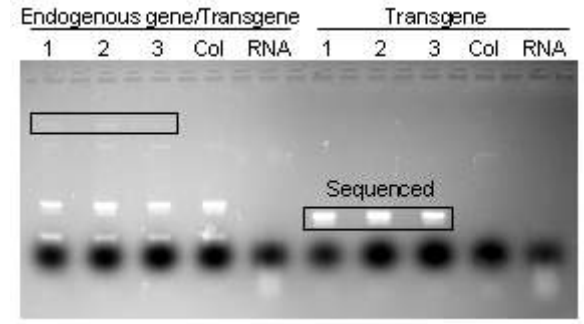
Figure 29. Phylogenetic tree of Aux/IAA family. The multialignment was generated by ClusterW (http://www.ebi.ac.uk/Tools/clustalw2/index.html). Protein sequences were downloaded from T-DNA Express (http://signal.salk.edu/cgi-bin/tdnaexpress) and if a gene has more than one model, the one with the longest protein product was used. Phylogenetic tree were generated with Java view (http://www.jalview.org/) using neighbor-Joining method based on percentage of whole protein sequence identity.

REULSTS AND DISCUSSION

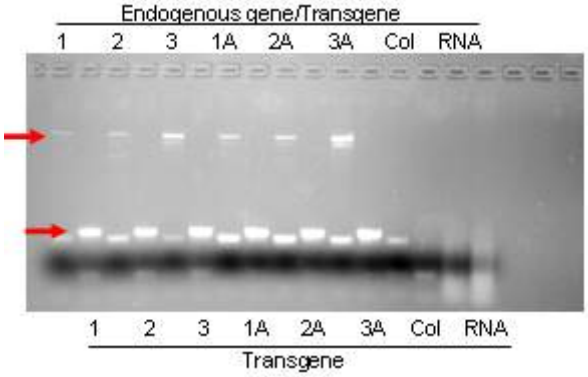
Tagging genes of the *Aux/IAA* family with *GFP*

To determine the expression patterns and subcellular localizations of members of the *Aux/IAA* gene family, 28 *Aux/IAA* genes were tagged at their C-terminal with *GFP* using the recombineering-based system (Table 5). *IAA18* was not included in this study since there is not and end-sequenced *JAtY* clone available that contains this gene. Expression of the transgenes was confirmed for 3 *IAA-GFP* constructs (*IAA4-GFP*, *IAA5-GFP* and *IAA19-GFP*) by RT-PCR (Figure 30). *IAA5* has been reported to be quickly induced by IAA (Conner et. al., 1990; Abel et. al., 1995) and real-time RT-PCR shows that *GFP*-tagged *IAA5* and the endogenous gene respond to auxin treatment in a similar way (Figure 31).

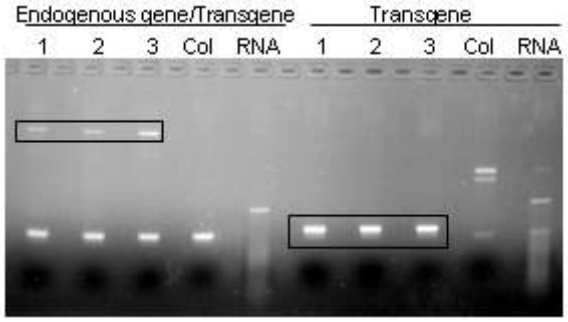
In spite of detecting the trangenes' mRNAs of several Aux/IAAs, no fluorescence was observed from any of the *GFP* constructs. There could be several reasons for this. First of all, these studies were limited to *Arabidopsis* plants at the seedlings stage where some *IAA* genes may not be expressed or might have very low expression level. In fact, based on microarray data most *Aux/IAA* genes have very low expression levels in the root (Birnbaum et. al., 2003; Stepanova et. al., 2005). Secondly, Aux/IAA proteins are extremely short-lived (Abel et. al., 1994) and do not accumulate *in vivo*. Thus, it is likely that the *GFP* used in the study is sensitive enough to detect these unstable proteins with low expression level.



A (*IAA4-GFP*)



B (*IAA5-GFP*)



C (*IAA19-GFP*)

Figure 30. RT-PCR result of GFP-tagged *IAA* gene expression. (A) Transgene expression could be detected from 3 independent *IAA4-GFP* lines (1, 2 and 3) using primer pairs specific for transgene (T) and primer pairs amplifying both the endogenous gene and transgene.(E/T)In the right part of graph, bands specific for transgene amplification using T primer pairs are boxed and their identity further confirmed by DNA sequencing. PCR bands specific for transgene when using E/T primer pairs are boxed in the left part of graph. Col: Columbia wildtype used as negative control. RNA: one of the RNA samples without reverse transcription used as negative control. (B) Expression of *IAA5-GFP* was detected from 3 independent lines before (1, 2 and 3) or after auxin treatment (1A, 2A, and 3A). Arrows indicate the bands specific for the transgene amplified using either E/T primer pairs or T primer pairs. (C) Expression of *IAA19-GFP* was detected from 3 independent lines (1, 2 and 3). Bands specific for transgene amplification using E/T and T primer pairs are boxed in the left part of graph and the right part of graph respectively.

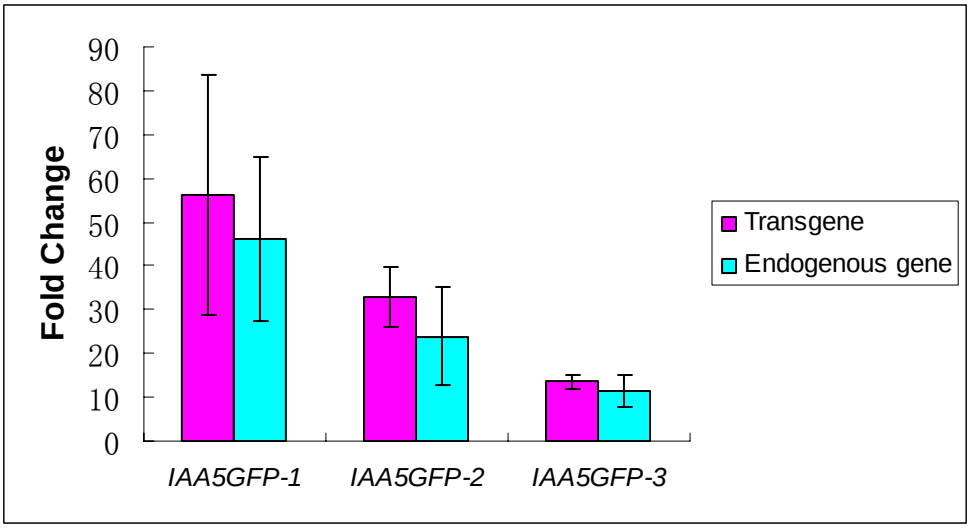


Figure 31. Induction of *IAA5-GFP* expression after auxin treatment. Intron spanning primers were designed to amplify endogenous *IAA5* gene and *IAA5-GFP* transgene. Fold change was determined by comparing the expression level in pooled *IAA5-GFP* T2 seedlings treated with 0.1μM IAA for 1h and expression level in mock-treated seedlings. The experiment was performed with 3 biological replicates (3 independent *IAA5-GFP* transgenic lines were used and the names of these lines are displayed on the bottom panels). The result indicates that *IAA5* endogenous gene and *IAA5-GFP* transgene show similar response to auxin treatment.

Tagging gain-of-function mutants of the Aux/IAA family with GFP

It has been shown that the conserved domain II is required for rapid turnover of Aux/IAA proteins by mediating interaction with SCF^{TIR1} ubiquitin-liagse complex (Gray et. al., 2001; Tian et. al., 2003). Observations from fusion of domain II to *firefly luciferase* (*LUC*) or *β-glucuronidase* (*GUS*) reveal that domain II confers instability to the reporters (Worley et. al., 2000; Gray et. al., 2001; Ramos et. al., 2001) and is sufficient to recapitulate *in vivo* degradation of fusion proteins in response to auxin (Zenser et. al., 2003). Also, *axr3-1/iaa17* with a domain II mutation has been shown to have a slower degradation and seven-fold greater lifetime compared with endogenous wild-type *IAA17* (Ouellet et. al., 2001). Up to date, dominant or semi-dominant mutations have been identified for 10 *Aux/IAA* family members (Yang et. al., 2004) and analysis of these

dominant mutants has substantially contributed to our current understanding of the *Aux/IAA*s' functions.

The core residues of domain II required for rapid degradation is present in 23 of the 29 *Arabidopsis* Aux/IAA proteins, but entirely or partially missing in IAA20, IAA30, IAA31, IAA32, IAA33, and IAA34 (Figure 32). It has been reported that proline hydroxylation is critical for Aux/IAA recognition by TIR1 receptor (Tan et. al., 2007) and that mutation of either one of the two prolines in the core sequence GWPPV to a serine stabilizes the Aux/IAA proteins (Fukaki et. al., 2002; Hamann et. al., 2002).

Since we were unable to detect any fluorescence from GFP-tagged wildtype IAA constructs, 10 dominant Aux/IAA mutations with the Pro→Ser amino acid exchange (details in Table 5) were generated by 70-mer oligonucleotide-mediated homologous recombination or recombineering using *galK* selection (see methods section). The resulting *IAA* mutants were then tagged in their C-terminus with GFP using the recombineering system (Table 5). However, no fluorescence was observed from any of the *GFP* lines generated with the domain II mutation. A possible explanation for this is that the accumulation of stabilized Aux/IAA proteins will cause detrimental auxin-related developmental phenotypes. Thus when selecting for healthy plants in the herbicide containing media, we were indirecty selected transgenic lines with relatively low level of transgene expression.

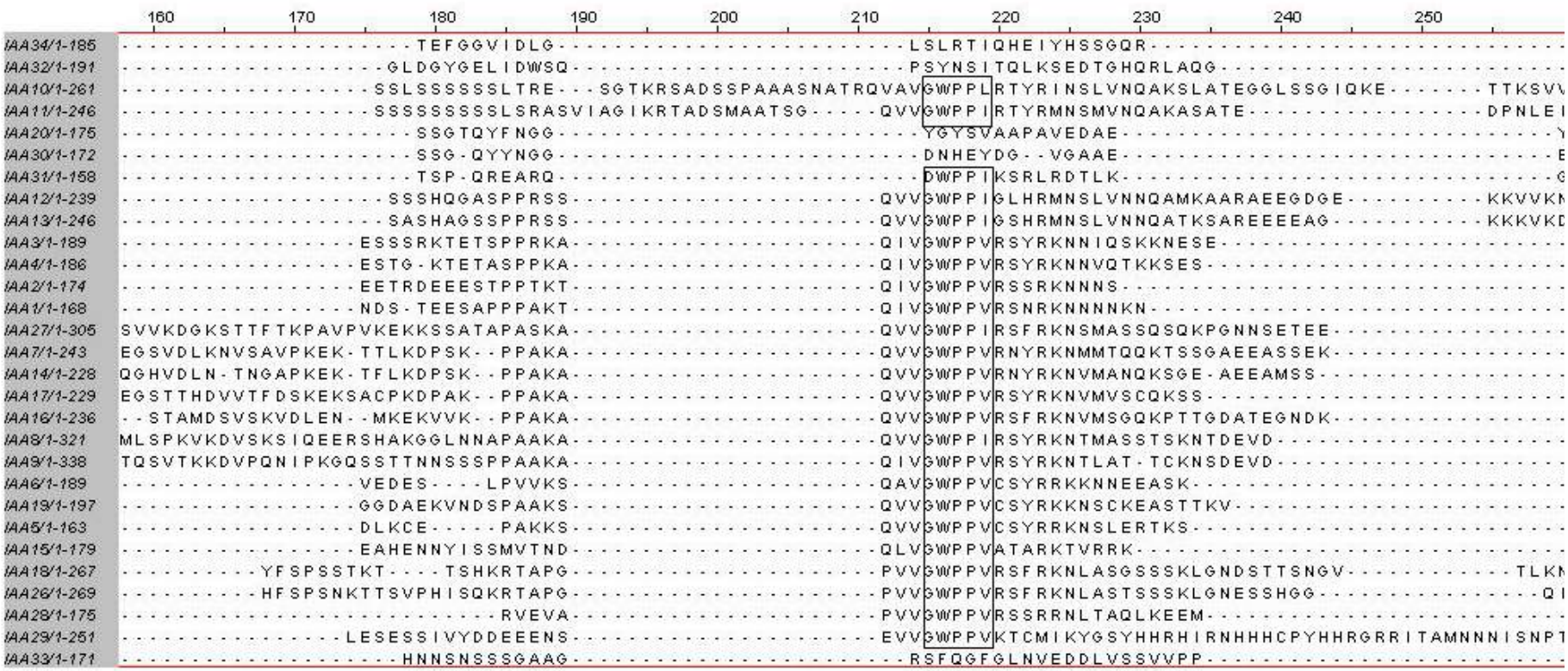


Figure 32. Protein alignment of the Aux/IAA domain II. The multi-alignment was generated by ClusterW and the image was acquired using Java view. Only amino acids around predicted domain II were included and core sequences GWPPV of domain II are boxed.

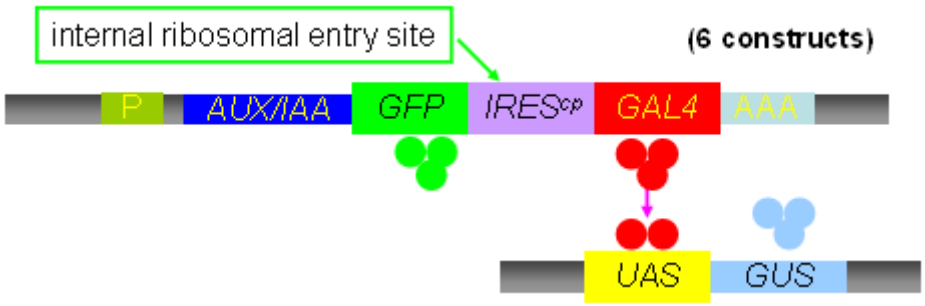
Tagging *Aux/IAA* family genes with *GFP-IRES^{cp}-GAL4*

To determine the physiological relevance of a specific gene expression pattern, *GAL4-UAS* system has been extensively employed to manipulate the expression pattern of a GOI and analyze the physiological consequences of such alterations (Phelps et. al., 1998). In most cases, however, this has been done by using randomly generated *GAL4-GFP* enhancer trap lines or transcriptional fusions that make use of the promoter of the genes of interest (Laplaze et. al., 2005; Casamitjana-Martinez et. al., 2003). In the former approach, the endogenous gene whose expression has been observed and imposed to a reporter gene is not typically known. On the other hand, in the latter case of promoter transcriptional fusions, regulatory sequences other than the promoters are usually not included. We have used a bi-cistronic *GFP-IRES^{cp}-GAL4* system to overcome the limitations of the two aforementioned systems. The general idea is to generate a bi-cistronic mRNA that will contain both the full length mRNA of the GFP-tagged GOI and the *IRES^{cp}-GAL4* RNA sequences. The levels and expression patterns of this chimeric RNA will be determined by "all the regulatory sequences" of the targeted gene (including potential miRNA-mediated regulation) that are present in the engineered *JAtY* clone. Importantly, this RNA will encode two independent proteins, the GFP-tagged protein corresponding to the endogenous gene and the GAL4 protein. The GAL4 protein would then bind to the upstream activation sequence (*UAS*) and promote the transcription of *GUS* reporter gene (or any other reporter gene under the control of the *UAS* sequences). The schematic plot of the system is shown in Figure 33A. Critical to the success of this system is that both coding regions in the chimeric RNA are efficiently translated. Among the different possible alternatives to generate this construct we have selected the internal

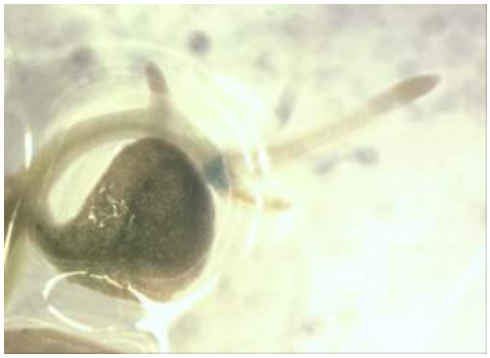
ribosomal entry site from the tobacco virus coat protein (IRES^{cp}, Skulachev et. al., 1999). This translation enhancer has been extensively tested in *Arabidopsis* and proven to be highly efficient (Yamamoto et. al., 2003). The *GFP-IRES^{cp}-GAL4* system may have several main advantages compared with the whole gene-*GFP* translational fusions described above. First, it may be more sensitive, since the GAL4 transcription factor may act as an amplification element. On the other hand, and as discussed above, the GAL4-based system, can also be used for the precise manipulation of GOI expression *in planta*. Furthermore, it may provide a practical solution in the analysis of genes with very sort-lived proteins, since the expression pattern of GAL4 would perfectly mimic that of the tagged gene, but the turnover of the GAL4 protein will not be influence by that of the GOI. And finally, the system integrates GFP tagging to offset the limitation that GAL4 itself cannot be used to examine subcellular localization.

Six *Aux/IAA* genes have been tagged with *GFP-IRES^{cp}-GAL4* at their C-terminal using the recombineering system (depicted from now on as "*GFPGAL4*" constructs). The resulting constructs (Table 5) were transformed into plants containing *UAS-GUS* reporter. As expected from our previous experiments, no GFP fluorescence could be detected, and unfortunately, GUS staining was only observed from a few transgenic T1 seedlings of 2 constructs indicating the efficiency of translation of *GAL4* cistron is too low to be detected. *IAA5-GFPGAL4* (Figure 33B) and *IAA26-GFPGAL4* (Figure 33C) were shown to be expressed at hypocotyl-root junction and upper root respectively. However, these expression patterns could not be reproduced in T2 or T3 seedlings. The most likely explanation for the failure of the system is the relatively low expression level of the targeted genes and low efficiency of the *IREScp* in this experimental context. Thus, how

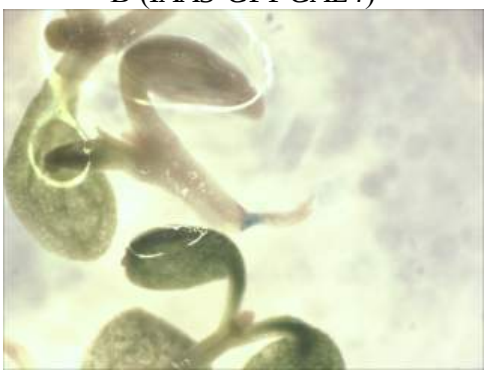
efficiently IRES^{cp} can start translation of GAL4 in *GFP-IRES^{cp}-GAL4* system still needs to be further examined.



A



B (*IAA5-GFPGAL4*)



C (*IAA26-GFPGAL4*)

Figure 33. Tagging of *Aux/IAAs* with *GFP-IRES^{cp}-GAL4* system. (A) Schematic representation of the *GFP-IRES^{cp}-GAL4* tagging cassettes. A bi-cistronic mRNA containing the *GFP*-tagged gene and the *GAL4* sequences separated by the *IRES^{cp}* will be produced mimicking the expression patter of the tagged *Aux/IAA* genes. P stands for promoter of *Aux/IAA* genes and AAA represents the polyadenylation tail at the 3' end of mRNA. Gal4 specifically binds to *UAS* and activate *GUS* gene transcription. (B-C) GUS staining of *UAS-GUS* lines transformed with *IAA5* and *IAA26* tagged with *GFP-IRES^{cp}-GAL4* at their C-terminal. Microscopy images were taken from representative T1 seedlings.

