Genetic and environmental modulation of phenotypic variation in *Arabidopsis thaliana*

Dissertation

der Mathematisch-Naturwissenschaftlichen Fakultät

der Eberhard Karls Universität Tübingen

zur Erlangung des Grades eines

Doktors der Naturwissenschaften

(Dr. rer. nat.)

vorgelegt von
Mrs Sridevi Sureshkumar
aus Dharmapuri, Indien

Tübingen
2011

Tag der mündlichen Qualifikation:	26. 01. 2011
Dekan :	Prof. Dr. Wolfgang Rosenstiel
1.Berichterstatter:	Prof. Dr. Detlef Weigel
2.Berichterstatter:	Prof. Dr. Klaus Harter

Summary

Pre and post-embryonic plant development is modulated by multiple environmental cues and endogenous factors. Since plants are sessile organisms, they are highly adaptive to varied environmental conditions. Climate change associated changes in environmental conditions have predicted negative impacts on agricultural crop production. Among the environmental factors light and temperature are the major factors, which modulate plant growth and development throughout their life cycle. Responding to these parameters during development is essential for the plants to sustain their ecology niches. One of the crucial developmental events is switching the vegetative phase into reproductive phase, where plant flower and set seeds to progress to the next generations. Mutant analysis and studies on natural genetic variation have shown that multiple environmental signals regulate the flowering time in model plant *Arabidopsis thaliana*. During my doctoral studies I had exploited the natural *Arabidopsis thaliana* wild strains as model to study for the light and temperature implications on genetic architecture and extreme phenotypic variation.

During a phenotypic screen of wild *Arabidopsis* strains in multiple conditions, we found some of the strains displayed an unusual early flowering behaviour, compared to laboratory strain Col-0 at short day conditions (16 hours dark and 8 hrs of light). I had carried out the genetic characterisation of some of the interesting strains to understand the genetic basis of these flowering time differences and environmental interactions. I combined the F2 analysis with the analysis of recombinant inbred lines to discover the

genetic basis of early flowering behaviour. Genetic analysis of these stains revealed that they do not carry a same genetic mechanism for the early flowering behaviour. The F2 distribution of some strains indicated that more than one locus affect the phenotype. The previous research has been shown that deletion in gene called *FLOWERING LOCUS M* (*FLM*), lead to early flowering phenotype. I had carried out genotyping these lines for *FLM* deletions. One of the stain called Ei-6 displayed a deletion in this locus, implied that early flowering nature of the Ei-6 is due to *FLM* locus. Est-1 X Col-0 recombinant inbred lines the flowering time was mapped to two independent Quantitative trait loci (QTL) under short day conditions. We have learnt from Est-1 X Col-0 that the generally accepted correlation between days to flowering and total leaf number, which are the proxies for measuring flowering time may not hold true all the time and there is a genetic basis of the number of leaves produced per day.

I had extended the QTL approach to study the temperature regulation of flowering time. Two RIL lines C24 X Col-0 and Nd-1 X Col-0, used for the case study and measured the flowering time at higher temperature 27°C in short day. QTL analysis for flowering time at short day conditions in the Nd-1 x Col-0 had shown that the Nd-1 strain carries a deletion of *FLM* locus. The effect of this QTL was diminished at higher temperatures indicating that temperature effects are mediated through the same genetic cascade in which *FLM* operates. In addition to C24 X Col-0 RIL for flowering time mapped to chromosome IV around *FRIGIDA (FRI)* but failed to detect *FLC* suggesting that there may not be a functional difference between the Col-0 and C24 *FLC* alleles as

previously thought. During the analysis of flowering time, we also revealed interesting transgression of leaf serration, which mapped to two interacting loci.

My second chapter I had carried out molecular basis of flowering time in one of the early flowering strain called Fr2. We mapped the QTL region to *PHYTOCHROME C (PHYC)* and the sequencing of a *PHYC* cDNA from Fr-2 revealed a nonsense mutation in the first exon, which converts the K299 codon to a stop codon. In addition to the stop codon, the Fr-2 *PHYC* allele is highly polymorphic with 12 non-synonymous changes compared to Col-0. Ten of these substitutions were present in another common laboratory strain, L*er*. *phyc* mutants were never picked up in the L*er* background, which lead us to suspect the activity of *PHYC* allele. Through sequencing and by comparing the phenotypic effects of alleles across population, we demonstrate that there are two functionally distinct haplogroups for *PHYC* in *Arabidopsis* that display a latitudinal cline in their distribution. The more active Col-0 *PHYC* haplogroup was more frequent at northern cline, compare to L*er* haplogroups. In addition, we reveal alternatively splicing *at PHYC*, which adds additional complexity.