Tagging *Aux/IAA* genes with 3 copies of *Ypet* (*Typet*)

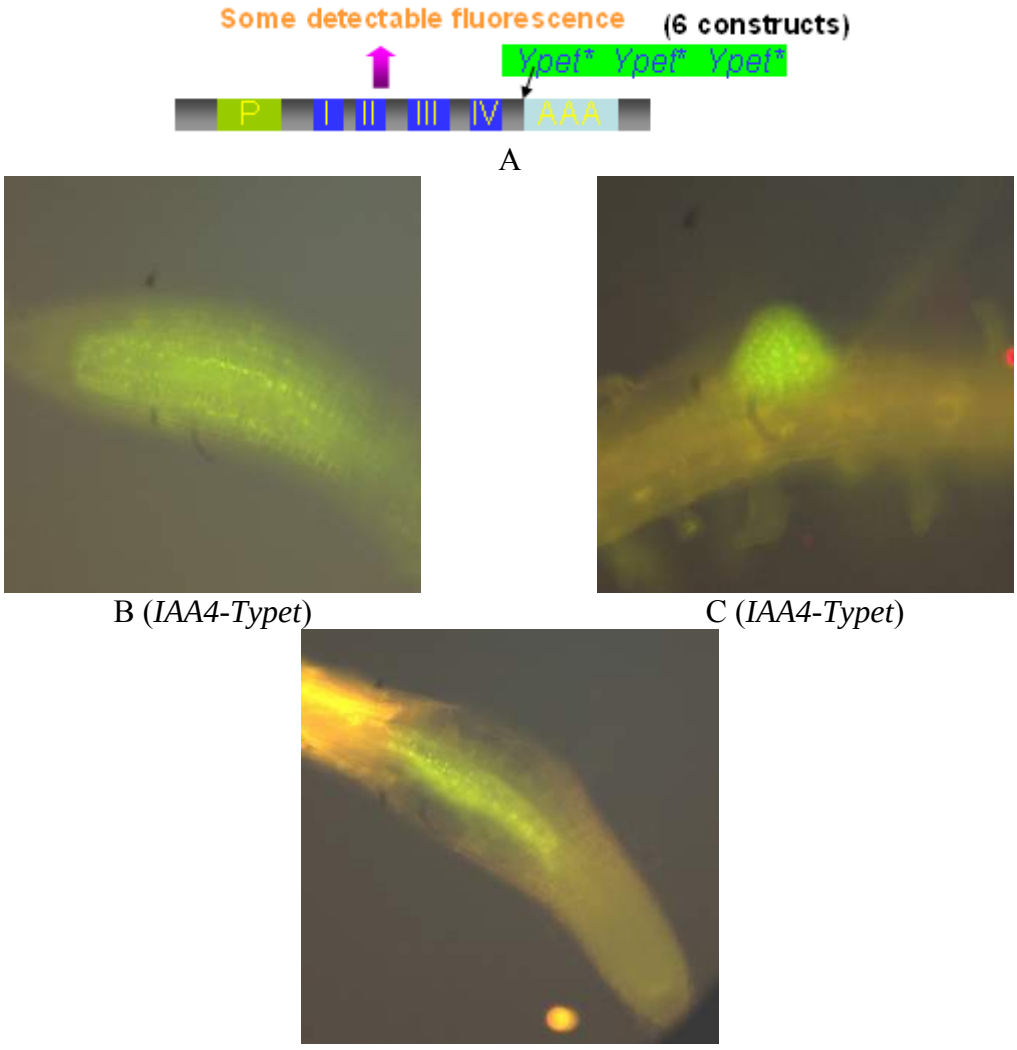
Ever since *GFP* was adapted to study protein localization, a constant search for new and improved variants has been conducted (Heim et. al., 1994; 1995). *Ypet*, an improved version of *YFP* (a derivative of *GFP*), has been shown to have reduced chloride sensitivity, faster folding and increased fluorescence intensity (Nguyen et. al., 2005). Since one of possible reason for failing to detect fluorescence from all our *GFP*-tagged *Aux/IAA* constructs is the relatively low abundance of the GFP-tagged protein, we speculated that by increasing the brightness of the fluorescent protein we may be able to detect these low-abundance proteins. To further improve our chances of success in this difficult task, the *Ypet* sequence (Nguyen et. al., 2005) was modified according to the codon usage in *Arabidopsis*. The codon-optimized *Ypet* (*Ypet**) along with linker A and G sequences used in *galK/GFP* cassettes (figure 7B) was synthesized to form a new tagging cassette, *Sypet* (Table 6). In the mean time, adaptor sequences were used to link the three individual *Ypet** (sequences slightly different) into a larger fluorescent protein (*Typet*, Table 6) and avoid possible conformational alterations of the individual *Ypets*. Six *Aux/IAA* genes were then tagged with *Typet* at their C-terminus (Table 5, Figure 34A, depicted from now on as "*Typet*" constructs). However, fluorescence could only be detected in two T1 lines of *IAA4-Typet* construct and one T1 line of *IAA5-Typet* under extreme conditions (overnight auxin treatment) (Figure 34 B-D). Therefore, improvement the intensity of fluorescence alone was not sufficient for detecting of the expression of the low abundance and highly unstable Aux/IAA proteins.

Table 6. Sequences of Sypet and Typet cassettes.

Sypet cassette
(Linker G) GGAGGTGGAGGTGGAGCT-(ypet*) ATGTCTAAGGGTGAAGAGTTGTTCACTGGTGTTGTTCCTATCCTTGTGGAGCT TGATGGTGATGTTAACGGACACAAGTTCTCTGTTTCTGGTGAAGGTGAAGGTG ATGCTACTTACGGAAAGCTTACTCTTAAGCTCCTCTGCACTACTGGAAAGCTT CCAGTTCCTTGGCCTACTCTTGTTACTACTCTCGGATACGGACTTCAATGTTTC GCTAGATACCCTGATCATATGAAGCAGCACGATTTCTTCAAGTCTGCTATGCC TGAAGGATACGTGCAAGAGAGAACCATCTTCTTCAAGGATGATGGAAACTAC AAGACTAGAGCTGAGGTTAAGTTCGAGGGTGATACTCTCGTTAACAGGATCG AGCTTAAGGGAATCGATTTCAAAGAGGATGGAAACATCCTTGGACATAAGCT CGAGTACAACTACAACTCTCACAACGTGTACATCACTGCTGATAAGCAGAAG AACGGTATCAAGGCTAACTTCAAGATCAGACACAACATTGAGGATGGTGGAG TTCAACTTGCTGATCACTACCAACAGAACACTCCTATTGGAGATGGACCTGTT CTTCTCCCTGATAACCACTACCTTTCTTACCAGTCTGCTCTCTTCAAGGATCCT AACGAGAAGAGGGATCATATGGTTCTTCTCGAGTTCCTTACTGCTGCTGGAAT TACTGAGGGAATGAACGAGCTTTATAAG- *(Linker A) GGTGCTGCTGCGGCCGCTGGGGCC*
Typet cassette

Table 6 (Continued)

(Linker G) GGAGGTGGAGGTGGAGCT -(*Ypet*1*)
ATGTCTAAGGGTGAAGAGTTGTTCACTGGTGTTGTTCCTATCCTTGTGGAGCT
TGATGGTGATGTTAACGGACACAAGTTCTCTGTTTCTGGTGAAGGTGAAGGTG
ATGCTACTTACGGAAAGCTTACTCTTAAGCTCCTCTGCACTACTGGAAAGCTT
CCAGTTCCTTGGCCTACTCTTGTTACTACTCTCGGATACGGACTTCAATGTTTC
GCTAGATACCCTGATCATATGAAGCAGCACGATTTCTTCAAGTCTGCTATGCC
TGAAGGATACGTGCAAGAGAGAACCATCTTCTTCAAGGATGATGGAAACTAC
AAGACTAGAGCTGAGGTTAAGTTCGAGGGTGATACTCTCGTTAACAGGATCG
AGCTTAAGGGAATCGATTTCAAAGAGGATGGAAACATCCTTGGACATAAGCT
CGAGTACAACTACAACTCTCACAACGTGTACATCACTGCTGATAAGCAGAAG
AACGGTATCAAGGCTAACTTCAAGATCAGACACAACATTGAGGATGGTGGAG
TTCAACTTGCTGATCACTACCAACAGAACACTCCTATTGGAGATGGACCTGTT
CTTCTCCCTGATAACCACTACCTTTCTTACCAGTCTGCTCTCTTCAAGGATCCT
AACGAGAAGAGGGATCATATGGTTCTTCTCGAGTTCCTTACTGCTGCTGGAAT
TACTGAGGGAATGAACGAGCTTTATAAG-
(Adapter1) GCTGCTGCAGCTGCTGCAGCTGCTGCAGCT-(*Ypet*2*)
ATGTCTAAGGGTGAAGAACTCTTTACCGGTGTTGTGCCTATCTTGGTTGAATT
GGACGGTGATGTGAATGGTCATAAGTTCTCAGTTAGCGGAGAGGGTGAAGGT
GACGCAACATACGGTAAGTTGACCCTCAAGCTTTTGTGTACAACTGGAAAGT
TGCCTGTTCCATGGCCAACACTCGTTACTACATTGGGTTACGGACTCCAGTGC
TTTGCTAGGTATCCAGACCACATGAAGCAACATGACTTTTTCAAGAGCGCTAT
GCCAGAGGGATATGTTCAAGAGCGTACTATTTTCTTCAAGGACGACGGTAAT
TACAAAACACGTGCTGAAGTGAAGTTTGAGGGTGACACCTTGGTGAACAGAA
TCGAGTTGAAGGGTATTGACTTTAAAGAGGACGGTAACATCTTGGGTCACAA
GTTGGAGTATAACTACAACAGCCATAACGTCTACATTACCGCAGACAAGCAA
AAAAACGGTATTAAGGCAAATTTCAAAATCCGTCATAACATCGAAGATGGTG
GTGTTCAGTTGGCAGACCATTACCAGCAAAACACACCTATCGGTGATGGTCCT
GTGTTGCTCCCAGACAACCATTACCTTAGCTACCAGAGCGCTTTGTTTAAGGA
CCCAAACGAGAAACGTGATCACATGGTGCTTTTGGAGTTTTTGACCGCTGCTG
GAATCACAGAGGGAATGAATGAGTTGTACAAG-
(Adapter2)GCAGCAGCTGCAGCTGCTGCTGCTGCAGCT-(*Ypet*3*)
ATGAGCAAAGGTGAAGAGTTATTCACGGGAGTGGTTCCAATTTTGGTGGAGT
TGGATGGGGACGTCAACGGTCATAAGTTTTCTGTGAGTGGAGAAGGCGAGGG
TGATGCAACATATGGAAAACTCACCCTCAAGTTGCTCTGTACTACCGGTAAAC
TTCCTGTGCCATGGCCTACTTTGGTTACAACCCTTGGATATGGATTGCAATGC
TTCGCTCGTTATCCGGATCACATGAAGCAGCACGACTTTTTTAAGTCAGCCAT
GCCAGAAGGTTACGTTCAGGAAAGGACGATCTTTTTTAAGGACGACGGGAAC
TATAAGACAAGGGCAGAAGTGAAGTTCGAAGGGGATACCCTTGTTAATCGTA
TCGAACTCAAAGGCATTGACTTCAAAGAAGATGGGAATATCCTCGGGCACAA
ACTTGAGTACAATTATAACAGCCACAATGTTTACATCACGGCCGACAAACAA
AAGAATGGGATTAAGGCCAACTTTAAGATTAGGCATAATATCGAGGATGGGG
GAGTGCAATTGGCCGATCATTATCAGCAGAATACCCCAATTGGTGATGGACC
AGTGCTTTTGCCGGATAACCATTACTTGTCCTACCAATCTGCTCTTTTTAAAGA
CCCGAATGAGAAGCGTGACCATATGGTTTTGTTGGAATTCCTCACAGCTGCTG
GTATTACCGAAGGCATGAATGAACTTTACAAG-
(Linker A) *GGTGCTGCTGCGGCCGCTGGGGCC*



A

B (*IAA4-Typet*)

C (*IAA4-Typet*)

D (*IAA5-Typet* + overnight auxin treatment)

Figure 34. Tagging of *Aux/IAA* genes with *Typet*. (A) Structure of constructs. *Typet*, composed of three codon-optimized Ypet (Ypet*) was introduced right before the stop codon of the *Aux/IAAs*. (B) Fluorescent microscopy image of *IAA4-Typet Arabidopsis* root. (C) Expression of *IAA4-Typet* at a lateral root initiation site. (D) Expression of *IAA5-Typet* in *Arabidopsis* root with overnight auxin treatment.

Tagging of dominant *aux/iaa* mutants with part of domain IV replaced by *Typet*

As discussed before, the attempts to tag dominant mutants of *Aux/IAA* gene were unsuccessful the transgenic plants with high expression levels of stabilized Aux/IAA protein are unlikely to survive the selection process. However, it has been shown that increased repressor activity of mutations in domain II can be partially suppressed by a second site mutations in domain I and III of Aux/IAA proteins (Tiwari et. al., 2001).

Importantly for us, in a suppressor screen of the dominant *axr3/iaa17-1 mutant*, the strongest intramolecular suppressor turned out to correspond to a deletion of half of domain IV of the protein (Rouse et. al., 1998). Since domain IV is also involved in dimerization of Aux/IAA proteins and conserved across the whole family (Figure 35), we speculated that, if domain IV of a stabilized Aux/IAA protein is partially disrupted, the ability of this protein to interact with other Aux/IAA proteins and with ARFs would be reduced, reversing the increased transcriptional repression activity of domain II dominant mutations in these *Aux/IAA* genes. To test this possibility, four constructs were made by mutating the second proline to serine in the core sequence GWPPV of domain II (depicted from now on as "M2" mutations) and then replacing 22 amino acids (Table 5) in conserved domain IV by *Typet* (indicated as "M2R4-Typet" mutations) in *IAA4*, *IAA5*, *IAA19* and *IAA26* (Figure 36 A).

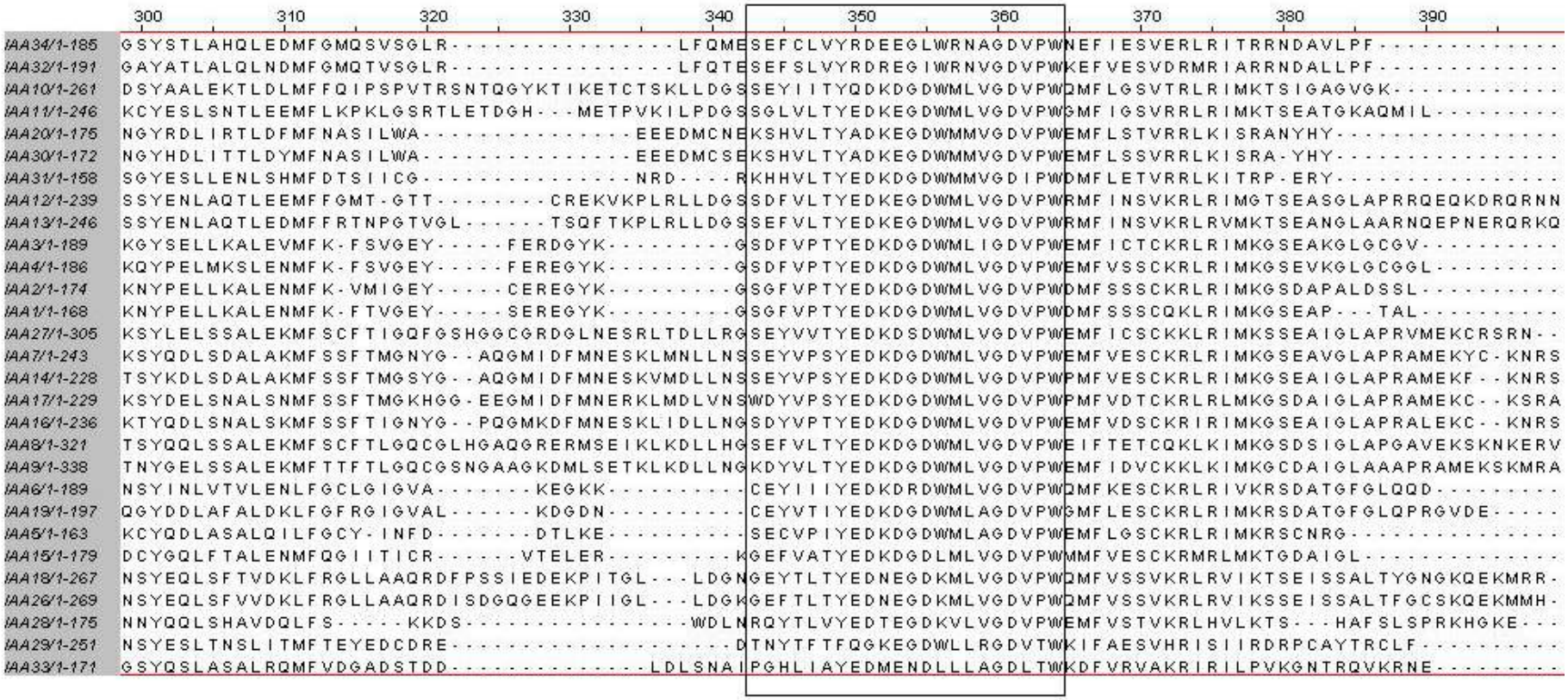


Figure 35. Protein alignement of domain IV of Aux/IAA proteins. The multialignment was generated by ClusterW and the image was acquired using Java view. Amino acids of predicted domain IV and part of domain III were included. Box in the figure displays 22 amino acids which are conserved in domain IV across Aux/IAA family and were targeted in R4-Typet mutations.

Strong fluorescence could be detected from the *IAA4M2R4-Typet*, *IAA19 M2R4-Typet*, and *IAA26M2R4-Typet* transgenic lines (Figure 36 E-G). For the *IAA5M2R4-TYpet* lines, fluorescence could only be detected after auxin treatment (Figure 37 A-B). The expression pattern of *IAA4*, *19* and *26M2R4-Typet* constructs are consistent with reconstructed spatiotemporal gene expression data (Cartwright et. al., 2009) and both *IAA19M2R4-Typet* and *IAA5M2R4-Typet* can be induced by auxin treatment in a similar manner as the corresponding endogenous gene (Figure 37C). These results clearly indicate recombineering-mediated tagging system can be used to explore expression pattern of the low-expressed and short-lived Aux/IAAs proteins. Although these proteins can only be tagged in their mutated form, which may not reflect the authentic accumulation pattern, the fact that targeted GOI is regulated in its full genomic context makes this approach probably the best practical way to examine the expression of Aux/IAA family and their response to auxin.

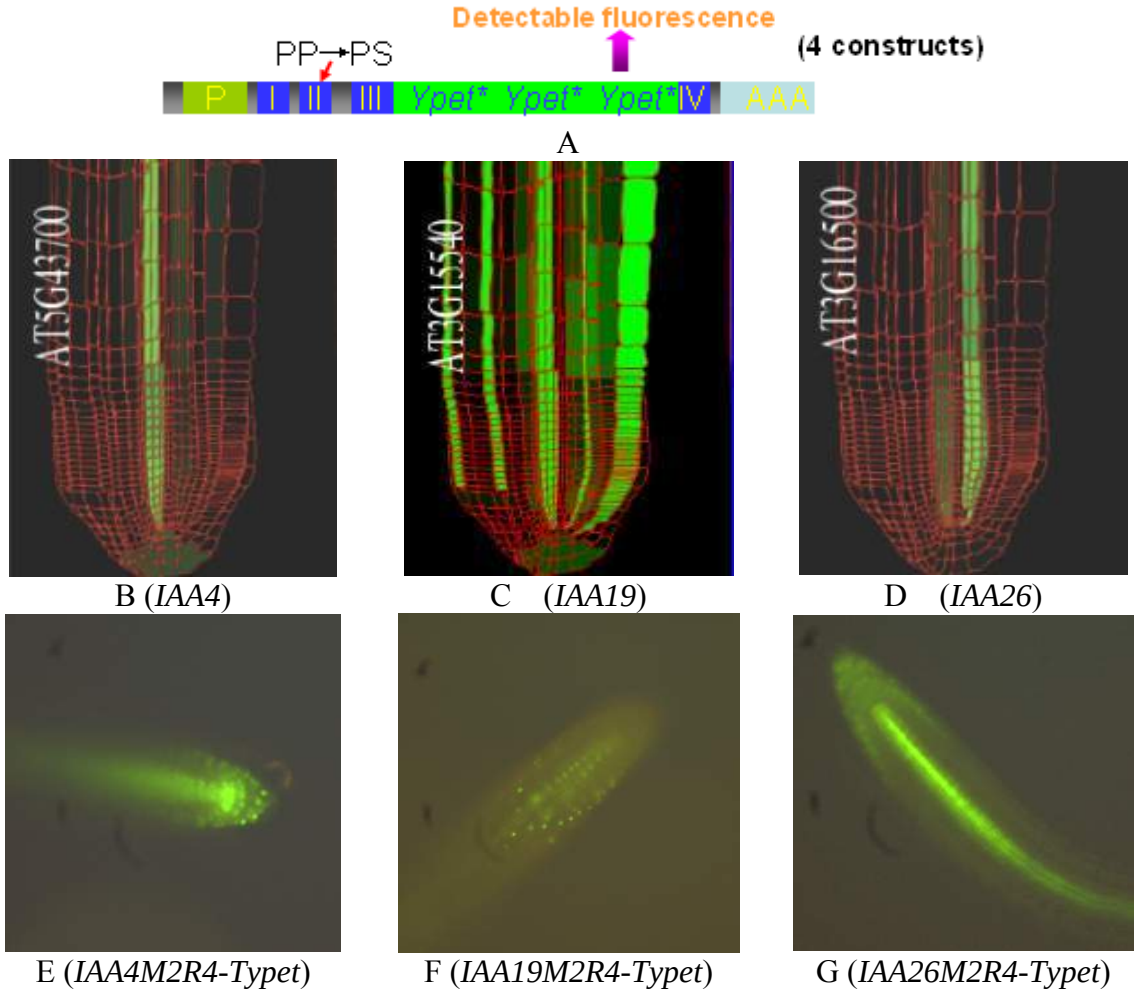
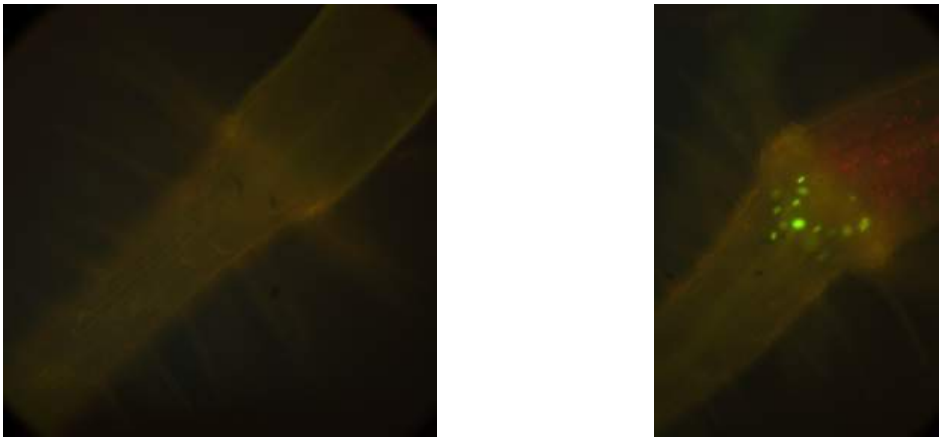
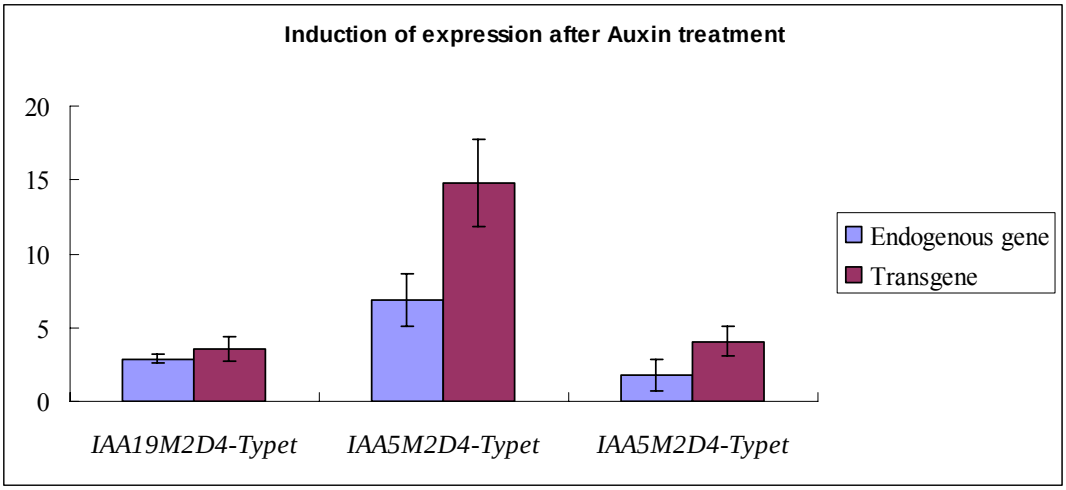


Figure 36. Tagging of dominant Aux/IAA mutants with part of domain IV replaced by Typet. (A) Structure of the constructs. The second proline in conserved domain II was mutated to serine and *Typet* was introduced within domain IV to replace 22 amino acids conserved across the Aux/IAA family (Table 5). (B-D) Predicted expression pattern of *IAA4*, *IAA19* and *IAA26* in *Arabidopsis* root. The images were generated with Deconvolved Root Expression Map Visualizer (Cartwright et. al., 2009). (E-G) Fluorescent microscopy images of *IAA4M2R4-Typet*, *IAA19 M2R4-Typet* and *IAA26 M2R4-Typet* roots. Images were taken from representative T2 seedlings.



A (*IAA5M2R4-Typet*) B (*IAA5M2R4-Typet*+ 3h auxin treatment)



0.1μM IAA/1h 0.1μM IAA/1h 0.1μM IAA/3h

C

Figure 37. Induction of *IAAM2R4-Typet* expression by auxin treatment. (A-B) Fluorescent microscope image of *IAA5M2R4-Typet* expression at the hypocotyl-root junction before and after 3h auxin treatment. (C) Real-time PCR results of auxin-induced *IAA19M2R4-Typet* (1h treatment) and *IAA5M2R4-Typet* (1h and 3h treatment). For both endogenous gene and transgene, fold change was determined by comparing expression level in pooled *IAAM2R4-Typet* T2 seedlings treated with 0.1μM IAA and 0.1% ethanol for the same amount of time (1h or 3h). A housekeeping gene *AT5g22440* was used for normalization. Three biological replicates (3 independent T2 seedling pools) were performed. Average and standard errors of the fold changes are shown in the figure. The names of constructs and corresponding treatment are displayed on the bottom panels.

Subcellular localization of Aux/IAA proteins

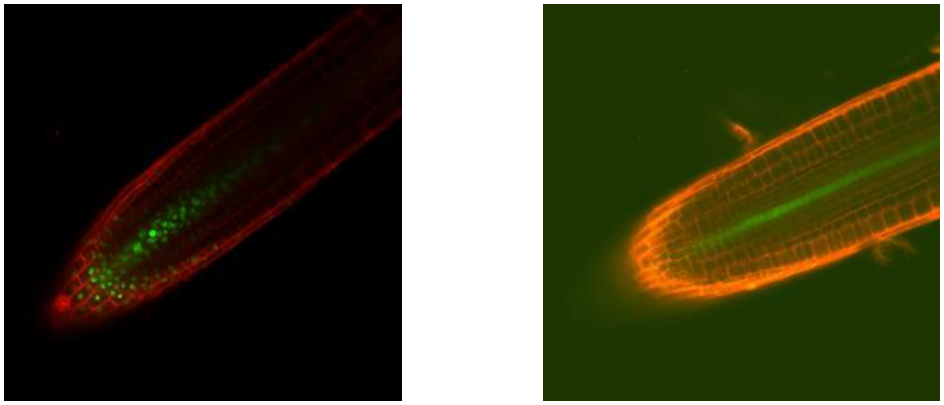
Nuclear-localization sequences (NLS) have been found at the end of the conserved domain IV of Aux/IAA proteins (Abel et. al.1994). A small basic patch containing at least

one Lys residue and often directly in front of an Arginine located between domain I and II has also been shown to be part of a bipartite NLS (Dreher et. al., 2006). The nuclear localization of Aux/IAA proteins has been experimentally verified by cDNA fusions with GFP (Abel et. al., 1994; 1995; Tiwari et. al., 2001; Dreher et. al., 2006; Muto et. al., 2007). However, in all these studies the fusion proteins were driven by 35S promoter, and therefore, the observed localization may or may not accurately reflect that of the endogenous protein. Due to the low expression level and rapid turn-over, no studies have reported the localization of "wildtype" Aux/IAA proteins using translational fusion driven by native promoter except that of *slr/iaa14* that under the control of *IAA14* promoter was shown to localize in the nucleus (Fukaki el. al., 2005).

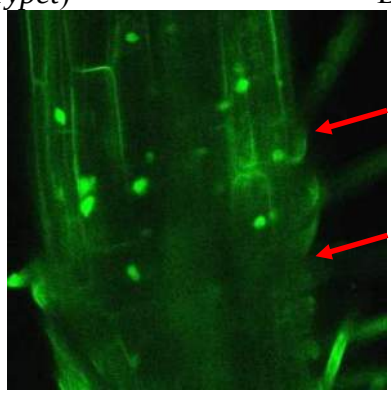
Our *M2R4-Typet* constructs provide a robust way to examine the expression pattern of the *Aux/IAA* genes and subcellular localization of corresponding tagged proteins. As shown in Figure 38, *IAA4M2R4-Typet* is expressed in root cap, quiescent center and vasculature cells; *IAA26M2R4-Typet* is expressed in vasculature cells; *IAA5M2R4-Typet* is expressed at hypocotyl-root junction and also in root epidermis cells; finally *IAA19M2R4-Typet* is expressed at pericycle cells, which is consistent with previous studies (Tatematsu et. al., 2004).

Furthermore, our results also suggest that IAA4, IAA5 and IAA19 are nuclear localized, as it was expected form the predictions (Figure 38, Figure 39). The subcellular localization of IAA26M2R4-Typet was difficult to asses due to the expression of this gene in the vasculature tissues where cells are very small in size and far from the root surface, what makes it difficult to image with the confocal. In summary, the strategy of tagging the Aux/IAA gene family by mutating domain II and replacing part of domain IV worked efficiently and reliably, and was possibly the best way to determine the

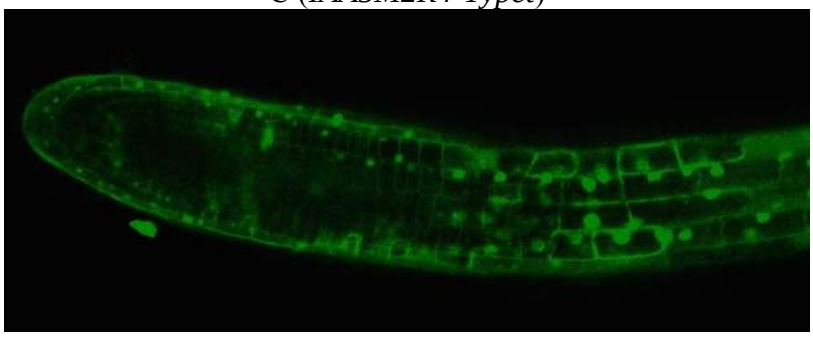
expression pattern and subcellular localization of the Aux/IAAs. Hence, the same strategy was employed for tagging the rest of the Aux/IAA family members except that *Sypet* instead of *Typet* was used (Table 5). Four *Aux/IAAs*, *IAA20*, *IAA30*, *IAA33* and *IAA34* were not included in these experiments because they do not show a conserved domain II. The information provided by this type of fusion protein would dramatically improve our understanding of the spatial-temporal expression pattern of Aux/IAA gene family and the sub-cellular localization of corresponding protein products. Combination of this information with precise maps of auxin response levels (readily obtainable using *DR5-GFP* lines) and detailed spatial-temporal expression of the ARF-GFP (constructs already made) will be fundamental in better understanding the molecular mechanism of auxin response.



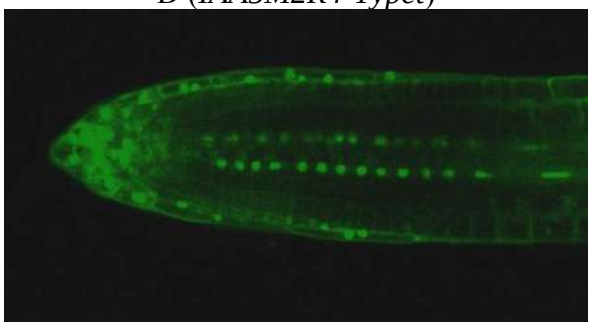
A (*IAA4M2R4-Typet*) B (*IAA26M2R4-Typet*)



C (*IAA5M2R4-Typet*)



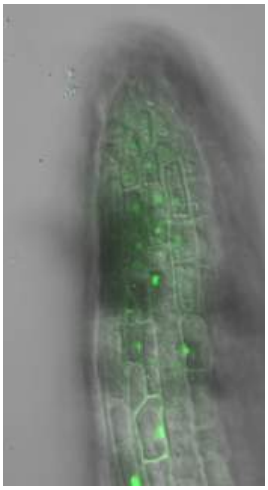
D (*IAA5M2R4-Typet*)



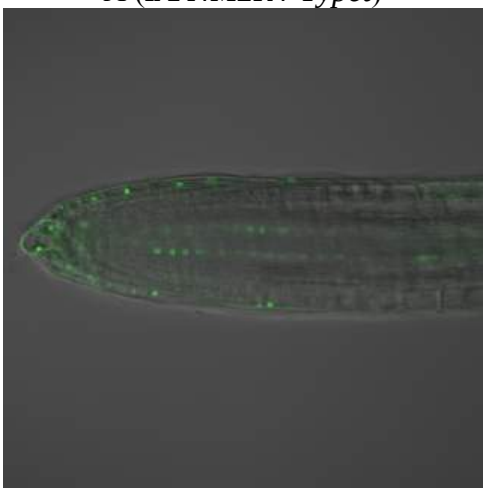
E (*IAA19M2R4-Typet*)

Figure 38. Confocal images showing expression of *IAA-M2R4-Typet* constructs. A-B: Expression of *IAA4M2R4-Typet* and *IAA26M2R4-Typet* in *Arabidopsis* roots. Propidium iodide staining (in red, showing the cell wall) and *Ypet* fluorescence (in green) were combined. C-D: Expression of *IAA5M2R4-Typet* in the hypocotyl-root junction and root respectively. E: Root expression of *IAA19M2R4-Typet*.

A (*IAA4M2R4-Typet*)



B (*IAA5M2R4-Typet*)



C (*IAA19M2R4-Typet*)

Figure 39. Confocal images showing the expression of *IAA4M2R4-Typet*. A: *IAA5M2R4-Typet*. B:*IAA19M2R4-Typet*. C: in *Arabipdosis* roots. DIC images and Typet fluorescent images were combined by selecting the DIC images as the background.

Expression of *Aux/IAA* genes in the lateral root initiation site

It has been shown that auxin acts as a local morphogenetic trigger and its local accumulation in root pericycle cells is a necessary and sufficient signal to respecify these cells into lateral root founder cells (Dubrovsky et. al., 2008). Although still incomplete, the molecular pathways by which auxin triggers lateral root formation has been elucidated and many of the signal transduction components that mediate auxin-regulated lateral root initiation have been identified and characterized. Gain-of-function *slr-1/iaa14* mutant

(Fukaki el. al., 2005) does not form lateral root as a result of pericycle founder cells failing to undergo formative divisions. On the other hand, the loss-of-function *iaa14-1* (Okushima et. al., 2005) mutant has no obvious phenotype, suggesting that other members of *Aux/IAA* gene family might also be involved. The gain-of-function mutants *iaa1/axr5* (Yang et. al., 2005), *shy2-2/iaa3* (Tian et. al., 1999), *iaa28-1* (Rogg et. al., 2001), and *iaa19/msg2* (Tatematsu et. al., 2004) are all impaired in lateral root development, however none of these mutants can block lateral root initiation completely. We observed that *IAA26M2R4-Typet* (Figure 40) and *IAA4M2R4-Typet* (Figure 41) are also expressed at lateral root initiation sites suggesting a possible role for these genes in this important auxin regulated process. Moreover, *IAA4M2R4-Typet* is consistently expressed during all the stages of lateral root development including initiation, patterning and emergence, activation of new meristem and elongation of new root (Figure 41). Once again, these results demonstrate that gene expression and protein subcellular localization information provided by recombineering-based gene tagging can substantially enhance our understanding of how these proteins are involved in auxin-mediated processes and how they regulate numerous aspects of growth and development in *Arabidopsis*.

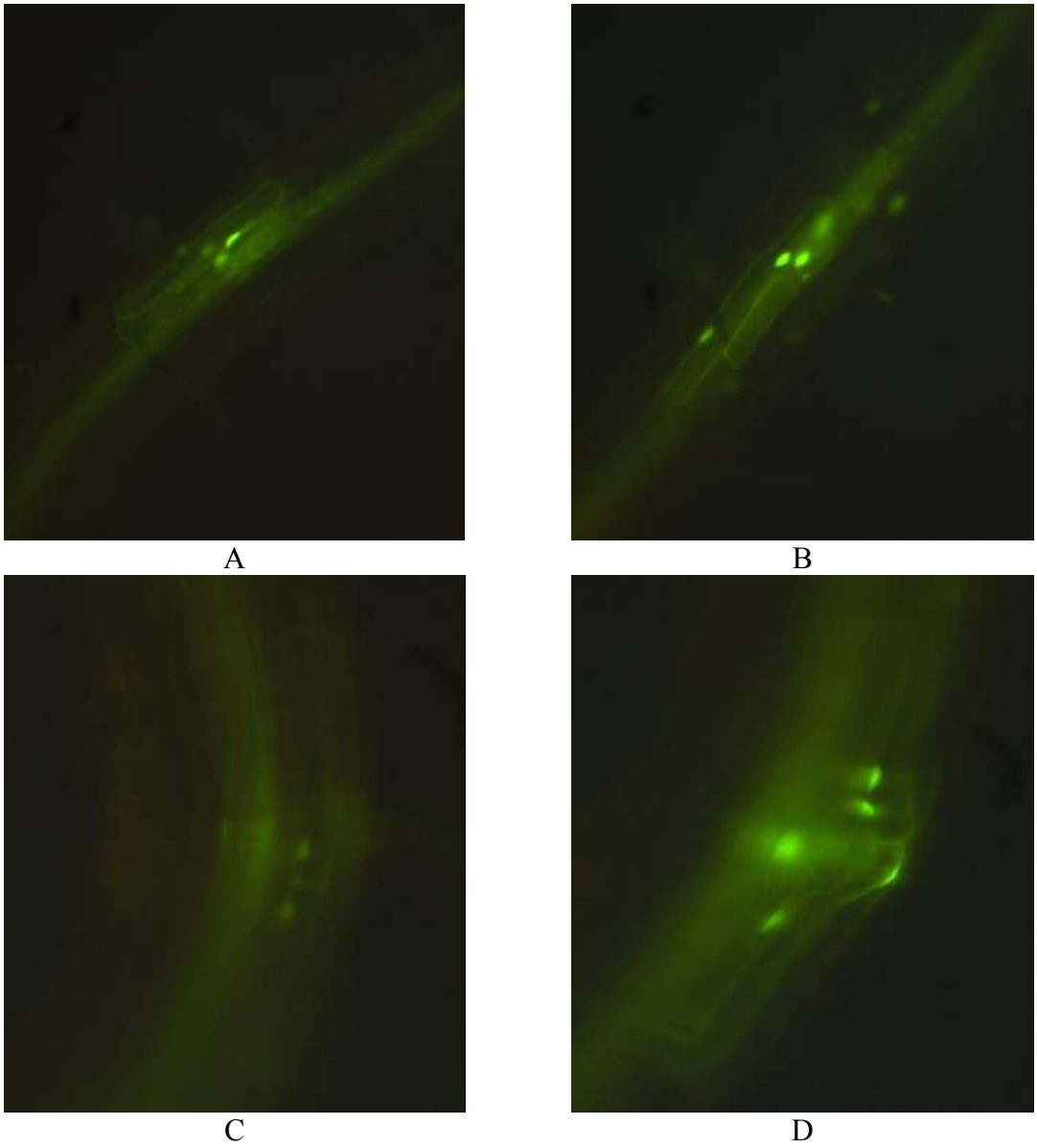


Figure 40. Expression of *IAA26M2R4-Typet* at the lateral root initiation sites. Fluorescent microscopy images at different stages of the lateral root development are shown (A-D, younger stages to old stages).