Hidden genetic variation has interesting evolutionary implications and many phenotypes are nullified in the normal conditions. The aim of chapter 3 is to understand the extreme phenotypes in the higher temperature, which plants rarely experience. We have mapped a conditional growth defect to a gene called *ISOPROPYL MALATE ISOMERASE LARGE SUBUNIT 1(IIL1)* in the one of strain called Bur-0 that displayed abnormal growth arrest phenotype compare to other reference strains. *IIL1* gene possesses a long intronic trinucleotide repeats TTC/GAA, similar to the GAA/TTC

expansion observed un the human hereditary disease Friedreich ataxia. We demonstrate through transgenics that over expressed Bur-0 cDNA could rescue the phenotype. Through suppressor screens, we demonstrate that the expanded repeat is the polymorphism that underlies the growth arrest phenotype. This is the strong evidence that the triplet repeat mediated genetic defects can be seen in plants as well. Given that this is the first example of a triplet expansion associated genetic defect described outside humans, the feasibility to do genetic screens and the striking parallels this system displays with that of Friedreich ataxia make the *IIL1* case an excellent model for studying some of the fundamental aspects of triplet expansion associated genetic defects.

Introduction

Phenotypic variation

Phenotypic variations are resources for exploring genetic variation in the species. Deep insights about phenotypic variation and adaptation across the geographic regions started from Darwin's origin of species and subsequently phenotypic variation has been addressed in multiple species. However the genetic and molecular basis of phenotypic variation is poorly understood. Sometimes demographic factors and anchoring environment can buffer the traits from under selection (such as size, growth, and reproduction). For example, studies in sticklebacks (*Gasterostreus aculeatus*) reveal that the three spine fish adapted quickly from the marine to the fresh water due to the adaptive genetic variation in major effective gene called *Eda* locus, causing reduced armoured plate phenotypes in the fresh water fish (SCHLUTER *et al.*). Genetic loci *achaete-scute* region was found to be strongly associated with bristle number in *Drosophila melanogaster* populations (MACKAY and LANGLEY 1990).

Traditional method of analysing phenotypic variation involves quantitative genetic analysis including the identification and characterisation of the Quantitative Trait Locus (QTL). QTL studies largely reported that the substantial amount of genetic variation controlled by multiple loci. The quantitative variation in the phenotypic increases the genetic complexity in the populations and gives better probabilities for optimal adaptations (ORR 1998). Although QTL is one of the long established genetic tools used in laboratory and breeding, the identification of genes underlying the QTLs has become a reality only very recently. Fruit yield and organ size is important for the

domestic crops, the first QTL controlling the fruit size has been cloned between two wild variates tomato(*Lycopersicon esculentum*) where structural change in *fw2.2* allele (FRARY *et al.* 2000).

Plants are sessile and this forces them to adapt for environmental changes in contrast to animals, which can avoid a harsh environment. This makes plants an ideal model to study adaptive genetic variation. While QTL approaches are in principle possible in any organism, a well-characterized model organism provides additional advantages for QTL analysis. The model plant *Arabidopsis thaliana*, with its extensive genomic resources and with the variety of strains that show extensive phenotypic variation across several traits provides an ideal organisms to address some of the fundamental issues associated with phenotypic variation and selection. Natural occurring *Arabidopsis thaliana* habituated around the northern hemisphere, and colonized in northern America and central Asia and extended to west Asia. (More details please refer my thesis chapter 1 and chapter 2). *Arabidopsis thaliana* natural populations or so called ecotypes or wild strains harbors a rich phenotypic variation and maintained in different traits such as seed size, maturation and leaf morphology and reproduction (Flowering) in wild (KOORNNEEF *et al.* 2004). Extensive studies form wild *Arabidopsis* strains provide a genetically tractable model for evolutionary functional genomics (NORDBORG *et al.* 2005). In my doctoral research I had explored phenotypic variation in *Arabidopsis thaliana* to understand the genetic basis of phenotypic variation under varied environments. We have under taken a systemic analysis of flowering and light and temperature responses in different environments and identified potentially interesting

strains that can be used for understanding the genetic bases of novel flowering for light and temperature response.