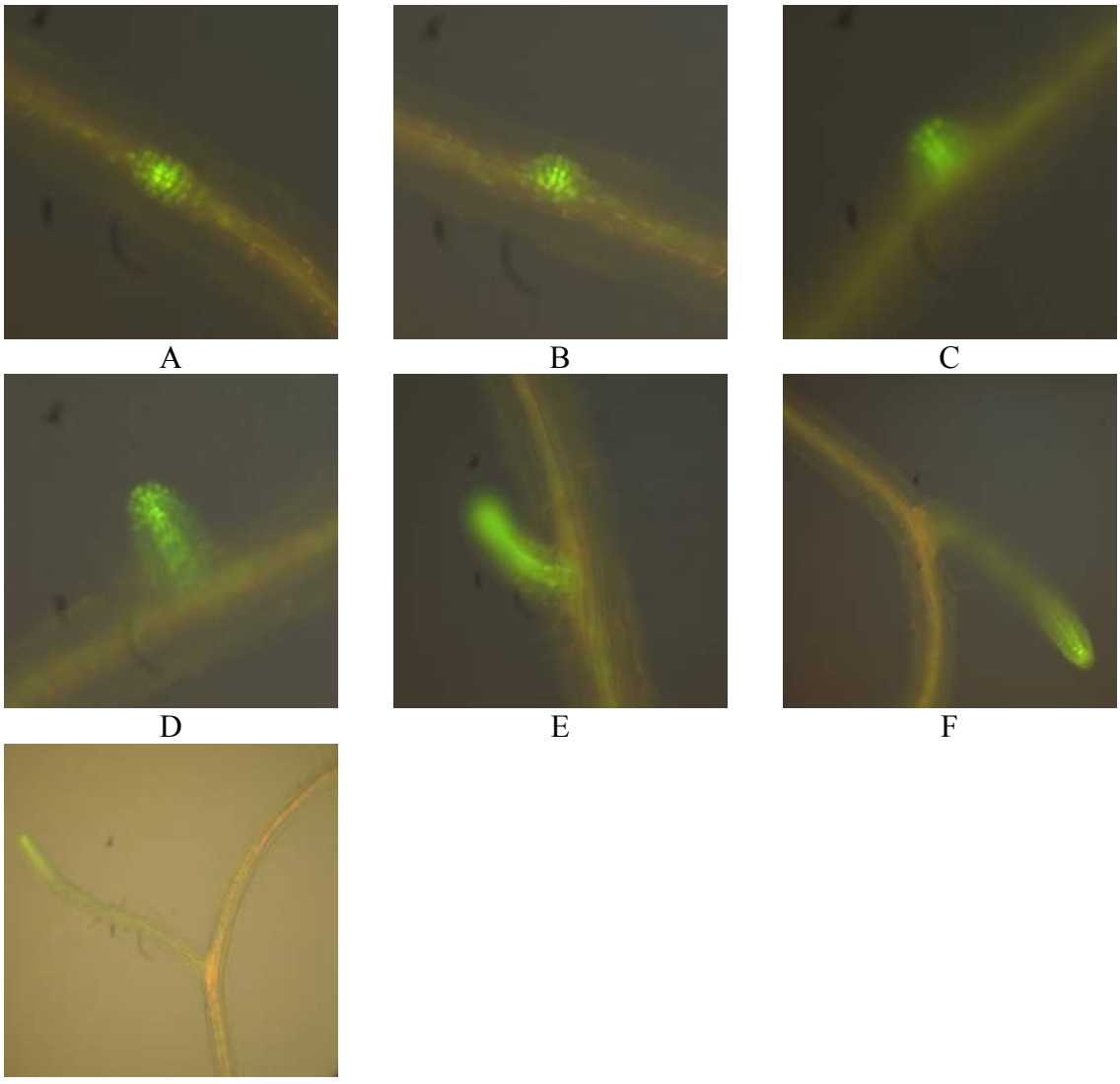


Figure 41. Expression of *IAA4M2R4-Typet* during lateral root development. Fluorescent microscopy images of lateral roots at different developmental stages are shown (A-H, younger stages to old stages).

MATERIALS AND METHODS

Reporter gene constructs

IAAM2-GFP constructs were made by introducing a single amino acid change in domain II of *IAA-GFP* constructs (Table 5). *IAA1M2-GFP*, *IAA12M2-GFP*, *IAA14M2-GFP*, *IAA3M2-GFP* and *IAA2M2-GFP* were generated by oligonucleotide-mediated homologous recombination to substitute a proline with serine (70-mer oligo sequences are

listed in Table 7). To achieve the highest recombination efficiency, oligos were designed to form four mismatch base pairs when bonding the target sequence with the priority of C-C>T-T, T-C, A-G>A-C, A-A, G-G, G-T. The actual sequence changes made in each case are listed below: *IAA12*, CCACCA→AGCCCT; *IAA1*, CCTCCA→CCGAGC; *IAA14*, CCACCG→CCCAGT; *IAA3*, CCACCA→CCCAGC; *IAA2*, CCACCA→CCCAGC. The recombinants were identified by PCR with primer pair RECTEST-F/OUTTEST-R or RECTEST-R/OUTTEST-F and the mutations were confirmed by DNA sequencing with PCR products generated with primer pair OUTTEST-F/OUTTEST-R (Table 7). The rest of *IAAM2-GFP* constructs were made by recombination with *galK* selection. The first recombination was conducted using the *galK* universal cassette. The 69-mer forward oligos (MGAL-F, Table 8) include (5'-3'): 50 bp homologous to the sequence upstream the second proline codon, a single nucleotide T and the 18 bp homology to linker G sequence. The 74-mer reverse oligos (MGAL-R, Table 8) include (5'-3') 50 bp homologous to the sequence downstream the first nucleotide in the codon of the second proline, and 24 bp of homology to linker A. The *galK* insertions were tested with primer pairs MUTTEST-F/MUTTEST-R (Table 8). The second recombination (removing *galK* insertion) was conducted with the PCR product amplified using MUTTEST-F/MUT-R primer pairs (Table 8) and *Arabidopsis* Columbia genomic DNA as template. 60-mer long oligos MUT-R are composed by (5'-3'): 9 bp homologous to sequence upstream the codon of the second proline, a single nucleotide A and 50 bp downstream the first nucleotide in the second proline codon. Removal of *galK* was tested with PCR using MUTTEST-F/MUTTEST-R pair and mutations were confirmed by DNA sequencing.

The *IRES^{cp}-GAL4* construct was obtained from Genomic Sciences Center, Saitama, Japan. *GFP* was subcloned in frame before the *IRES^{cp}* and the upstream Linker G, downstream

Linker A were added to this tag. *GFP- IRES^{cp}-CAL4* cassette amplified with REC_F and REC_R primers (Table 4) were used to replace *galK* inserted at the C-terminal of the selected genes using recombination system. The *GFP-IRES^{cp}-GAL4*-tagged *JAtY* clones were then used to transform hemizygote UAS:GUS lines.

The *IAAs-M2R4-Sypet* (Table 5) and *IAAs-M2R4-Typet* constructs were made in two consecutive steps. In first step, mutations of domain II in original TACs containing target gene were generated either with oligonucleotide-mediated homologous recombination (Table 7) or recombination with *galK* selection (Table 8). In second step, 66 nucleotides in domain IV (Table 5) were replaced by the *Typet* or *Sypet* cassette through recombination with *galK* selection. *galK*, *Sypet* and *Typet* cassettes were amplified using R4-F and R4-R oligos (Table 9). R4-F oligos contain 50 bp homologous to the sequence upstream deletion region and 18 bp of homology to linker G. R4-F oligos are composed by 50 bp homologous to the sequence downstream deletion region and 24 bp of homology to linker A. The positive colonies were confirmed by PCR using primers D4-TF and D4-TR (Table 9). *For IAA-M2R4-Sypet* constructs, DNA fragments were amplified and sequenced using primers D4-TF and D4-TR. For *IAA-M2R4-Typet* constructs, DNA fragments were amplified with primers D4-TF and D4-TR and sequenced with D4-TF, D4-TR, Typet-seq-F1 (5'-AGCTATGTCTAAGGGTGAAGAACTC-3'), Typet-seq-F2 (5'- CTGCTGCTGCTGCAGCTATGAGC-3'), Typet-seq-R1 (5'-TCACCTTCACCCTCTCCGCTAACT-3'), and Typet-seq-R2 (5'-CACCCTCGCCTTCTCCACTCACAG-3') primers.

Real-time RT-PCR

Plates with seeds were stratified for 2 days at 4°C in the dark, exposed to light for 1 hour at room temperature and then incubated for 3 d in the dark at 22°C. For the auxin

experiments, plates with 3-day-old etiolated seedlings were opened under safe green light (Kodak green filter KOFSLl3810/1521632) and sprayed with 1 μM IAA in a 0.01% ethanol solution or with 0.01% ethanol alone. Seedlings were then incubated for an additional 1 hour or 3 hours at 22°C in the dark.

Total RNA was extracted following acid guanidinium thiocyanate-phenol-chloroform extraction (Chomczynski et. al., 1987). 400 ng of each RNA sample were reverse transcribed in a 20 μl volume using TagMan reverse transcription reagents (Applied Biosystems, CA) according to the manufacturer's recommendations.

Intron-spanning forward primers were designed to avoid co-amplification of genomic DNA. For the RT-PCR analysis of *IAAs-GFP* expression, the PCR primers RT-F and RT-R (Table 10) were used to amplify both transgenes and endogenous genes. The PCR primers RT-F and GFPRNATR-1 (5'-TTGTGCCCATTAACATCACC-3') were used to amplify transgenes. For anxin-induced expresson of *IAA5-GFP* and *IAAs-M2R4Typet*, real-time PCR was performed using an ABI 7900 System (Applied-Biosystems, CA) in the presence of SYBR-green according to the manufacturer's recommendations. All reactions were repeated 3 times and the median Ct values were used for further analysis. A housekeeping gene, At5g44200, encoding nuclear cap binding protein, CBP20, was employed as an internal control (At5g44200F: 5'-AATCGCCATGGAAGAGGAGA-3' and At5g44200R: 5'-GAATCGTGGGTTCTTCTCCG-3'). The fold change of transgene and endogenous gene before and after auxin treatment was calculated with equation: $2^{\Delta Ct_{gene}}/2^{\Delta Ct_{control}}$ and $\Delta Ct = Ct_{auxin} - Ct_{mock}$ (Livak et. al., 2001). Experiments were carried out for three independently prepared total RNA samples. The induction by auxin of *IAA5-GFP* construct was examined using the same primers as for RT-PCR. For *IAA5M2R4-Typet*, primers 5'- CGGATGCTATATCAATTTTGATGA -3'and 5'-

TGTAAGGCTCACTCACATTCAC -3' were used to amplify endogenous gene, whereas primers 5'- CGGATGCTATATCAATTTTGATGA -3' and 5'-TTTCCGTAAGTAGCATCACC-3' were used to amplify transgene. For *IAA19M2R4-Typet*, primers 5'- GTTTCCGTGGCATCGGTG -3'and 5'-TTAAATTAATGAACCAGCTCCTTG -3' were used to amplify endogenous gene, whereas primers 5'- GTTTCCGTGGCATCGGTG -3' and 5'-TTTCCGTAAGTAGCATCACC-3' were used to amplify transgene.

Microscopy

Fluorescence was imaged with a LSM 710 confocal workstation (Carl Zeiss MicroImaging Inc, Germany). GFP was excited at the 488 nm and propidium iodide was excited at the 543 nm. Emission signals were detected at 500 to 530 nm for *GFP* and >560 nm for propidium iodide by sequential scanning.

For GUS staining, seedling were fixed with ice-cold 90% acetone and then stained for GUS activity by incubation in X-Gluc substrate solution containing 20 mg/ml X-Gluc,12.5 mM $K_3Fe(CN)_6$, 12.5 mM $K_4Fe(CN)_6$, 10% Triton X-100 and 50 mM NaH_2PO_4 (pH7.0) at 37°C for 16 h. GUS staining was then examined with Zeiss Axioplan (Carl Zeiss MicroImaging Inc, Germany) and images were acquired by SPOT Insight™ (SPOT™ Imaging Solutions, MI).

Table 7. Primers used for generating IAA dominant mutations by oligonucleotide-mediated homologous recombination.

NAME	RECOMBINATION OLIGO	RECTEST-F	RECTEST-R	OUTTEST-F	OUTTEST-R
IAA12	ACCAAACTGTTCATCCTGTGTAACCCAATAGGGCTCCATCCTACCACTTGACTGCCACAACAAACCAAGA	CAGTCAAGTGGTAGGATGGAGC	CCTGTGTAACCCAATAGGGCT	TCTTCCTCTCACCAAGGAGC	AACATCTCCAACAAGCATCC
IAA1	TTTGTTGTTGTTGTTCTTACGGTTAGATCTCACGCTCGGCCATCCAACGATTTGTGTTCTGAAGTTTACA	AAATCGTTGGATGGCCGAGC	TACGGTTAGATCTCACGCTC	CACCAATGGGCTTAACCTTA	GAATTTGTTTTTGCCTCGAC
IAA14	ATTAGCCATAACATTTTTCCGGTAGTTCCTCACACTGGGCCAACCCACCACTTGTGCTCTGCATAGAATA	AAGTGGTGGGTTGGCCCAGT	TCCGGTAGTTCCTCACACTG	GACCCTTCTAAGCCTCCTGC	CCATCACTTTACTCTCGTTCA
IAA3	ACTCTGAATGTTGTTCTTCCTGTAAGATCTAACGCTGGGCCATCCAACAATCTGAGCCCTGCAAAAAGGA	AGATTGTTGGATGGCCCAGC	TCCTGTAAGATCTAACGCTG	AGAGCTGAGGCTGGGATTAC	CCTGCAACAACAAACAGAACC
IAA2	CACACTGTTGTTGTTCTTACGGGAAGATCTCACGCTGGGCCAACCAACGATTTGAGTTCTGTAAATTTGA	AAATCGTTGGTTGGCCCAGC	TACGGGAAGATCTCACGCTG	AAAGCAATGGCGTACGAGAA	ACTCCGATCCGAATCAAACC

Table 8. Primers used for generating IAA dominant mutations by homologous recombination using galK selection.

GENE	MGAL-F	MGAL-R	MUT-R	MUTTEST-F	MUTTEST-R	SEQ-F
IAA7	CCTATCTAAATAATGTTGTTCTATGTAGAGCACAAGTGGTGGGATGGCCATGGAGGTGGAGGTGGAGCT	ACTACTGGTCTTCTGCTGAGTCATCATGTTCTTCCTGTAGTTCCTCACAGGGCCCCAGCGGCCGCAGCAGCACC	ACTACTGGTCTTCTGCTGAGTCATCATGTTCTTCCTGTAGTTCCTCACAGATGGCCATCC	CCAATACATACATGCGTACAAGC	GGCCAATGCATCAGAAAGAT	TGCATAACTTTGATTCTTTGAAGT
IAA8	GTGGTAATCTGTGTCTATTGTTTTGCAGGGCACAGGTTGTTGGTTGGCCTTGGAGGTGGAGGTGGAGCT	GTTCTTCGAAGTAGAAGAAGCCATTGTATTCTTCCGGTATGATCTGATTGGGCCCCAGCGGCCGCAGCAGCACC	GTTCTTCGAAGTAGAAGAAGCCATTGTATTCTTCCGGTATGATCTGATTGAAGGCCAACC	CTCATGCTAAGGGTGGCTTG	AGGGTGAAGCAGCTGAACAT	TTGCTGACACTTGGGATGAG
IAA9	GTGTTTTTTCGGATTCTGGAATTTTCAGGGCACAAATTGTCGGTTGGCCTTGGAGGTGGAGGTGGAGCT	ACTGTTCTTACAAGTAGTGGCCAATGTGTTCTTCCTGTAGGATCTCACTGGGCCCCAGCGGCCGCAGCAGCACC	ACTGTTCTTACAAGTAGTGGCCAATGTGTTCTTCCTGTAGGATCTCACTGAAGGCCAACC	TGCCACAAAACATACCCAAA	CTGAAGAAAGCTCCCCGTAG	AGAACAATGCATCGGGAAAC
IAA10	GATTTAATATCTTTTGGTAATGGAGCAGACAAGTTGCTGTAGGTTGGCCGTGGAGGTGGAGGTGGAGCT	TAAAGACTTTGCTTGATTGACCAAACTGTTGATTCTGTAAGTCCGTAGAGGGCCCCAGCGGCCGCAGCAGCACC	TAAAGACTTTGCTTGATTGACCAAACTGTTGATTCTGTAAGTCCGTAGAGACGGCCAACC	CAGAATCAAAGCAACTTAGAGGA	CTCCGTCCATTGTCACCTTC	TCTCTGACCAGGGAAAGTGG
IAA11	TTGAGCAATTTGAAACCTTTGAAAATGGCAGGCAAGTTGTGGGATGGCCATGGAGGTGGAGGTGGAGCT	TGAAGCCTTAGCTTGGTTAACCATACTGTTCATTCTGTAAGTCCTTATTGGGCCCCAGCGGCCGCAGCAGCACC	TGAAGCCTTAGCTTGGTTAACCATACTGTTCATTCTGTAAGTCCTTATTGATGGCCATCC	CCTGTATAGTGATTCTCTTGGTATCTG	CCTTCCAATAGGAATGCCATC	GAAAAGATTCGACCTTTGTTACG
IAA13	CCTGACTTGATGTTTGCTTTGTTTGTAGCAGTCAAGTTGTTGGATGGCCTTGGAGGTGGAGGTGGAGCT	CTTTGTAGCTTGGTTATTAACCAAACTGTTCATCCTGTGTGACCCTATAGGGCCCCAGCGGCCGCAGCAGCACC	CTTTGTAGCTTGGTTATTAACCAAACTGTTCATCCTGTGTGACCCTATAGAAGGCCATCC	CTCTGAACATAGTTCGTTTAGGAGT	AGCAACTCCATCCATGTTCAC	TGGTTCTAAACGTGCTGCTG
IAA15	TGATCTTCTTAATTTATTGTGGTTTTAGTGACCAGTTGGTGGGCTGGCCGTGGAGGTGGAGGTGGAGCT	CGCCACTTTCACGTATTTCCGCCTCACTGTTTTCCTCGCTGTCGCTACCGGGCCCCAGCGGCCGCAGCAGCACC	CGCCACTTTCACGTATTTCCGCCTCACTGTTTTCCTCGCTGTCGCTACCGACGGCCAGCC	GGGAAGCACACGAGAACAAT	TTCTAGAGCGGTGAAAAGCTG	TTGGAGGAACTGAGCTGACC
IAA16	AAAAAAAAAAATCATTACGAAATTTCAGGGCACAAGTTGTGGGATGGCCATGGAGGTGGAGGTGGAGCT	GGTGGTCGGTTTTTGGCCGGACATGACGTTCTTGCGGAAAGATCGTACCGGGCCCCAGCGGCCGCAGCAGCACC	GGTGGTCGGTTTTTGGCCGGACATGACGTTCTTGCGGAAAGATCGTACCGATGGCCATCC	CAACCACCTAACATCCGACA	CGTCCATGCTAACCTTCACA	TTTGGTTTTGCTCCATGTTTC

Table 8 (Continued)

IAA17	CGTTTTTCCGTTTTAAAAAAAATATAGGGCACAAGTTGTGGGATGGCCATGGAGGTGGAGGTGGAGCT	GCTTGATTTTTGGCAGGAAACCATCACGTTCTTCCGGTATGATCTCACCGGGCCCCAGCGGCCGCAGCAGCACC	GCTTGATTTTTGGCAGGAAACCATCACGTTCTTCCGGTATGATCTCACCGATGGCCATCC	CTCTTCTTAAATCGTCCAATGTC	GCATTAGAAAGCTCATCGTAGC	CATTTTTGTAAAGGTGACTGCAA
IAA27	TTGATGTTTTTTTTTTTTGATTGAGCAGAGCACAAGTGGTGGGATGGCCATGGAGGTGGAGGTGGAGCT	TTTCTGAGATTGAGAAGAAGCCATTGAGTTCTTCCTGAATGATCTTATTGGGCCCCAGCGGCCGCAGCAGCACC	TTTCTGAGATTGAGAAGAAGCCATTGAGTTCTTCCTGAATGATCTTATTGATGGCCATCC	TGCTACAGCTCCAGCTTCAA	AAGGAGCACCTTCCATACTCA	AGGTGCCAAAAGGGTGTTCT
IAA28	CTAGATATGGCTAAAACAGGGTTGAGGTAGCTCCAGTGGTGGGATGGCCGTGGAGGTGGAGGTGGAGCT	CATCTCCTCCTTTAGTTGTGCCGTTAGGTTTCTCCGGGATGATCTCACCGGGCCCCAGCGGCCGCAGCAGCACC	CATCTCCTCCTTTAGTTGTGCCGTTAGGTTTCTCCGGGATGATCTCACCGACGGCCATCC	CTGGTAATTCTACAGTTTTCGGAGA	GGCATGTGAAAGCTGTTGG	AGCTCGACCAAAGAAACATCA
IAA29	TTTTAATATTTAGTGATGAGGAGGAAAATAGCGAAGTTGTTGGATGGCCATGGAGGTGGAGGTGGAGCT	AATATGACGATGATGATAACTACCATACTTTATCATACATGTTTTTACTGGGCCCCAGCGGCCGCAGCAGCACC	AATATGACGATGATGATAACTACCATACTTTATCATACATGTTTTTACTGATGGCCATCC	CAATAGTTTACGAGTGAGTACCTTGAC	GATCCCACAGTAGCCGTTGT	TTTGGGTTAGGGAATGTGGA
IAA31	TGTGTTTTAAATCTTTATTTTTTTGTAGAGAAGCAAGACAAGATTGGCCATGGAGGTGGAGGTGGAGCT	ACGAAGAAGACGACGACCCTTTAGTGTATCTCTTAATCTAGACTTTATCGGGCCCCAGCGGCCGCAGCAGCACC	ACGAAGAAGACGACGACCCTTTAGTGTATCTCTTAATCTAGACTTTATCGATGGCCAATC	CACATTTCCTTCTACTTCTCCACA	AATAGACTCTCGTAGCCTGAGAAT	GCTCTGCAGCTTCAAGGTCT
IAA32*	TCAATATCTATATGTGAATCAAAGTGATAGGTCTCGACGGGTATGGAGAGGGAGGTGGAGGTGGAGCT	TGGTGTCCAGTGTCCTCACTCTTCAACTGTGTAATGCTGTTATAGCTGGGGGCCCCAGCGGCCGCAGCAGCACC	TGGTGTCCAGTGTCCTCACTCTTCAACTGTGTAATGCTGTTATAGCTGGGCTCTCCATACCCGTCGAG	AATAGCCGTGCTTTCACCTC	GCCATCCAGATTTACCTTTACG	AGTTGAGATGTGACTTTGTGC
IAA5	TTGATTTGAAATGTGAACCGGCGAAAAAGAGTCAAGTTGTGGGTTGGCCATGGAGGTGGAGGTGGAGCT	CGAACTTTTGGTCCGTTCGAGACTGTTCTTTCTCCGGTACGAACAAACCGGGCCCCAGCGGCCGCAGCAGCACC	CGAACTTTTGGTCCGTTCGAGACTGTTCTTTCTCCGGTACGAACAAACCGATGGCCAACC	TAATCTTGGACTCGAAATCACC	AATCCTTAACCTCTTGCACG	-
IAA6	TTGAAGATGAATCACTGCCGGTTGTGAAGAGTCAAGCGGTGGGATGGCCATGGAGGTGGAGGTGGAGCT	AGCTTTCGATGCTTCCTCATTGTTCTTCTTTCTCCTGTAAGAACACACAGGGCCCCAGCGGCCGCAGCAGCACC	AGCTTTCGATGCTTCCTCATTGTTCTTCTTTCTCCTGTAAGAACACACAGATGGCCATCC	TAGCGAAATATCAGTATGCG	CTAACCTATGCCAAGACAGC	TTCCAACAAGAGAAGTGCTTGA
IAA4	CTAATCTATGCTATTTCTTGTTTTATAGGGCTCAGATTGTTGGATGGCCATGGAGGTGGAGGTGGAGCT	AGATTCACTCTTCTTTGTCTGAACATTGTTCTTTCTGTAAGATCTAACTGGGCCCCAGCGGCCGCAGCAGCACC	AGATTCACTCTTCTTTGTCTGAACATTGTTCTTTCTGTAAGATCTAACTGATGGCCATCC	GAGATTGGGATTACCAGGGA	TCCTTTCATGATCCTTAGCC	-

Table 8 (Continued)

IAA19	AGAAGGTTAATGATTCGCCGGC GGCGAAAAGCCAGGTGGTGGG GTGGCCATGGAGGTGGAGGTGG AGCT	CGTGGTCGAAGCTTCCTTACAG CTGTTTTTCTTCCGGTAAGAACA AACCGGGCCCCAGCGGCCGCAG CAGCACC	CGTGGTCGAAGCTTCCTTACA GCTGTTTTTCTTCCGGTAAGA ACAAACCGATGGCCACCC	CAAGAAATG GAGAAGGAA GG	TAACCGATG CCACGGAAA CC	-
IAA26	TTTGATATCTGGTTCAAAGAACT GCTCCTGGTCCAGTGGTGGGTT GGCCTTGGAGGTGGAGGTGGAG CT	TAGCTTTGAAGAGCTTGTGCTC GCTAGATTCTTTCTGAACGAAC GAACCGGGCCCCAGCGGCCGCA GCAGCACC	TAGCTTTGAAGAGCTTGTGCT CGCTAGATTCTTTCTGAACGA ACGAACCGAAGGCCAACC	ATTAGGGAC CCATCTTCT GC	CCTTTCCCA TCTAATAAT CC	-

*: *IAA32M2* mutant was generated by deleting 6 amino acids of domain II by recombinnering. In the first recombination, *galK* cassette amplified with MGAL-F and MGAL-R was used to replace coding region of 6 amino acids targeted for deletion. 68-mer forward prime MGAL-F includes 50 bp of homology to sequence upstream the deletion site and 18 bp homologous to link G. downstream The 74-mer reverse prime MGAL-R includes 50 bp of homology to sequence downstream the deletion site and 24 bp homologous to link A. In the second recombination, the *galK* cassette was deleted with the PCR product amplified using MUTTEST-F/MUT-R primer pairs and *Arabidopsis* Columbia genomic DNA as template. 68-mer long oligo MUT-R is composed by 50 bp homologous to sequence downstream the deletion site and 18 bp homologous the sequence upstream the deletion site.

Table 9. Primers used for replacing part of domain IV with Sypet or Typet.

Name	R4 -F	R4-R	R4T-F	R4-R
IAA1	TCAAGTTCACAGTAGGTGAATATTCCGAGAGAGAA GGCTACAAAGGATCTGGAGGTGGAGGTGGAGCT	GCTTCGGATCCTTTCATGATTCTGAGTTTTTGACAAGATG AAGAGAACATGGCCCCAGCGGCCGCAGCAGCACC	TGGACGGAGCTC CATATCTC	TTGCTTTTTCTTT CTTTACACCA
IAA2	TCAAAGTCATGATTGGTGAATATTGTGAGAGAGAAG GATACAAAGGATCTGGAGGTGGAGGTGGAGCT	GCGTCGGATCCCTTCATGATTCTGAGTCTCTTACAAGAA GAAGAGAACATGGCCCCAGCGGCCGCAGCAGCACC	CAACAACAGTGT GAGCTACGTG	TCCGAATCAAAC CAAAAAGG
IAA3	TGTTCAAATTCTCTGTGGGAGAGTACTTTGAGAGAG ATGGATATAAAGGTGGAGGTGGAGGTGGAGCT	TGATTTTGTGGAAATTGTTTTCAGACAGGTATATGAAAG AAGACTTACTCGGCCCCAGCGGCCGCAGCAGCACC	GCATGAGGGTCA AGGAATCT	CCTTTGGCTTCTG ATCCTTTC
IAA7	GAATGATAGATTTCATGAACGAGAGCAAGCTAATG AATCTGCTGAATAGCGGAGGTGGAGGTGGAGCT	GTTATAAGTTATAAAAAAAAGTATATGAAACGATTAAAT TTGACTCACTCGGCCCCAGCGGCCGCAGCAGCACC	CAAAACTCATGA ACCTTACCG	CCTCTTGTCTCTT GCCAACC
IAA8	AAGGGAGGGAAAGAATGAGCGAGATTAAACTGAAG GATCTTCTTCATGGAGGAGGTGGAGGTGGAGCT	GAAACAAAACCACAGTTATTAAAGCTGAGAGTTCCAAA AGAAGCTTACTCGGCCCCAGCGGCCGCAGCAGCACC	GCTTCACCCTTGG TGAGAAT	CAGATTAGGGGG AAAGATGTTGG
IAA9	CTGGGAAGGATATGCTTAGTGAGACCAAGCTCAAG GATCTTTTGAATGGAGGAGGTGGAGGTGGAGCT	AGTAGAGAAAGCTTCTAACAAGTTTTGAATATAAGAGA GGAAACTTACTCGGCCCCAGCGGCCGCAGCAGCACC	GAGCTTTCTTCAG CCTTGGA	CTTCAGCTTCTTG CAGACATC
IAA10	GATACAAGACAATTAAGGAAACATGTACTTCGAAA TTACTGGATGGCTCAGGAGGTGGAGGTGGAGCT	ATTACAAATACAGATGATGAATGGGAAGTTATAAGGGT AGTCACTTACTGGGCCCCAGCGGCCGCAGCAGCACC	GATTCCTTCTCCT GTAACAAGTCA	CCAGCTCCAATTG ATGTCTTC
IAA11	AAACCGATGGTCACATGGAAACACCGGTCAAGATA CTACCAGATGGGTCTGGAGGTGGAGGTGGAGCT	CTTACACGGGACAAAGCATCAAAAGCATAAGAACATAT GTATGCATACCCGGCCCCAGCGGCCGCAGCAGCACC	GTGCTTTTCGAAT CTCATGG	CCAGTAGCCTCC GATGTTTT
IAA12	GTACTACTTGTCGAGAAAAGGTTAAACCTTTAAGGC TTTTAGATGGATCAGGAGGTGGAGGTGGAGCT	TAATCAGGTTTTGAACTACTGCAATACTGCTTAATATCC GAATCTTACCTGGCCCCAGCGGCCGCAGCAGCACC	GAAATGTTCTTTG GAATGACAGG	CGAGAAATCAAG AAACCAGAAGA
IAA13	TCGGGTTAACCAGTCAGTTCACTAAACCGTTGAGGC TTTTAGATGGATCGGGAGGTGGAGGTGGAGCT	TTATTGGAAGTGTAAGAGATAAATAAACAAGGTTTTTAG ATCTATTACCTGGCCCCAGCGGCCGCAGCAGCACC	GTGAACATGGAT GGAGTTGC	ACCGAGTCCATT AGCTTCAG
IAA14	GGATGATAGATTTCATGAACGAGAGTAAAGTGATG GATCTGTTGAACAGTGGAGGTGGAGGTGGAGCT	AATAATTTTGGTCATAATAATGAAAGCAAGAAGAATGA GGAAACTCACGGGGCCCCAGCGGCCGCAGCAGCACC	CCATGGGTATGC ATTTTCAG	AAGATGGCTGTG CGTGTTTA
IAA15	ACTTGGGGTTTTGGATGGTGGTAGGTAGAGTGACAG AGTTGGAGAGGAAGGGAGGTGGAGGTGGAGCT	AAGGAAAAGTACGTAAAAGCAACACAAAGGGTAATTCC GAGAACTCACATGGCCCCAGCGGCCGCAGCAGCACC	GACAGCTTTTCAC CGCTCTA	TTCCTTCACAGCA GAATAATACAG
IAA16	GAATGAAAGATTTCATGAATGAGAGTAAATTGATCG ATCTTCTAAACGGAGGAGGTGGAGGTGGAGCT	TTTATAATAATAATTCATATATAAATTTAAAGATATAAA AATACTCACTCGGCCCCAGCGGCCGCAGCAGCACC	CTCCAACGCCTTA AGCAAAA	CGATTGCTTCTGA TCCCTTC
IAA17	GAATGATAGACTTCATGAATGAGAGGAAATTGATG GATTTGGTGAATAGCGGAGGTGGAGGTGGAGCT	AAAGCTCTAATTAAATGGTAATTAATAAGGGAGAAGAA GAAGACGTACGGGGCCCCAGCGGCCGCAGCAGCACC	GCACCATAAAAG AAAACCACCA	AAACGCTTGCAT GTATCGAC
IAA27	GTGGCAGAGATGGGTTAAACGAGAGTCGCTTGACT GATCTCTTGCGTGGTGGAGGTGGAGGTGGAGCT	TGTTAATTTACATTGTTGAGATTGATCTTTTAAAAATCAA AGACTTACTCGGCCCCAGCGGCCGCAGCAGCACC	CGTTTTCTAGTCT CTCCTTCTTTGA	CAGCTTCTTGCAG GAGCATA
IAA28	CACATGCCGTTGACCAACTCTTCTCTAAGAAAGATT CGTGGGATCTAAACGGAGGTGGAGGTGGAGCT	AGTTCCAGAAAAAAAAATGGAATTAGTTTTGTTTGATTT AAAACTTACTCGGCCCCAGCGGCCGCAGCAGCACC	AAGGAGAGTGAT GAAGAGAAGGAA	TCTGATGAGAAA TGGCCAAA
IAA29	CTATTGAATCTAATTTTGGGTTATCAGATGAAGATT GCGACAGAGAAGATGGAGGTGGAGGTGGAGCT	CATTACTCAATAAGTCATTTTTCTTGAAATATAGAGAAG TAAACTTACTTGGCCCCAGCGGCCGCAGCAGCACC	CGAGTCCCTCACT AACTCCTTG	TGCACACGGTCG ATCTCTAA

Table 9 (Continued)

IAA31	AAGTATATCCATGCATTGTCTTGGCTATAACAGGCG GTAATCGAGATCGAGGAGGTGGAGGTGGAGCT	TAACCGAGGAAATTATACCAATGTAAGATCAAGATATA AGATTCATACTCGGCCCCAGCGGCCGCAGCAGCACC	GCGTATTCTCAGG CTACGAGA	CCTTGACCGATCC AAGTTTT
IAA32	AATCTGCAGGGATGCAGACCGTGTCGGGATTGAGGT TGTTCCAGACTGAGGGAGGTGGAGGTGGAGCT	TTAACTTAAAATGCTTAGTTTTTGTATAGAGAAAATGTA AGAGCTTACTTGGCCCCAGCGGCCGCAGCAGCACC	CGCAACTCTTGCT CTTCAGC	CCCCAGAAGAAG CTTGGAC
IAA4	TGTTTAAATTCTCTGTGGGAGAATATTTTGAGAGAG AAGGATATAAAGGCGGAGGTGGAGGTGGAGCT	GTGTCATAAGTTAACTAATTTGAGAGATAAATGAAAATT TGTACTTACTCGGCCCCAGCGGCCGCAGCAGCACC	GAGATTGGGATT ACCAGGGA	TCCTTTCATGATC CTTAGCC
IAA5	TTAACTTTATAATCTTTATATATGAAAATGAAGATG ATACGTTGAAGGAAGGAGGTGGAGGTGGAGCT	AATAAAAGAAAGTAAATTTGAATGTTATAGAGGTCGCT AGAAACTTACTCGGCCCCAGCGGCCGCAGCAGCACC	TAATCTTGGACTC GAAATCACC	AATCCTTAACCTC TTGCACG
IAA6	TAACACATGTATATATAAACTATGCAGGAGTGGCGA AGGAGGGTAAGAAGGGAGGTGGAGGTGGAGCT	GTAAATTTGATAATTGAGCATTGTAATGCAAGCTAAACA TAAACTTACTGGGCCCCAGCGGCCGCAGCAGCACC	TCTTGCTGGAGAC CAAAACC	TCTTGCTGGAGAC CAAAACC
IAA19	TTACATATATACATATTTGTGCAGGTGTGGCCTTGA AAGATGGTGACAACGGAGGTGGAGGTGGAGCT	CATTTTCTTTTTGTAGTGAGATATCTTAAGTCAAAAGATA GAACATACCCGGCCCCAGCGGCCGCAGCAGCACC	TCTACAACCAAA CCAACACCC	AACCAGCTCCTTG CTTCTTG
IAA26	CTGATGGTCAAGGAGAAGAGAAACCTATCATTGGA TTATTAGATGGGAAAGGAGGTGGAGGTGGAGCT	CAAAACACTGAATCAGATTCAGAGAACAAAAGGGGAAA AGATTCTTACTGGGCCCCAGCGGCCGCAGCAGCACC	CAAAGAACTGCT CCTGGTCC	TCACACGCAGTCT CTTCACAG

Table 10. Primers used for RT-PCR of IAA-GFP constructs.

GENE	**RT-F**	**RT-R**
IAA4	GGTGATGTTCCTTGGGAGATG	ACCCGAGCAGTTTCAGAGC
IAA5	GAGATGTTCCTTGGGAAATG	TGTAAGGCTCACTCACATTCAC
IAA19	GTGATGTACCTTGGGGGATG	TTAAATTAATGAACCAGCTCCTTG

CHAPTER 3

Optimization of Recombineering-based Tagging System for a systematic *Arabidopsis* gene functional analysis

SUMMARY

In order to generate a genome-wide collection of TAC-based gene translation fusions in *Arabidopsis*, a high-throughput recombineering protocol needs to be implemented. To facilitate the systematic production of the whole-gene GFP fusions, we optimized several of the rate limiting steps of the previously developed gene-by-gene recombineering procedure. Specifically, we evaluated the possibility of 1) preparing competent cells in pools of individual *JAtY* clones, 2) introducing new version of fluorescence proteins to improve the sensitivity and, therefore, widen the utility of the system, and 3) constructing the *Sypet-FRT-Amp-FRT* cassette to improve recombination efficiency and accelerate the procedure. Finally we have also developed a program (*Arabidopsis* Tagging (v.1.0)) to automatically select the most suitable TAC clone for any given *Arabidopsis* gene, and generate the sets of PCR primers required for tagging and verification. Based on the preliminary data generated, a 96-well-format recombineering pipeline for *Arabidopsis* is proposed. We believe these improvements have results in a highly efficient and robust 96-well recombineering protocol. The development of this procedure is essential step towards the generation of whole-genome collection of FP-tagged constructs, and therefore an important contribution in the field of *Arabidopsis* functional genomics.

RESULTS AND DISSCUSSION

Recombination using competent cells generated with pooled cells

Preparing high-quality competent cells is critical for achieving high rates of recombination. However, current procedures require that the competent cells for each *JAtY* clone (one per gene) to be prepared and electroporated individually. Of course, competent cells can be prepared by growing the *SW102-JAtY* clones in 96 deep well plates and by carrying out all subsequent centrifugation and washing steps in such type of plates. However, in our hands, competent cells prepared in this way display relatively low transformation efficiencies (data not shown). An alternative approach to achieve high throughput would be to take advantage of the high specificity of homologous recombination, and prepare the competent cells as pools of individual *JAtY* clones. As the first step to evaluate this possibility the efficiency of the second recombination with deoxy-galactose counter-selection using competent cells of a pool of 2 *JAtY* clones was carried out (Figure 42). High efficiencies were achieved in all reactions (87.5%-100%) indicating the feasibility to prepare competent cells by pooling several *JAtY* clones together.

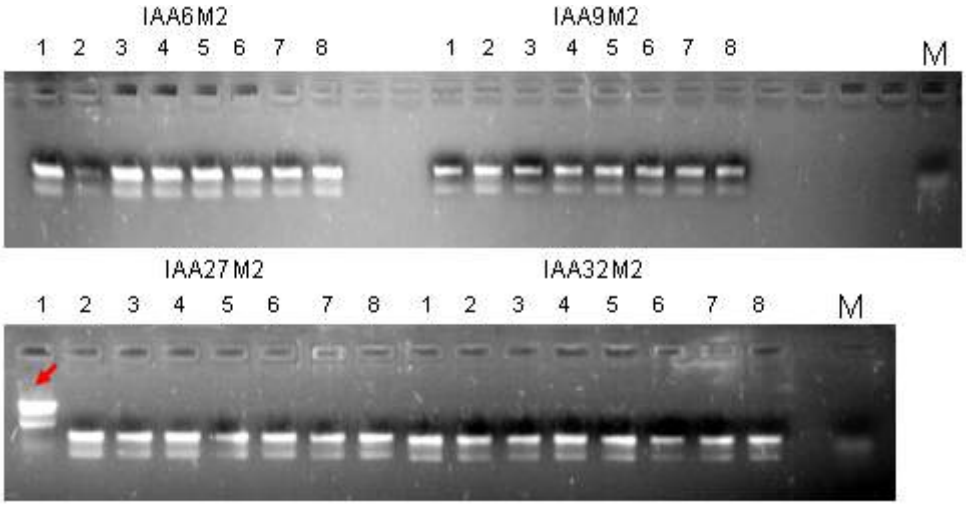


Figure 42. Efficiency of recombination when using a pool of two *JAtY* clones. Pool 1: IAA6 and IAA9; Pool 2: IAA27 and IAA32. 8 independent colonies for each construct were tested by PCR using the MUTTEST-F/ MUTTEST-R primers (Table 8) after recombination (replacement of *galK* by DNA fragment amplified using MUTTEST-F and MUT-R primers, Table 8) and the PCR products were running on a 1% agarose gel. M: DNA maker. The presence of a band slightly larger than that of the marker indicates the recombination was successful. Arrow shows the only one case (IAA27M2 lane 1) the recombination failed as indicated by the large size of the PCR product.

Tagging *Arabidopsis* genes with *Sypet-FRT-Amp-FRT* cassette

galK selection has been optimized for use in our recombineering-based gene tagging system. However, the long selection period (4-5 days for first and second recombination) and variable efficiencies associated with the second recombination, limit its use in a high-throughput strategy. As an alternatively to the inefficient second recombination step (replacement of *galK* by the desired tag), a selectable marker flanked with locus of *lox*P or Flippase Recognition Target (*FRT*) sites and the nonselectable tag (such as GFP) can be inserted together as part of a single cassette. Removal of the selectable marker can then be achieved by the expression of the Cre or Flippase (Flp) proteins. Arabinose inducible *flp* gene has been introduced into BAC host cells to facilitate BAC modification using (*FRT*) sites (Lee et. al., 2001). Similarly, a recombineering strain (SW105) carrying an

$_L$-arabinose-inducible *Flp* gene has been constructed to facilitate the modification of BACs clones (Warming et. al., 2005). In fact, several groups working with C. elegans have implemented this strategy in their recombineering procedures (Sarov et. al., 2006; Tursun et. al., 2009). Their results motivated us to generate the *Sypet-FRT-Amp-FRT* cassette (Figure 43) in order to bypass the low efficience of the second recombination step. This cassette has three important novelties. 1) It incorporates the A and G adaptors making it compatible with our *galK*/*GFP* system. 2) It utilizes the codon-optimized *Ypet* making it ideal for expression studies in *Arabidopsis*, and 3) It utilizes an ampicillin resistant gene making the selection process very rapid and robust.