Genetic basis for natural variation in *Arabidopsis thaliana*

Combination of recent break trough sequence technologies and QTL analysis expanded the knowledge of the genome wide genetic variation in *Arabidopsis thaliana* naturally occurring strains (CLARK *et al.* 2007b). Genetic basis for natural variation in Arabidopsis thaliana has been addressed in physiological traits, biochemical and developmental traits. One of the first studies with natural strains of *Arabidopsis* revealed a QTL for an early day length insensitive phenotype mapped to the photoreceptor *CRYPTOCHROME 2* and a natural variant in light sensitivity encodes for a photoreceptor phytochrome A. Subsequently several other examples exist in *Arabidopsis* arguing for the exploitation of *Arabidopsis* natural variation as a genetic resource. For example, allelic variation in the *DELAY OF GERMINATION 1(DOG1)* locus contributes to germination variation between two different strains (BENTSINK *et al.* 2006). Genome wide association studies in the Arabidopsis strains showed a regulatory variation in the *HKT1* gene assimilates different sodium uptake and also demonstrated that the clinal variation in the allelic frequency for salt tolerance (RUS *et al.* 2006). Enhanced fitness has been conferred by natural occurring variation in the Circadian clock has been extensively studied using *Arabidopsis* strains (MICHAEL *et al.* 2003; SWARUP *et al.* 1999). The root length variation in two different accessions resulted in variation in *BRX* locus controlling the root size (MOUCHEL *et al.* 2004). Inbred depression has been addressed trough natural variation (BOMBLIES *et al.* 2007).

Genetic basis for flowering time variation

Mutants analysis had shown that more than 80 loci could affect flowering time in Arabidopsis. Flowering is a crucial developmental process for the plant, and the timing of flowering should empower maximum reproductive success. Multiple environmental cues play a role in the crucial decision of the timing of flowering. Among the cues light and temperature a major role in the plant development. Naturally occurring wild strains showed an extensive variation in there flowering behaviour, where most of them flower early in long days and some of them need exposure winter conditions (vernalization) for earlier flowering. Two major loci *FRIGIDA (FRI)* and *FLOWING LOCUS C (FLC)* are the major determinants of this life history behaviour (GAZZANI *et al.* 2003; JOHANSON *et al.* 2000). Most of the early flowering strains carry a deletion at *FRI* locus. In contrast to *FRI*, alleles of other loci that modulate variant flowering responses tend to be rare (ALONSO-BLANCO *et al.* 2009). The QTL mapping and subsequent cloning have identified a dominant allele for the blue light receptor *CRYPTOCHROME 2 (CRY2*) between two natural strains (EL-DIN EL-ASSAL *et al.* 2001) and a naturally occurring deletion in the floral repressor *FLOWERING LOCUS M(FLM)* to be conferring early flower under short day conditions (WERNER *et al.* 2005b). Natural variation in light sensitivity has also been studied and a dominant variant of the photoreceptor *PHYTOCHROME A (PHYA)* has been shown to underlie a reduced red light response in the wild strain Lm-2 (MALOOF *et al.* 2001). In addition, a naturally occurring mutation in *PHYTOCHROME D (PHYD)* has also been conferring early flowering (AUKERMAN *et al.* 1997).

Genetic basis for temperature regulation of flowering time

Temperature is one of the major environmental cues effect the flowering time and other phenotypes. However, we know very little about temperature regulation flowering time. Blazquez et al showed temperature may modulate flowering via the autonomous pathway genes (more detail please refer chapter 2). Recently studies showed the thermal induction on flowering time requires another MADS box gene called *SHORT VEGETATIVE PHASE (SVP)* (LEE *et al.* 2007). *SVP* happened to found function within the thermosensory pathway and downstream of *FCA* and *FVE* autonomous pathway genes. Studies with mutants reveal that thermal induction is independent of *CO*. It also appears to be independent of the SA-dependent stress pathway or the hormonal pathway but mediated via the floral integrator *FT* (BALASUBRAMANIAN *et al.* 2006). Recent studies from Wigge lab , suggests a role for histones and chromatin remodeling in thermal sensing (KUMAR and WIGGE 2010).

Lessons from hidden genetic variation

Hidden genetic variation has interesting evolutionary implications and many phenotypes are nullified in the normal conditions. The part of genetic variation does not translate into phenotypic differences, in given environment. When changes in the environmental conditions could trigger novel phenotypes. It has been believed that the hidden genetic variation has a genetic influence on the evolutionary traits (LE ROUZIC *et al.* 2007). By subjecting strains to abnormal conditions we reveal interesting cryptic phenotypes and show that some striking observations that have implications for human genetic conditions (Chapter 3). Repeat expansions underlies more than 40 neurological hereditary diseases

in the humans. Based on their occurrence in the gene the repeat expansion disorders are divided into two major classes namely coding region disorders and non-coding region diseases. Triplets in coding regions typically lead to poly-glutamine (CAG) and poly-alanine (GCU GCC GCA GCG) tracts and in non-coding regions triplet expansion can modulate gene expression (e,g., Friedreich ataxia) (GATCHEL and ZOGHBI 2005). So far triplet repeat mediated genetic disease addressed only in humans.

In this thesis, I have addressed some of the basic aspects of the molecular basis of phenotypic variation. In the first chapter, I discuss the genetic architecture of variation in flowering time revealed under non-inductive short day conditions. I show, unlike the vernalization response, the genetic basis of flowering time variation in the photoperiodic response is complex and does not appear to have many major effect loci contributing to this variation. In the second chapter, I specifically characterize one strain for its early flowering behavior and show that it is due to a mutation in the photoreceptor PHYTOCHROME C. In addition, I demonstrate that *PHYC* is present in two forms in *Arabidopsis* that are functionally distinct and suggest that *PHYC* might be under adaptive selection. In the last chapter, I discuss one of the cryptic genetic variations revealed under high temperature conditions and show the genetic basis for the same. This study has led to the identification of the first non-human example for a triplet expansion associated genetic defect, which shares striking parallels with that of the human condition Friedreich ataxia. In each of the chapters, I have also given specific introductions related to the topic under investigation.

Chapter: 1

Quantitative genetic analysis of flowering time variation in wild strains of *Arabidopsis thaliana*.

Summary

Flowering is a crucial developmental event in the life cycle of flowering plant and is regulated by both internal and environmental cues. Most of the flowering plants produce flowers according to favorable seasonal changes in the day length and temperature. Light and temperature modulate several aspects of plant development. Although it is well known environmental perturbations have strong implications for plant growth and productivity, especially in the context of global warming, we know relatively little on how plants display diverse responses to temperature and changes in the light conditions. In this chapter, I provide an overview of our current knowledge and identify strains that show unusual light response measured as variation in flowering time under short day conditions. Through genetic studies, I show that there is no common genetic basis for early flowering under short day conditions. I have identified strains such as Ei-6 and En-1 that flowers early compared to the common laboratory strains Col-0 and L*er*. The earliness in Ei-6 was mapped to a *FLOWERING LOCUS M (FLM)* deletion and also our analysis revealed a potential genetic modifier for *FLM*. Through Quantitative trait locus (QTL) analysis, I show that the genetics of flowering under short days has a complex basis and multiple regions contribute to this variation. I extend this studies by including

another variable (temperature) in short days and show that *FLM* also modulates variation in temperature sensitivity. I also demonstrate variability in growth rate using a recombinant inbred line and identify a genomic region underlying this variation. Finally use a vegetative trait (leaf shape) to study the genetic architecture of leaf shape and reveal interesting transgressive variation. Testing for the linked leaf shape markers for flowering time revealed that leaf shape is independent to flowering time. QTL underlying some of these phenotypes are also described.

Introduction

Flowers are remarkable distinguishing features of the angiosperms that differentiate them from the non-flowering plant species. The successful life cycle of plant takes place in two distant phases, called vegetative phase or rosetta and reproductive phase. The vegetative phase is mainly involved in leaf production and in the reproductive phase floral meristems are initiated, which in turn give raise to different floral organs sepals, petals, stamens and carpel. Flowering is a step-by-step process decided by a combination of environmental cues and intrinsic factors. Flowering, being the first step towards seed production, is important to ensure successful reproduction of the plants.