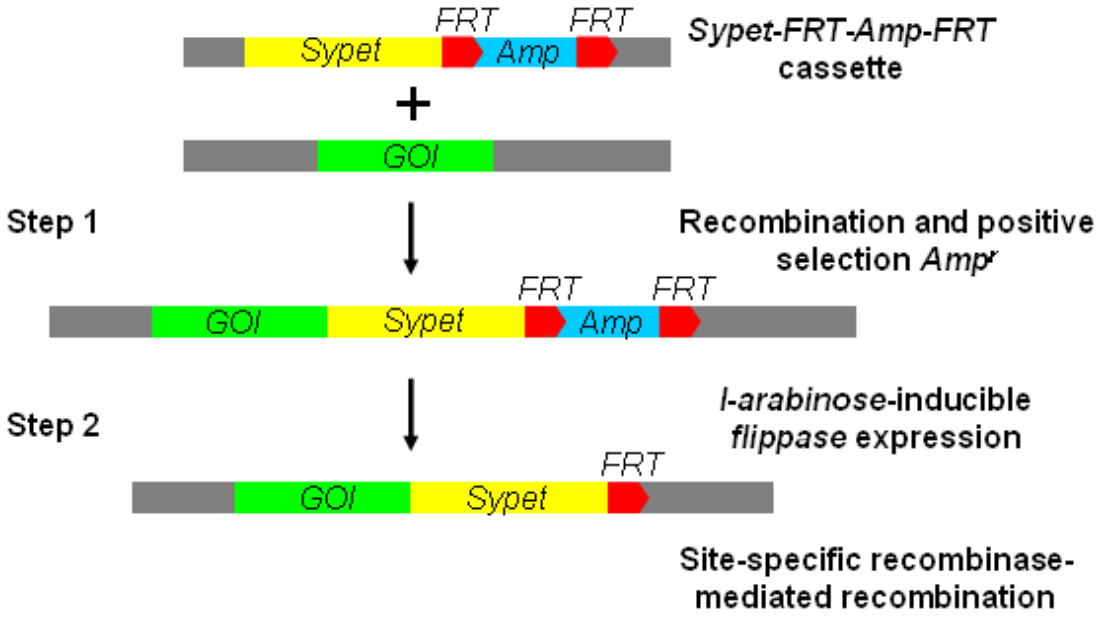


Figure 43. Schematic representation of recombineering-based *Sypet-FRT-Amp-FRT* tagging system. In Step 1, the Sypet fragment is introduced along with the selectable marker Amp which is flanked by *FRT* sites. Recombinants are selected for the presence of *Amp*^{r}. In Step 2, *Amp* is removed by the $_L$-*arabinose*-inducible site-specific flippase-mediated recombination. Unlike the 'seamless' method, a single *FRT* site (11 amino acids) is left behind after recombination.

The utility of this system was tested when building IAA7, 10, 11, 17, 28-M2Sypet constructs. In the first step to insert *Sypet-FRT-Amp-FRT* cassette, only 2 out of 40 recombination reactions failed (Figure 44A), indicating the high efficiency of using *Amp*

as selectable marker. The second step, site-specific *flippase*-mediated recombination has even higher efficiency with *Amp* successfully removed in all reactions (Figure 44B). Although a single *FRT* site is retained after recombination unlike the 'seamless' method, the shorter selection period (2 days for each step) and high efficiency of each step enable the high-throughput application of this approach.

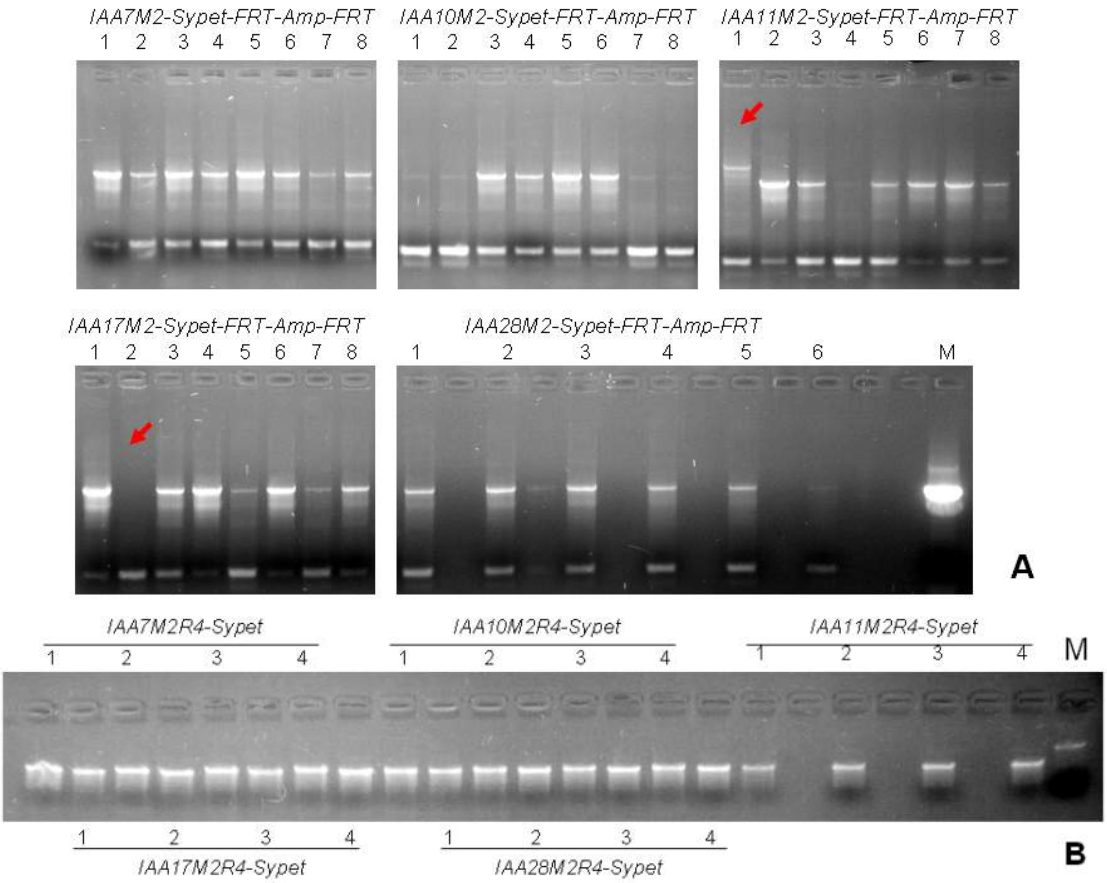


Figure 44. Efficiency of recombineering-based gene tagging with *Sypet-FRT-Amp-FRT* cassette. For each step, competent cells were made by pools of individual *JAtY* clones (Pool 1: IAA7, IAA10 and IAA11; Pool 2: IAA17 and IAA28). The recombinants were tested by PCR and PCR products were examined by 1% agarose gel. M: DNA maker showing expected size of *Sypet-FRT-Amp-FRT* insertion. (A) Efficiency of step 1 (λ-Red mediated recombination) to insert *Sypet-FRT-Amp-FRT* cassette. For each construct, 6 or 8 independent colonies were selected after recombination and tested by PCR. Arrows show the two cases (*IAA11M2-Sypet-FRT-Amp-FRT* lane 1 and *IAA17M2-Sypet-FRT-Amp-FRT* lane 2) the recombination failed as indicated by the large size of the PCR product or missing *Sypet-FRT-Amp-FRT* insertion. (B) Efficiency of step 2 (site-specific recombinase-mediated excision of the selectable marker). For each construct, four independent colonies were tested after recombination.

96-well-format recombineering pipeline for *Arabidopsis*

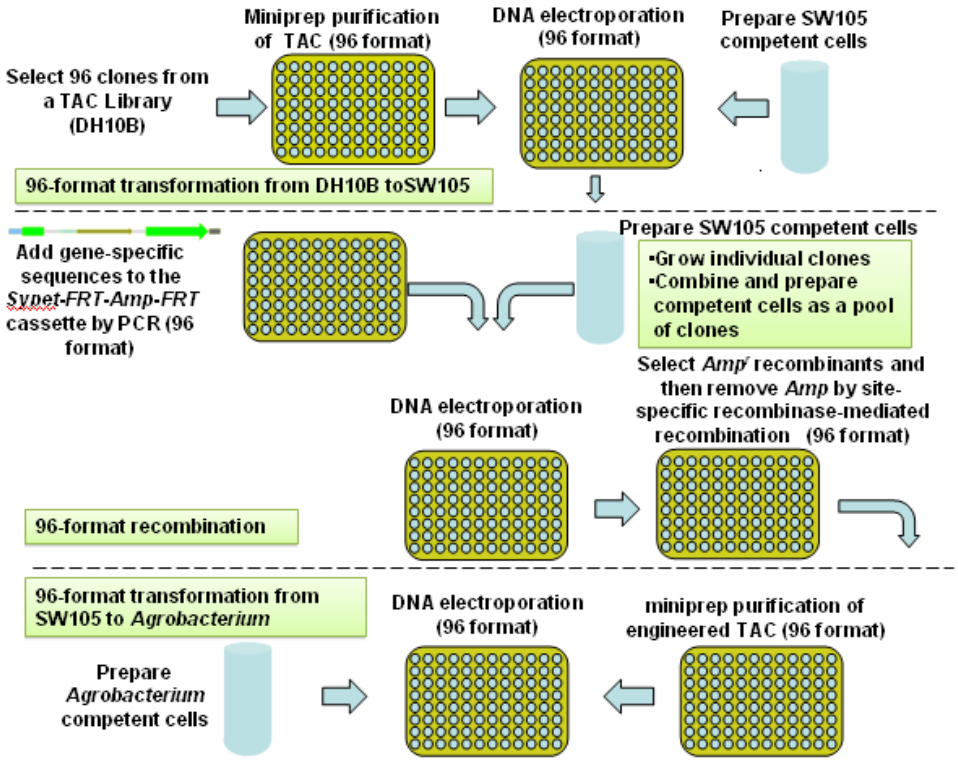


Figure 45. Schematic diagram of 96-well-format recombineering pipeline for *Arabidopsis*. 96 *JAtY* clones containing 96 GOI are selected and grown in a 96-well plate. 96-well-format miniprep is prepared to purify TACs and 96-well-format electroporation is performed to transfer the TACs into *SW105* competent cells. The *Sypet-FRT-Amp-FRT* cassettes amplified in 96-well plate are targeted to the TACs in recombineering reactions performed in 96-well plates with competent cells made by pool of individual clones. 96-well-format transformation is conducted to transform modified TACs from *SW105* into *Agrobacterium* cells to be ready for transforming plants.

Based on the gene-by-gene recombineering-based tagging approach already extensively tested in this study, as well as, the improvements described above (preparation of competent cells in pools and substitution of the classical *galK*-based two step recombineering by the sigle step *Sypet-FRT-Amp-FRT based system*), we propose a 96-well-format recombineering pipeline that should be suitable for the generation of a genome-wide collection of whole-gene translation fusions in *Arabidopsis* (Figure 45). This procedure has been successfully employed in the Alonso-Stepanova lab to tag 3×96 genes (Perez-Amador et. al. unpublished).

Web tool developed for recombineering-based gene tagging system

Identifying the best available *JAtY* clone for a gene of interest, and designing the primers needed for recombineering are important and time consuming steps that could be automated. To make this potentially rate limiting step fast and simple and automated, a web tool application, *Arabidopsis* Tagging (v.1.0) (http://www.arabidopsislocalizome.org/), was developed. In the following sections the use of this application is described.

Input

Arabidopsis Tagging (v.1.0) accepts three types of inputs. For a single gene, in the textbox below the Input Option 1, users can either type in a locus identifier (*At1g01010*, etc.) by choosing type GENE, or input a gene name or any key word that described the function of the gene (*IAA1*, etc.) by choosing type FUNCTION. The tagging position (N-terminal or C-terminal) can also be selected. By default the program will return the results for tagging at the C-terminus. If the input type is specified by Function, the program will display the results for all of the *Arabidopsis* genes whose description contains the keyword. For example, any of the following keywords, F19P19.32, F19P19_32, *IAA3*, INDOLE-3-ACETIC ACID 3, SHORT HYPOCOTYL 2 or *SHY2*, will generate the same output as entering the corresponding gene identifier *At1g04240*, whereas entering IAA will result in an output containing all of the *Arabidopsis Aux/IAA* genes.

For multiple genes, users can paste locus identifiers (At1g01010, etc.) with the desired tagging position (N or C; by default the program will return the results for tagging at the C-terminus) separated by tabs or spaces in the textbox below the Input Option 2.

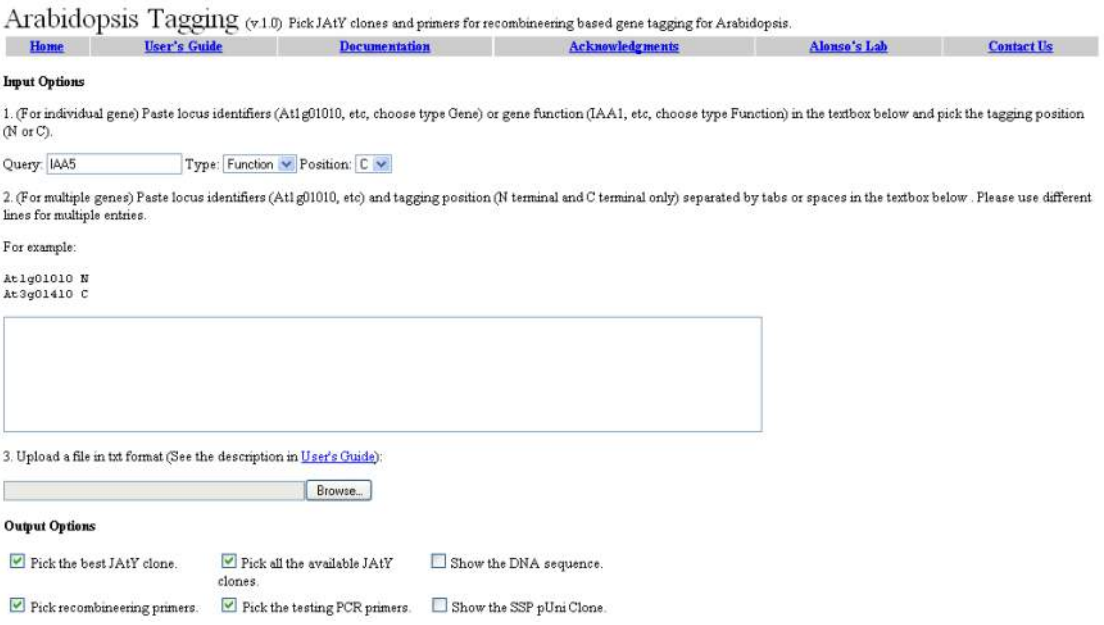
In addition, a single file with a gene list (one gene per line, including locus identifiers and desired tagging position) can be uploaded. The program will return the results for every gene entry in the file. The output can be viewed on the screen or downloaded as a file through the link provided in the output. The program only accepts txt format input files with size smaller than 50 kb.

Output

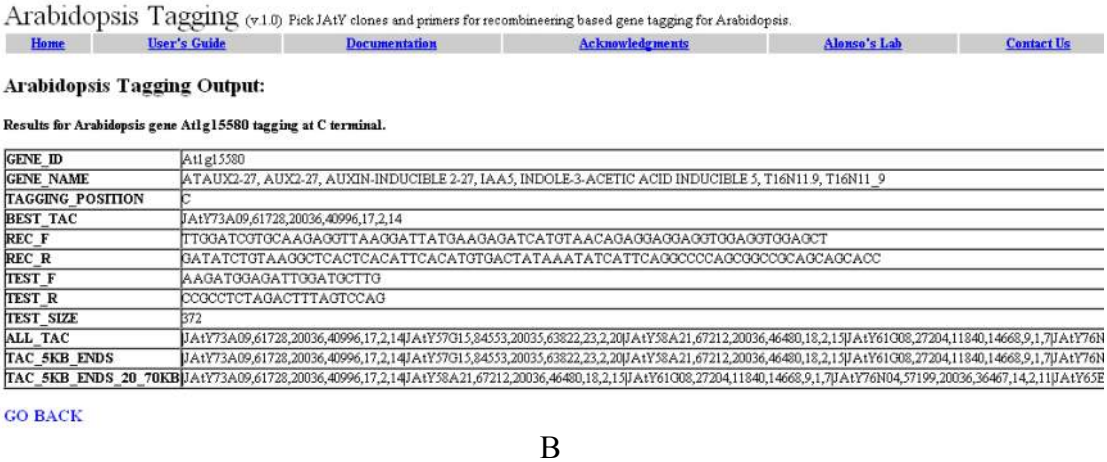
By default, the program will return output listed below： Gen_ID: locus identifier; GENE_NAME: gene symbol; TAGGING_POSITION: N for N-terminal and C for C-terminal; BEST_TAC: information on the best available *JAtY* clone (based on the criteria described below) including clone ID, size of insert, distance in Kb from the 5' end of the TAC clone to the gene of interest, distance from the 3'end of the TAC clone to the gene of interest, total number of genes in the clone, number of genes upstream of the gene of interest, and number of genes downstream of the selected gene; REC_F: forward primer for homologous recombination; REC_R: reverse primer for homologous recombination; TEST_F: forward primer for PCR testing of recombineering-modified clones; TEST_R: reverse primer for PCR testing of recombineering-modified clones; TEST_SIZE: expected size of the PCR product amplified with the TEST primers using *Arabidopsis* genomic DNA as a template. Users can also retrieve additional information by marking the check boxes below the Output Options. The additional information contains: ALL_TAC: information for all of the *JAtY* clones containing the gene of interest; TAC_5KB_ENDS: information for all of the *JAtY* clones where the gene of interest is located at least 5 Kb from either end of the TAC; TAC_5KB_ENDS_20_70KB: information for all of the *JAtY* clones with insertion size between 20 to 70 Kb where the gene of interest is located at least 5 Kb from both ends of the TAC; DNA_SEQ: DNA

sequence of the gene of interest including 500 bp upstream of the start codon and 500 bp downstream of the stop codon, with exon sequences indicated by upper case letters; UNICLONE: SSP pUni Clone ID (Yamada et. al., 2003).

As an example, the input and output of *Arabidopsis* Tagging (v.1.0) for the *Arabidopsis* gene *IAA5* are shown in Figure 46.



A



B

Figure 46. Interface of *Arabidopsis* Tagging (v.1.0) and an input and output example. (A) Gene *IAA5* was used to demonstrate how to input queries using this web tool. (B) Output generated for *IAA5* corresponding to the input options specified in A.

Data Processing

All of the DNA sequences were downloaded using T-DNA Express: *Arabidopsis* iSect Tool (http://signal.salk.edu/cgi-bin/tdnaexpress), including 500 bp upstream of the start codon and 500 bp downstream of the stop codon. If a gene had more than one splicing form, sequences supported by an SSP pUni Clone (Yamada et. al., 2003) were chosen. If none of the transcripts are verified by the SSP pUni Clone, the first DNA sequence in the *Arabidopsis* iSect Tool output was used.

For N-terminal tagging, start codon and 47 bp upstream of the start codon plus the linker G sequence (5'-GGAGGTGGAGGTGGAGCT-3') are used as a forward primer for recombination (68 bp). Reverse complement of the 50 bp downstream of the start codon followed by the linker A sequence (5'-GGCCCCAGCGGCCGCAGCAGCACC-3') are used as a reverse primer for recombination (74 bp).

For C-terminal tagging, a reverse complement of the stop codon and 47 bp downstream of the stop codon plus the linker G sequence are used as a reverse primer for recombination (74 bp). 50 bp upstream of the stop codon plus the linker A sequence are used as a forward primer for recombination (68 bp).