Arabidopsis is a useful model organism to study genetic variation.

Arabidopsis thaliana, a short plant that belongs to *brassicaceae* family, first used extensively by Laibach, is mostly restricted to the northern hemisphere and it has been believed that during the glaciations this plant had colonized to several European countries and as well as naturalized in different part of the world (HOFFMANN 2002). *Arabidopsis* is a well-established genetic model, and the availability of hundreds of naturally occurring wild strains (Referred to as accessions or ecotypes or strains by different labs) and the extensive variation in phenotypes opens up opportunities to study ecological and evolutionary aspects of phenotypic variation. This also provides excellent opportunities to perform Quantitative traits locus (QTL) analysis as well as to understand G X E (genotype by environment) interactions. (ALONSO-BLANCO *et al.* 2005; AXEISSON *et al.* 2001; EL-DIN EL-ASSAL *et al.* 2001)

Four different major genetic pathways controlling flowering time

The transition to flowering is an important event in the plant life cycle and regulated by several environmental factors including photoperiod, light quality, vernalization, growth temperature, as well as biotic and abiotic stress. Extensive genetic analyses of laboratory-induced mutations as well as the analysis of natural variants have identified four different genetic pathways controlling flowering (PUTTERILL 2001; PUTTERILL *et al.* 2004). The photoperiodic pathway receives inputs from the circadian clock and light quality, increases the nuclear protein CONSTANS (CO), which in turn activates the expression of floral integrators *FLOWERING LOCUS T (FT)*. Vernalization (promotion of flowering by cold temperatures, see below) promotes flowering by enabling stable repression of *FLOWERING LOCUS C* (*FLC*), a potent suppressor of flowering. In winter annual accessions of *Arabidopsis*, *FLC* levels are high due to activation by *FRIGIDA* (*FRI*). Vernalization leads to epigenetic modification of the *FLC* chromatin in the winter annuals to ensure flowering in spring (GENDALL *et al.* 2001; LEVY *et al.* 2002; SUNG and AMASINO 2004). A series of autonomous pathway genes promote flowering in a photoperiod-independent manner function, also via suppression of *FLC* levels. Finally, the hormone gibberellic acid (GA) is essential in a fourth pathway, which controls flowering redundantly with the photoperiod pathway. All of these pathways appear to converge on a small number of integrators, the flowering-time genes *FLOWERING LOCUS T* (*FT*) and *SUPPRESSOR OF OVEREXPRESSION OF CONSTANS 1* (*SOC1*), and the floral meristem identity gene *LEAFY* (*LFY*) and *APETALA 1* (MOURADOV *et al.* 2002; SIMPSON and DEAN 2002). It is worth noting that among the environmental

parameters that modulate flowering, vernalization and photoperiod is reasonably explored at the molecular level, while the mechanisms that mediate flowering response to other variables such as temperature, stress and nutrition are not very well understood.

Flowering response to changes in ambient temperature in *Arabidopsis*.

Another variable that modulates flowering in *Arabidopsis* is ambient temperature. Studies reveal that elevated temperatures can induce flowering potently (BALASUBRAMANIAN *et al.* 2006; BLAZQUEZ *et al.* 2003; HALLIDAY *et al.* 2003; LEE *et al.* 2007). High levels of *FLC* suppressed the thermal induction, but loss of *FLC* did not ameliorate the late flowering observed under low temperature conditions in short days indicating thermal induction is not mediated via *FLC.* Studies with mutants reveal that thermal induction is independent of *CO* but mediated via the floral integrator *FT.* It also appears to be independent of the SA-dependent stress pathway or the hormonal pathway. While some of the recent studies implicate a role for floral repressors in thermal induction, and a role for chromatin remodeling in thermosensing, the genetic basis for variation in thermal response remains to be explored (BALASUBRAMANIAN *et al.* 2006; KUMAR and WIGGE 2010; LEE *et al.* 2007).