The PCR testing primers were generated using T-DNA Express: *Arabidopsis* iSect Tool (PrimerS) (http://signal.salk.edu/cgi-bin/tdnaexpress) using the following options: PRIMER_OPT_SIZE=21; PRIMER_MIN_SIZE=18; PRIMER_MAX_SIZE=28; PRIMER_OPT_TM=61; PRIMER_MIN_TM=53; PRIMER_MAX_TM=71; PRIMER_MIN_GC=20.0; PRIMER_MAX_GC=80.0; PRIMER_GC_CLAMP=1; EXT5=250; EXT3=250 and FORMAT_OUTPUT=1. For N-terminal tagging, a 500 bp DNA sequence containing 247 bp upstream of the start codon, the start codon and 250 bp downstream of the start codon are used as an input, with the middle 200 bp replaced by

200 ‘N’ to make sure the primers generated flank the insertion site. For C-terminal tagging, a 500 bp DNA sequence containing 250 bp upstream of the stop codon, the stop codon and 247 bp downstream of the stop codon are used as an input, with the middle 200 bp replaced by 200 ‘N’ to make sure the primers generated flank the insertion site. The first pair of primers returned is chosen as output.

To identify the best *JAtY* clone containing the gene of interest, all of the clones containing the gene of interest were first selected (Category I). Next, the list was trimmed by removing those clones that have fewer than 5 kb from either end of the gene of interest to either end of the TAC clone (Category II). The next cutoff (Category III) was made by selecting the TAC clones with insertion size in the range from 20 to 70 kb. Among those, the best *JAtY* clone was chosen based on having the biggest minimum distance from either end of the gene of interest to either end of the clone. If all of the clones in Category II are smaller than 20 kb, the largest available clone was chosen as the best clone. Similarly, if all of the clones in Category II are larger than 70 kb, the smallest available clone was chosen as the best clone. If none of the clones pass the Category II requirements, direction of the gene is then considered and the best clone was identified as that having the greatest distance from the start codon to the upstream end of the *JAtY* clone.

Programming and database design

Perl script was written to download the DNA sequences and description of genes to local computer, design oligos for recombination, generate the input sequence for online primer designing tool and retrieve testing the primers from output (arabidopsistagging.pl). The same perl script was also used to choose best available *JAtY* clone based on criteria mentioned above. Data generated by per script was then stored in a database with Locus

Identifer as key. A CGI script (arabidopsistagging.cgi) was written to display the interface of web tool, acquire the input information, query the database and generate the output for users to review. All these scripts and files required to run them can be downloaded from (http://www.arabidopsislocalizome.org/arabidopsistagging.zip).

CONCLUSION

Ever since the genomic sequencing of *Arabidopsis* was completed more than 10 years ago, understanding the function of all the genes in this genome has become the next major challenge in plant biology. Among the many layers that comprise a gene's function, the detailed spatial and temporal expression map is among the most challenging to obtain. Nevertheless, and due to the critical importance of this type of gene expression information, tremendous efforts have been dedicated to this task in several model organisms. One of the main problems of generating this detailed expression maps is the need of capturing all regulatory elements of a gene. Utilization of recombineering technology has proven invaluable in achieving the goal of generating the gene modification, so their activity can be easily monitored, preserving their genomic context. This methodology is applicable to large and rare transcripts, which are difficult to obtain as cDNAs, keeps complete repertoire of potential splice variants encoded by a gene, and most importantly, enables the expression of the transgene from its native genomic environment including the presence of native transcription and translation regulatory sequences.

Our results indicate recombineering-based gene tagging using the λ-Red system and the *JAtY* library of *Arabidopsis* TAC clones, is a good combination to generate the whole-gene GFP fusions required for detailed spatial-temporal gene expression studies. We developed a recombineering-based gene tagging protocol and examined the efficiency and fidelity of each step in the recombineering procedure, as well as, some of the factors that may affects such parameters. Our results demonstrate that tags can be precisely inserted to a desired location of GOI with high efficiency. We discovered that good transformation efficiency can be achieved when transforming plants with *JAtY* clones by

simply replacing sucrose with glucose in infiltration media. In addition, ~70% transgenic lines contain the tagged GOI and ~80% of transgenic lines have single-copy insertion. Fusion proteins generated using recombineering are expressed with the expected pattern, and located in correct subcellular localization. Our results also indicate that the recombineering-based system can be adopted as a general strategy, not only for the generation of whole-gene reporter fusion constructs, but also for making all sorts of sequence modification, including point mutations, deletions and replacements, in the pseudogenomic context of a large TAC clone.

We exemplify the utility of this approach by tagging the low-expressed and short-lived *Arabidopsis* Aux/IAA proteins, and by studing the expression pattern and subcellular localization of this gene family. Tagged Aux/IAA proteins were nuclear-localized and expressed with specific pattern in *Arabidopsis* seedlings. Information of Aux/IAA localization facilitated by the constructs described in this work will greatly help to understand auxin-mediated responses in plants.

We have also made significant advances towards the optimization of the procedure and the development of a high-throughput protocol. Specifically, our preliminary data indicate that competent cells could be made by pooling individual *JAtY* clones. The use of the new recombineering *Sypet-FRT-Amp-FRT* cassette will significantly improve the efficiency and speed of the process. Thus, we proposed a high-throughput pipeline that enables recombineering to be carried out in a 96-well format, an indispensable step towards the generation of a large collection of whole-gene translation fusions. We believe that the rapid innovation in automated tissue culture, and robotic confocal imaging, together with the development of new genetic tools such as the one described in this work will make

possible to meet the grand challenges of the post-genome era in *Arabidopsis* and other model organism.

REFERENCES

Abel S, Nguyen MD, Theologis A (1995) The PS-IAA4/5-like family of early auxin-inducible mRNAs in Arabidopsis thaliana. *J Mol Biol* **251:** 533-549

Abel S, Oeller PW, Theologis A (1994) Early auxin-induced genes encode short-lived nuclear proteins. *Proc Natl Acad Sci U S A* **91:** 326-330

Alonso JM, Stepanova AN, Leisse TJ, Kim CJ, Chen H, Shinn P, Stevenson DK, Zimmerman J, Barajas P, Cheuk R, Gadrinab C, Heller C, Jeske A, Koesema E, Meyers CC, Parker H, Prednis L, Ansari Y, Choy N, Deen H, Geralt M, Hazari N, Hom E, Karnes M, Mulholland C, Ndubaku R, Schmidt I, Guzman P, Aguilar-Henonin L, Schmid M, Weigel D, Carter DE, Marchand T, Risseeuw E, Brogden D, Zeko A, Crosby WL, Berry CC, Ecker JR (2003) Genome-wide insertional mutagenesis of Arabidopsis thaliana. *Science* **301:** 653-657

Alper MD, Ames BN (1975) Positive selection of mutants with deletions of the gal-chl region of the Salmonella chromosome as a screening procedure for mutagens that cause deletions. *J Bacteriol* **121:** 259-266

Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant Arabidopsis thaliana. *Nature* **408:** 796-815

Baudin A, Ozier-Kalogeropoulos O, Denouel A, Lacroute F, Cullin C (1993) A simple and efficient method for direct gene deletion in Saccharomyces cerevisiae. *Nucleic Acids Res* **21:** 3329-3330

Bechtold N, Pelletier G (1998) In planta Agrobacterium-mediated transformation of adult Arabidopsis thaliana plants by vacuum infiltration. *Methods Mol Biol* **82:** 259-266

Bennett MD, Leitch IJ, Price HJ, Johnston JS (2003) Comparisons with Caenorhabditis (approximately 100 Mb) and Drosophila (approximately 175 Mb) using flow cytometry show genome size in Arabidopsis to be approximately 157 Mb and thus approximately 25% larger than the Arabidopsis genome initiative estimate of approximately 125 Mb. *Ann Bot* **91:** 547-557

Benzinger R, Enquist LW, Skalka A (1975) Transfection of Escherichia coli spheroplasts. V. Activity of recBC nuclease in rec+ and rec minus spheroplasts measured with different forms of bacteriophage DNA. *J Virol* **15:** 861-871

Birnbaum K, Shasha DE, Wang JY, Jung JW, Lambert GM, Galbraith DW, Benfey PN (2003) A gene expression map of the Arabidopsis root. *Science* **302:** 1956-1960

Brown PO, Botstein D (1999) Exploring the new world of the genome with DNA microarrays. *Nat Genet* **21:** 33-37

Bubner B, Gase K, Baldwin IT (2004) Two-fold differences are the detection limit for determining transgene copy numbers in plants by real-time PCR. *BMC Biotechnol* **4:** 14

Carrari F, Benech-Arnold R, Osuna-Fernandez R, Hopp E, Sanchez R, Iusem N, Lljavetzky D (2003) Genetic mapping of the Sorghum bicolor vp1 gene and its relationship with preharvest sprouting resistance. *Genome* **46:** 253-258

Cartwright DA, Brady SM, Orlando DA, Sturmfels B, Benfey PN (2009) Reconstructing spatiotemporal gene expression data from partial observations. *Bioinformatics* **25:** 2581-2587

Casamitjana-Martinez E, Hofhuis HF, Xu J, Liu CM, Heidstra R, Scheres B (2003) Root-specific CLE19 overexpression and the sol1/2 suppressors implicate a CLV-like pathway in the control of Arabidopsis root meristem maintenance. *Curr Biol* **13:** 1435-1441

Chalfie M, Tu Y, Euskirchen G, Ward WW, Prasher DC (1994) Green fluorescent protein as a marker for gene expression. *Science* **263:** 802-805

Chang IF, Szick-Miranda K, Pan S, Bailey-Serres J (2005) Proteomic characterization of evolutionarily conserved and variable proteins of Arabidopsis cytosolic ribosomes. *Plant Physiol* **137:** 848-862

Cheeseman IM, Desai A (2005) A combined approach for the localization and tandem affinity purification of protein complexes from metazoans. *Sci STKE* **2005:** pl1

Chomczynski P, Sacchi N (1987) Single-step method of RNA isolation by acid guanidinium thiocyanate-phenol-chloroform extraction. *Anal Biochem* **162:** 156-159

Clough SJ, Bent AF (1998) Floral dip: a simplified method for Agrobacterium-mediated transformation of Arabidopsis thaliana. *Plant J* **16:** 735-743

Coelho SM, Peters AF, Charrier B, Roze D, Destombe C, Valero M, Cock JM (2007) Complex life cycles of multicellular eukaryotes: new approaches based on the use of model organisms. *Gene* **406:** 152-170

Colon-Carmona A, Chen DL, Yeh KC, Abel S (2000) Aux/IAA proteins are phosphorylated by phytochrome in vitro. *Plant Physiol* **124:** 1728-1738

Comai L, Henikoff S (2006) TILLING: practical single-nucleotide mutation discovery. *Plant J* **45:** 684-694

Conner TW, Goekjian VH, LaFayette PR, Key JL (1990) Structure and expression of two auxin-inducible genes from Arabidopsis. *Plant Mol Biol* **15:** 623-632

Costantino N, Court DL (2003) Enhanced levels of lambda Red-mediated recombinants in mismatch repair mutants. *Proc Natl Acad Sci U S A* **100:** 15748-15753

Cutler SR, Ehrhardt DW, Griffitts JS, Somerville CR (2000) Random GFP::cDNA fusions enable visualization of subcellular structures in cells of Arabidopsis at a high frequency. *Proc Natl Acad Sci U S A* **97:** 3718-3723

Datsenko KA, Wanner BL (2000) One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proc Natl Acad Sci U S A* **97:** 6640-6645

Datta S, Costantino N, Court DL (2006) A set of recombineering plasmids for gram-negative bacteria. *Gene* **379:** 109-115

Dharmasiri N, Dharmasiri S, Estelle M (2005) The F-box protein TIR1 is an auxin receptor. *Nature* **435:** 441-445

Doyle JJ, Doyle JL (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **19**

Dreher KA, Brown J, Saw RE, Callis J (2006) The Arabidopsis Aux/IAA protein family has diversified in degradation and auxin responsiveness. *Plant Cell* **18:** 699-714

Dubrovsky JG, Sauer M, Napsucialy-Mendivil S, Ivanchenko MG, Friml J, Shishkova S, Celenza J, Benkova E (2008) Auxin acts as a local morphogenetic trigger to specify lateral root founder cells. *Proc Natl Acad Sci U S A* **105:** 8790-8794

Ejsmont RK, Sarov M, Winkler S, Lipinski KA, Tomancak P (2009) A toolkit for high-throughput, cross-species gene engineering in Drosophila. *Nat Methods* **6:** 435-437

Ellis HM, Yu D, DiTizio T, Court DL (2001) High efficiency mutagenesis, repair, and engineering of chromosomal DNA using single-stranded oligonucleotides. *Proc Natl Acad Sci U S A* **98:** 6742-6746

Escobar C, Hernandez LE, Jimenez A, Creissen G, Ruiz MT, Mullineaux PM (2003) Transient expression of Arabidopsis thaliana ascorbate peroxidase 3 in Nicotiana benthamiana plants infected with recombinant potato virus X. *Plant Cell Rep* **21:** 699-704

Fizames C, Munos S, Cazettes C, Nacry P, Boucherez J, Gaymard F, Piquemal D, Delorme V, Commes T, Doumas P, Cooke R, Marti J, Sentenac H, Gojon A (2004) The Arabidopsis root transcriptome by serial analysis of gene expression. Gene identification using the genome sequence. *Plant Physiol* **134:** 67-80

Forsbach A, Schubert D, Lechtenberg B, Gils M, Schmidt R (2003) A comprehensive characterization of single-copy T-DNA insertions in the Arabidopsis thaliana genome. *Plant Mol Biol* **52:** 161-176

Fukaki H, Nakao Y, Okushima Y, Theologis A, Tasaka M (2005) Tissue-specific expression of stabilized SOLITARY-ROOT/IAA14 alters lateral root development in Arabidopsis. *Plant J* **44:** 382-395

Fukaki H, Tameda S, Masuda H, Tasaka M (2002) Lateral root formation is blocked by a gain-of-function mutation in the SOLITARY-ROOT/IAA14 gene of Arabidopsis. *Plant J* **29:** 153-168

Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, Rick JM, Michon AM, Cruciat CM, Remor M, Hofert C, Schelder M, Brajenovic M, Ruffner H, Merino A, Klein K, Hudak M, Dickson D, Rudi T, Gnau V, Bauch A, Bastuck S, Huhse B, Leutwein C, Heurtier MA, Copley RR, Edelmann A, Querfurth E, Rybin V, Drewes G, Raida M, Bouwmeester T, Bork P, Seraphin B, Kuster B, Neubauer G, Superti-Furga G (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* **415:** 141-147

Gong S, Yang XW, Li C, Heintz N (2002) Highly efficient modification of bacterial artificial chromosomes (BACs) using novel shuttle vectors containing the R6Kgamma origin of replication. *Genome Res* **12:** 1992-1998

Gong S, Zheng C, Doughty ML, Losos K, Didkovsky N, Schambra UB, Nowak NJ, Joyner A, Leblanc G, Hatten ME, Heintz N (2003) A gene expression atlas of the central nervous system based on bacterial artificial chromosomes. *Nature* **425:** 917-925

Gray WM, Kepinski S, Rouse D, Leyser O, Estelle M (2001) Auxin regulates SCF(TIR1)-dependent degradation of AUX/IAA proteins. *Nature* **414:** 271-276

Guilfoyle TJ, Hagen G (2007) Auxin response factors. *Curr Opin Plant Biol* **10:** 453-460

Haas BJ, Wortman JR, Ronning CM, Hannick LI, Smith RK, Jr., Maiti R, Chan AP, Yu C, Farzad M, Wu D, White O, Town CD (2005) Complete reannotation of the Arabidopsis genome: methods, tools, protocols and the final release. *BMC Biol* **3:** 7

Hagen G, Guilfoyle T (2002) Auxin-responsive gene expression: genes, promoters and regulatory factors. *Plant Mol Biol* **49:** 373-385

Hamann T, Benkova E, Baurle I, Kientz M, Jurgens G (2002) The Arabidopsis BODENLOS gene encodes an auxin response protein inhibiting MONOPTEROS-mediated embryo patterning. *Genes Dev* **16:** 1610-1615

Hamann T, Mayer U, Jurgens G (1999) The auxin-insensitive bodenlos mutation affects primary root formation and apical-basal patterning in the Arabidopsis embryo. *Development* **126:** 1387-1395

Hamilton CM (1997) A binary-BAC system for plant transformation with high-molecular-weight DNA. *Gene* **200:** 107-116

Hamilton CM, Frary A, Lewis C, Tanksley SD (1996) Stable transfer of intact high molecular weight DNA into plant chromosomes. *Proc Natl Acad Sci U S A* **93:** 9975-9979

Heim R, Cubitt AB, Tsien RY (1995) Improved green fluorescence. *Nature* **373:** 663-664

Heim R, Prasher DC, Tsien RY (1994) Wavelength mutations and posttranslational autoxidation of green fluorescent protein. *Proc Natl Acad Sci U S A* **91:** 12501-12504

Himelblau E, Mira H, Lin SJ, Culotta VC, Penarrubia L, Amasino RM (1998) Identification of a functional homolog of the yeast copper homeostasis gene ATX1 from Arabidopsis. *Plant Physiol* **117:** 1227-1234

Hoskins RA, Nelson CR, Berman BP, Laverty TR, George RA, Ciesiolka L, Naeemuddin M, Arenson AD, Durbin J, David RG, Tabor PE, Bailey MR, DeShazo DR, Catanese J, Mammoser A, Osoegawa K, de Jong PJ, Celniker SE, Gibbs RA, Rubin GM, Scherer SE (2000) A BAC-based physical map of the major autosomes of Drosophila melanogaster. *Science* **287:** 2271-2274

Huh WK, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK (2003) Global analysis of protein localization in budding yeast. *Nature* **425:** 686-691

Huq E, Al-Sady B, Quail PH (2003) Nuclear translocation of the photoreceptor phytochrome B is necessary for its biological function in seedling photomorphogenesis. *Plant J* **35:** 660-664

Ichikawa T, Nakazawa M, Kawashima M, Iizumi H, Kuroda H, Kondou Y, Tsuhara Y, Suzuki K, Ishikawa A, Seki M, Fujita M, Motohashi R, Nagata N, Takagi T, Shinozaki K, Matsui M (2006) The FOX hunting system: an alternative gain-of-function gene hunting technique. *Plant J* **48:** 974-985

Imam AM, Patrinos GP, de Krom M, Bottardi S, Janssens RJ, Katsantoni E, Wai AW, Sherratt DJ, Grosveld FG (2000) Modification of human beta-globin locus PAC clones by homologous recombination in Escherichia coli. *Nucleic Acids Res* **28:** E65

Ioannou PA, Amemiya CT, Garnes J, Kroisel PM, Shizuya H, Chen C, Batzer MA, de Jong PJ (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* **6:** 84-89

Jamsai D, Orford M, Nefedov M, Fucharoen S, Williamson R, Ioannou PA (2003) Targeted modification of a human beta-globin locus BAC clone using GET Recombination and an I-Scei counterselection cassette. *Genomics* **82:** 68-77

Jefferson RA, Kavanagh TA, Bevan MW (1987) GUS fusions: beta-glucuronidase as a sensitive and versatile gene fusion marker in higher plants. *EMBO J* **6:** 3901-3907

Jenik PD, Barton MK (2005) Surge and destroy: the role of auxin in plant embryogenesis. *Development* **132:** 3577-3585

Junop MS, Yang W, Funchain P, Clendenin W, Miller JH (2003) In vitro and in vivo studies of MutS, MutL and MutH mutants: correlation of mismatch repair and DNA recombination. *DNA Repair (Amst)* **2:** 387-405

Kelley JM, Field CE, Craven MB, Bocskai D, Kim UJ, Rounsley SD, Adams MD (1999) High throughput direct end sequencing of BAC clones. *Nucleic Acids Res* **27:** 1539-1546

Kim BC, Soh MC, Kang BJ, Furuya M, Nam HG (1996) Two dominant photomorphogenic mutations of Arabidopsis thaliana identified as suppressor mutations of hy2. *Plant J* **9:** 441-456

Kouprina N, Eldarov M, Moyzis R, Resnick M, Larionov V (1994) A model system to assess the integrity of mammalian YACs during transformation and propagation in yeast. *Genomics* **21:** 7-17

Kumar A, Agarwal S, Heyman JA, Matson S, Heidtman M, Piccirillo S, Umansky L, Drawid A, Jansen R, Liu Y, Cheung KH, Miller P, Gerstein M, Roeder GS, Snyder M (2002) Subcellular localization of the yeast proteome. *Genes Dev* **16:** 707-719

Lafontaine D, Tollervey D (1996) One-step PCR mediated strategy for the construction of conditionally expressed and epitope tagged yeast proteins. *Nucleic Acids Res* **24:** 3469-3471

Lalioti M, Heath J (2001) A new method for generating point mutations in bacterial artificial chromosomes by homologous recombination in Escherichia coli. *Nucleic Acids Res* **29:** E14

Laplaze L, Parizot B, Baker A, Ricaud L, Martiniere A, Auguy F, Franche C, Nussaume L, Bogusz D, Haseloff J (2005) GAL4-GFP enhancer trap lines for genetic manipulation of lateral root development in Arabidopsis thaliana. *J Exp Bot* **56:** 2433-2442

Laxmi A, Pan J, Morsy M, Chen R (2008) Light plays an essential role in intracellular distribution of auxin efflux carrier PIN2 in Arabidopsis thaliana. *PLoS One* **3:** e1510

Lee EC, Yu D, Martinez de Velasco J, Tessarollo L, Swing DA, Court DL, Jenkins NA, Copeland NG (2001) A highly efficient Escherichia coli-based chromosome engineering system adapted for recombinogenic targeting and subcloning of BAC DNA. *Genomics* **73:** 56-65

Leyser O (2005) Auxin distribution and plant pattern formation: how many angels can dance on the point of PIN? *Cell* **121:** 819-822

Li XT, Costantino N, Lu LY, Liu DP, Watt RM, Cheah KS, Court DL, Huang JD (2003) Identification of factors influencing strand bias in oligonucleotide-mediated recombination in Escherichia coli. *Nucleic Acids Res* **31:** 6674-6687

Liscum E, Reed JW (2002) Genetics of Aux/IAA and ARF action in plant growth and development. *Plant Mol Biol* **49:** 387-400

Liu YG, Liu H, Chen L, Qiu W, Zhang Q, Wu H, Yang C, Su J, Wang Z, Tian D, Mei M (2002) Development of new transformation-competent artificial chromosome vectors and rice genomic libraries for efficient gene cloning. *Gene* **282:** 247-255

Liu YG, Shirano Y, Fukaki H, Yanai Y, Tasaka M, Tabata S, Shibata D (1999) Complementation of plant mutants with large genomic DNA fragments by a transformation-competent artificial chromosome vector accelerates positional cloning. *Proc Natl Acad Sci U S A* **96:** 6535-6540

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* **25:** 402-408

Mardis ER (2006) Anticipating the 1,000 dollar genome. *Genome Biol* **7:** 112

Maduro M, Pilgrim D (1996) Conservation of function and expression of unc-119 from two Caenorhabditis species despite divergence of non-coding DNA. *Gene* **183:** 77-85

Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH (1997) High throughput fingerprint analysis of large-insert clones. *Genome Res* **7:** 1072-1084

McPherson JD, Marra M, Hillier L, Waterston RH, Chinwalla A, Wallis J, Sekhon M, Wylie K, Mardis ER, Wilson RK, Fulton R, Kucaba TA, Wagner-McPherson C, Barbazuk WB, Gregory SG, Humphray SJ, French L, Evans RS, Bethel G, Whittaker A, Holden JL, McCann OT, Dunham A, Soderlund C, Scott CE, Bentley DR, Schuler G, Chen HC, Jang W, Green ED, Idol JR, Maduro VV, Montgomery KT, Lee E, Miller A, Emerling S, Kucherlapati, Gibbs R, Scherer S, Gorrell JH, Sodergren E, Clerc-Blankenburg K, Tabor P, Naylor S, Garcia D, de Jong PJ, Catanese JJ, Nowak N, Osoegawa K, Qin S, Rowen L, Madan A, Dors M, Hood L, Trask B, Friedman C, Massa H, Cheung VG, Kirsch IR, Reid T, Yonescu R, Weissenbach J, Bruls T, Heilig R, Branscomb E, Olsen A, Doggett N, Cheng JF, Hawkins T, Myers RM, Shang J, Ramirez L, Schmutz J, Velasquez O, Dixon K, Stone NE, Cox DR, Haussler D, Kent WJ, Furey T, Rogic S, Kennedy S, Jones S, Rosenthal A, Wen G, Schilhabel M, Gloeckner G, Nyakatura G, Siebert R, Schlegelberger B, Korenberg J, Chen XN, Fujiyama A, Hattori M, Toyoda A, Yada T, Park HS, Sakaki Y, Shimizu N, Asakawa S, Kawasaki K, Sasaki T, Shintani A, Shimizu A, Shibuya K, Kudoh J, Minoshima S, Ramser J, Seranski P, Hoff C, Poustka A, Reinhardt R, Lehrach H (2001) A physical map of the human genome. *Nature* **409:** 934-941

Mira H, Vilar M, Perez-Paya E, Penarrubia L (2001) Functional and conformational properties of the exclusive C-domain from the Arabidopsis copper chaperone (CCH). *Biochem J* **357:** 545-549

Misulovin Z, Yang XW, Yu W, Heintz N, Meffre E (2001) A rapid method for targeted modification and screening of recombinant bacterial artificial chromosome. *J Immunol Methods* **257:** 99-105

Mozo T, Dewar K, Dunn P, Ecker JR, Fischer S, Kloska S, Lehrach H, Marra M, Martienssen R, Meier-Ewert S, Altmann T (1999) A complete BAC-based physical map of the Arabidopsis thaliana genome. *Nat Genet* **22:** 271-275

Muday GK, DeLong A (2001) Polar auxin transport: controlling where and how much. *Trends Plant Sci* **6:** 535-542

Muto H, Watahiki MK, Nakamoto D, Kinjo M, Yamamoto KT (2007) Specificity and similarity of functions of the Aux/IAA genes in auxin signaling of Arabidopsis revealed by promoter-exchange experiments among MSG2/IAA19, AXR2/IAA7, and SLR/IAA14. *Plant Physiol* **144:** 187-196

Muyrers JP, Zhang Y, Benes V, Testa G, Ansorge W, Stewart AF (2000) Point mutation of bacterial artificial chromosomes by ET recombination. *EMBO Rep* **1:** 239-243

Muyrers JP, Zhang Y, Testa G, Stewart AF (1999) Rapid modification of bacterial artificial chromosomes by ET-recombination. *Nucleic Acids Res* **27:** 1555-1557

Nagpal P, Walker LM, Young JC, Sonawala A, Timpte C, Estelle M, Reed JW (2000) AXR2 encodes a member of the Aux/IAA protein family. *Plant Physiol* **123:** 563-574

Nguyen AW, Daugherty PS (2005) Evolutionary optimization of fluorescent proteins for intracellular FRET. *Nat Biotechnol* **23:** 355-360

Okushima Y, Overvoorde PJ, Arima K, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Lui A, Nguyen D, Onodera C, Quach H, Smith A, Yu G, Theologis A (2005) Functional genomic analysis of the AUXIN RESPONSE FACTOR gene family members in Arabidopsis thaliana: unique and overlapping functions of ARF7 and ARF19. *Plant Cell* **17:** 444-463

Osoegawa K, Tateno M, Woon PY, Frengen E, Mammoser AG, Catanese JJ, Hayashizaki Y, de Jong PJ (2000) Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res* **10:** 116-128

Ostergaard L, Yanofsky MF (2004) Establishing gene function by mutagenesis in Arabidopsis thaliana. *Plant J* **39:** 682-696

Ouellet F, Overvoorde PJ, Theologis A (2001) IAA17/AXR3: biochemical insight into an auxin mutant phenotype. *Plant Cell* **13:** 829-841

Overvoorde PJ, Okushima Y, Alonso JM, Chan A, Chang C, Ecker JR, Hughes B, Liu A, Onodera C, Quach H, Smith A, Yu G, Theologis A (2005) Functional genomic analysis

of the AUXIN/INDOLE-3-ACETIC ACID gene family members in Arabidopsis thaliana. *Plant Cell* **17:** 3282-3300

Parinov S, Sundaresan V (2000) Functional genomics in Arabidopsis: large-scale insertional mutagenesis complements the genome sequencing project. *Curr Opin Biotechnol* **11:** 157-161

Peterson KR, Navas PA, Li Q, Stamatoyannopoulos G (1998) LCR-dependent gene expression in beta-globin YAC transgenics: detailed structural studies validate functional analysis even in the presence of fragmented YACs. *Hum Mol Genet* **7:** 2079-2088

Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res* **29:** e45

Phelps CB, Brand AH (1998) Ectopic gene expression in Drosophila using GAL4 system. *Methods* **14:** 367-379

Ploense SE, Wu MF, Nagpal P, Reed JW (2009) A gain-of-function mutation in IAA18 alters Arabidopsis embryonic apical patterning. *Development* **136:** 1509-1517

Poser I, Sarov M, Hutchins JR, Heriche JK, Toyoda Y, Pozniakovsky A, Weigl D, Nitzsche A, Hegemann B, Bird AW, Pelletier L, Kittler R, Hua S, Naumann R, Augsburg M, Sykora MM, Hofemeister H, Zhang Y, Nasmyth K, White KP, Dietzel S, Mechtler K, Durbin R, Stewart AF, Peters JM, Buchholz F, Hyman AA (2008) BAC TransgeneOmics: a high-throughput method for exploration of protein function in mammals. *Nat Methods* **5:** 409-415

Ramos JA, Zenser N, Leyser O, Callis J (2001) Rapid degradation of auxin/indoleacetic acid proteins requires conserved amino acids of domain II and is proteasome dependent. *Plant Cell* **13:** 2349-2360

Remington DL, Vision TJ, Guilfoyle TJ, Reed JW (2004) Contrasting modes of diversification in the Aux/IAA and ARF gene families. *Plant Physiol* **135:** 1738-1752

Rensink WA, Buell CR (2004) Arabidopsis to rice. Applying knowledge from a weed to enhance our understanding of a crop species. *Plant Physiol* **135:** 622-629

Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* **17:** 1030-1032

Robinson SJ, Cram DJ, Lewis CT, Parkin IA (2004) Maximizing the efficacy of SAGE analysis identifies novel transcripts in Arabidopsis. *Plant Physiol* **136:** 3223-3233

Roca AI, Cox MM (1990) The RecA protein: structure and function. *Crit Rev Biochem Mol Biol* **25:** 415-456

Rogg LE, Lasswell J, Bartel B (2001) A gain-of-function mutation in IAA28 suppresses lateral root development. *Plant Cell* **13:** 465-480

Rouse D, Mackay P, Stirnberg P, Estelle M, Leyser O (1998) Changes in auxin response from mutations in an AUX/IAA gene. *Science* **279:** 1371-1373

Sarov M, Schneider S, Pozniakovski A, Roguev A, Ernst S, Zhang Y, Hyman AA, Stewart AF (2006) A recombineering pipeline for functional genomics applied to Caenorhabditis elegans. *Nat Methods* **3:** 839-844

Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Scholkopf B, Weigel D, Lohmann JU (2005) A gene expression map of Arabidopsis thaliana development. *Nat Genet* **37:** 501-506

Schwab R, Palatnik JF, Riester M, Schommer C, Schmid M, Weigel D (2005) Specific effects of microRNAs on the plant transcriptome. *Dev Cell* **8:** 517-527

Seki M, Narusaka M, Kamiya A, Ishida J, Satou M, Sakurai T, Nakajima M, Enju A, Akiyama K, Oono Y, Muramatsu M, Hayashizaki Y, Kawai J, Carninci P, Itoh M, Ishii Y, Arakawa T, Shibata K, Shinagawa A, Shinozaki K (2002) Functional annotation of a full-length Arabidopsis cDNA collection. *Science* **296:** 141-145

Sheng Y, Mancino V, Birren B (1995) Transformation of Escherichia coli with large DNA molecules by electroporation. *Nucleic Acids Res* **23:** 1990-1996

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. *Proc Natl Acad Sci U S A* **89:** 8794-8797

Skulachev MV, Ivanov PA, Karpova OV, Korpela T, Rodionova NP, Dorokhov YL, Atabekov JG (1999) Internal initiation of translation directed by the 5'-untranslated region of the tobamovirus subgenomic RNA I(2). *Virology* **263:** 139-154

Song J, Bradeen JM, Naess SK, Helgeson JP, Jiang J (2003) BIBAC and TAC clones containing potato genomic DNA fragments larger than 100 kb are not stable in Agrobacterium. *Theor Appl Genet* **107:** 958-964

Stepanova AN, Hoyt JM, Hamilton AA, Alonso JM (2005) A Link between ethylene and auxin uncovered by the characterization of two root-specific ethylene-insensitive mutants in Arabidopsis. *Plant Cell* **17:** 2230-2242

Swaminathan S, Ellis HM, Waters LS, Yu D, Lee EC, Court DL, Sharan SK (2001) Rapid engineering of bacterial artificial chromosomes using oligonucleotides. *Genesis* **29:** 14-21

Szostak JW, Orr-Weaver TL, Rothstein RJ, Stahl FW (1983) The double-strand-break repair model for recombination. *Cell* **33:** 25-35

Tan X, Calderon-Villalobos LI, Sharon M, Zheng C, Robinson CV, Estelle M, Zheng N (2007) Mechanism of auxin perception by the TIR1 ubiquitin ligase. *Nature* **446:** 640-645

Tatematsu K, Kumagai S, Muto H, Sato A, Watahiki MK, Harper RM, Liscum E, Yamamoto KT (2004) MASSUGU2 encodes Aux/IAA19, an auxin-regulated protein that functions together with the transcriptional activator NPH4/ARF7 to regulate differential growth responses of hypocotyl and formation of lateral roots in Arabidopsis thaliana. *Plant Cell* **16:** 379-393

Taylor CB (1997) Promoter fusion analysis: An insufficient measure of gene expression. *Plant Cell* **9:** 273-275

Theologis A, Huynh TV, Davis RW (1985) Rapid induction of specific mRNAs by auxin in pea epicotyl tissue. *J Mol Biol* **183:** 53-68

Tian GW, Mohanty A, Chary SN, Li S, Paap B, Drakakaki G, Kopec CD, Li J, Ehrhardt D, Jackson D, Rhee SY, Raikhel NV, Citovsky V (2004) High-throughput fluorescent tagging of full-length Arabidopsis gene products in planta. *Plant Physiol* **135:** 25-38

Tian Q, Reed JW (1999) Control of auxin-regulated root development by the Arabidopsis thaliana SHY2/IAA3 gene. *Development* **126:** 711-721

Till BJ, Colbert T, Codomo C, Enns L, Johnson J, Reynolds SH, Henikoff JG, Greene EA, Steine MN, Comai L, Henikoff S (2006) High-throughput TILLING for Arabidopsis. *Methods Mol Biol* **323:** 127-135

Till BJ, Colbert T, Tompa R, Enns LC, Codomo CA, Johnson JE, Reynolds SH, Henikoff JG, Greene EA, Steine MN, Comai L, Henikoff S (2003) High-throughput TILLING for functional genomics. *Methods Mol Biol* **236:** 205-220

Tiwari SB, Hagen G, Guilfoyle TJ (2004) Aux/IAA proteins contain a potent transcriptional repression domain. *Plant Cell* **16:** 533-543

Tiwari SB, Wang XJ, Hagen G, Guilfoyle TJ (2001) AUX/IAA proteins are active repressors, and their stability and activity are modulated by auxin. *Plant Cell* **13:** 2809-2822

Tursun B, Cochella L, Carrera I, Hobert O (2009) A toolkit and robust pipeline for the generation of fosmid-based reporter genes in C. elegans. *PLoS One* **4:** e4625

Ulmasov T, Hagen G, Guilfoyle TJ (1999) Activation and repression of transcription by auxin-response factors. *Proc Natl Acad Sci U S A* **96:** 5844-5849

Van Lijsebettens M, Vanderhaeghen R, De Block M, Bauw G, Villarroel R, Van Montagu M (1994) An S18 ribosomal protein gene copy at the Arabidopsis PFL locus

affects plant development by its specific expression in meristems. *EMBO J* **13:** 3378-3388

Venken KJ, Kasprowicz J, Kuenen S, Yan J, Hassan BA, Verstreken P (2008) Recombineering-mediated tagging of Drosophila genomic constructs for in vivo localization and acute protein inactivation. *Nucleic Acids Res* **36:** e114

Vision TJ, Brown DG, Tanksley SD (2000) The origins of genomic duplications in Arabidopsis. *Science* **290:** 2114-2117

Wang J, Sarov M, Rientjes J, Fu J, Hollak H, Kranz H, Xie W, Stewart AF, Zhang Y (2006) An improved recombineering approach by adding RecA to lambda Red recombination. *Mol Biotechnol* **32:** 43-53

Warming S, Costantino N, Court DL, Jenkins NA, Copeland NG (2005) Simple and highly efficient BAC recombineering using galK selection. *Nucleic Acids Res* **33:** e36

Weigel D, Ahn JH, Blazquez MA, Borevitz JO, Christensen SK, Fankhauser C, Ferrandiz C, Kardailsky I, Malancharuvil EJ, Neff MM, Nguyen JT, Sato S, Wang ZY, Xia Y, Dixon RA, Harrison MJ, Lamb CJ, Yanofsky MF, Chory J (2000) Activation tagging in Arabidopsis. *Plant Physiol* **122:** 1003-1013

Weijers D, Benkova E, Jager KE, Schlereth A, Hamann T, Kientz M, Wilmoth JC, Reed JW, Jurgens G (2005) Developmental specificity of auxin response by pairs of ARF and Aux/IAA transcriptional regulators. *EMBO J* **24:** 1874-1885

Worley CK, Zenser N, Ramos J, Rouse D, Leyser O, Theologis A, Callis J (2000) Degradation of Aux/IAA proteins is essential for normal auxin signalling. *Plant J* **21:** 553-562

Yamada K, Lim J, Dale JM, Chen H, Shinn P, Palm CJ, Southwick AM, Wu HC, Kim C, Nguyen M, Pham P, Cheuk R, Karlin-Newmann G, Liu SX, Lam B, Sakano H, Wu T, Yu G, Miranda M, Quach HL, Tripp M, Chang CH, Lee JM, Toriumi M, Chan MM, Tang CC, Onodera CS, Deng JM, Akiyama K, Ansari Y, Arakawa T, Banh J, Banno F, Bowser L, Brooks S, Carninci P, Chao Q, Choy N, Enju A, Goldsmith AD, Gurjal M, Hansen NF, Hayashizaki Y, Johnson-Hopson C, Hsuan VW, Iida K, Karnes M, Khan S, Koesema E, Ishida J, Jiang PX, Jones T, Kawai J, Kamiya A, Meyers C, Nakajima M, Narusaka M, Seki M, Sakurai T, Satou M, Tamse R, Vaysberg M, Wallender EK, Wong C, Yamamura Y, Yuan S, Shinozaki K, Davis RW, Theologis A, Ecker JR (2003) Empirical analysis of transcriptional activity in the Arabidopsis genome. *Science* **302:** 842-846

Yamamoto T, Moerschell RP, Wakem LP, Ferguson D, Sherman F (1992) Parameters affecting the frequencies of transformation and co-transformation with synthetic oligonucleotides in yeast. *Yeast* **8:** 935-948

Yamamoto YY, Tsuhara Y, Gohda K, Suzuki K, Matsui M (2003) Gene trapping of the Arabidopsis genome with a firefly luciferase reporter. *Plant J* **35:** 273-283

Yang X, Lee S, So JH, Dharmasiri S, Dharmasiri N, Ge L, Jensen C, Hangarter R, Hobbie L, Estelle M (2004) The IAA1 protein is encoded by AXR5 and is a substrate of SCF(TIR1). *Plant J* **40:** 772-782

Yang XW, Model P, Heintz N (1997) Homologous recombination based modification in Escherichia coli and germline transmission in transgenic mice of a bacterial artificial chromosome. *Nat Biotechnol* **15:** 859-865

Yates JR, III (2004) Mass spectral analysis in proteomics. *Annu Rev Biophys Biomol Struct* **33:** 297-316

Yu D, Ellis HM, Lee EC, Jenkins NA, Copeland NG, Court DL (2000) An efficient recombination system for chromosome engineering in Escherichia coli. *Proc Natl Acad Sci U S A* **97:** 5978-5983

Yu D, Sawitzke JA, Ellis H, Court DL (2003) Recombineering with overlapping single-stranded DNA oligonucleotides: testing a recombination intermediate. *Proc Natl Acad Sci U S A* **100:** 7207-7212

Zenser N, Dreher KA, Edwards SR, Callis J (2003) Acceleration of Aux/IAA proteolysis is specific for auxin and independent of AXR1. *Plant J* **35:** 285-294

Zhang P, Li MZ, Elledge SJ (2002) Towards genetic genome projects: genomic library screening and gene-targeting vector construction in a single step. *Nat Genet* **30:** 31-39

Zhang Y, Buchholz F, Muyrers JP, Stewart AF (1998) A new logic for DNA engineering using recombination in Escherichia coli. *Nat Genet* **20:** 123-128

Zhao S (2001) A comprehensive BAC resource. *Nucleic Acids Res* **29:** 141-143

Zhou CQ, Mai QY, Li T, Zhuang GL (2004) Cryopreservation of human embryonic stem cells by vitrification. *Chin Med J (Engl)* **117:** 1050-1055

Zybailov B, Rutschow H, Friso G, Rudella A, Emanuelsson O, Sun Q, van Wijk KJ (2008) Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. *PLoS One* **3:** e1994