Natural variation in flowering responses of *Arabidopsis.*

Based on flowering behavior strains can be classified into rapid cycling or winter annual strains. The rapid cycling (early flowering) strains finish their life span within a short span of period (usually within 3 months). The winter annual strains are promoted to flower by winter conditions, and flower on the completion of winter. The promotion of

flowering by winter conditions is referred to as vernalization. While flowering is accelerated by vernalization in the winter annuals, rapid cycling strains show no requirement of the vernalization for early flowering. The analysis of this flowering behavior in natural populations revealed that allelic variation at two epistatically interacting genes *FRIGIDA (FRI)* and *FLC* are the major determinants for this variation (BROCK *et al.* 2009; GAZZANI *et al.* 2003; JIANG *et al.* 2009; JOHANSON *et al.* 2000; KOORNNEEF *et al.* 2004; LEMPE *et al.* 2005; MICHAELS and AMASINO 2001; SHINDO *et al.* 2005; WERNER *et al.* 2005a) In addition, the studies reveal that natural allelic variation in *FRI* can provide rapid and predictable adaptive evolution in flowering time under spring conditions(LE CORRE *et al.* 2002; SCARCELLI and KOVER 2009). The resent studies have also connected allelic variation at *FRI* with fitness effect in the natural *Arabidopsis* populations (SCARCELLI *et al.* 2007). Nevertheless, recent studies revealed substantial *FRI/FLC* independent variation in flowering responses of *Arabidopsis* wild strains (LEMPE *et al.* 2005; SHINDO *et al.* 2005; SHINDO *et al.* 2007; WERNER *et al.* 2005a). While the variation in vernalization response appears to be mostly due to *FRI/FLC*, the molecular basis for variation in the other pathways appears to be more complex.

Quantitative trait locus (QTL) mapping and subsequent cloning of the QTLs have identified a dominant allele of the blue light receptor *CRYPTOCHROME 2 (CRY2)* to be conferring an early flowering day length insensitive phenotype in the Cvi strain of *Arabidopsis thaliana.* A single amino acid substitution in the CRY2 protein confers an early flowering response under short day conditions (EL-DIN EL-ASSAL *et al.* 2001). A naturally occurring deletion in a floral repressor *FLOWERING LOCUS M (FLM)* in the

Niederzens (Nd-1, possibly Niederzeuzheim and NOT Niederzens as it is commonly referred, Balasubramanian and Weigel, personal communication) also leads to early flowering under short days(WERNER *et al.* 2005a). In addition a naturally occurring 14 bp deletion in *PHYTOCHROME D (PHYD)* in Ws strain was found to cause early flowering in short days (AUKERMAN *et al.* 1997). Apart from the cloned QTLs, mutations in phytochromes and cryptochromes are known to display flowering phenotypes dependent on the environment (KOMEDA 2004; PUTTERILL *et al.* 2004). Further more, several QTL mapping experiments on recombinant inbreed lines (RILs) have detected QTLs conferring flowering time variation in an environment dependent manner and for many of these QTLs photoreceptors have been suggested as a candidate genes (LOUDET *et al.* 2002; WEINIG *et al.* 2002; WOLYN *et al.* 2004).

In spite of a strong interest in identifying the molecular players conferring natural variation in flowering time, only a small number of the genes have been shown to confer natural variation in flowering time in *Arabidopsis thaliana*. Here I have explored quantitative genetic approaches to investigate flowering time variation in *Arabidopsis thaliana*. In this chapter, I present my results of genetic studies of the strains that flower early under short day conditions. I have taken a systemic analysis of flowering and light responses in SD conditions. I show that the early flowering strains do not share a common genetic basis for their phenotype. I show that the early flowering in the Ei-6 strains is possibly due to a previously identified *FLM* deletion. I identify QTLs for flowering time under short day conditions in two different temperatures and provide an

overview of the genetic architecture flowering behavior non-inductive photoperiodic conditions in wild strains of *Arabidopsis thaliana*

Results summary

Among strains that displayed extensive variation in flowering time there were strains with weak photoperiodic responses compared to laboratory strains Col and L*er*. I choose six strains to dissect their genetic basis of their flowering time, Jm-1, En-1, Wei-0, Fr-2, Ei-6 and Np-0 along with my studies with 3 recombinant inbred lines (Est/Col, Nd-1/ Col-5 and C24/Col-0). Analysis of their F1 populations showed that there is no common genetic basis for their earliness. The earliness appears to be largely recessive compared to Col-0. QTL analysis revealed QTL on chromosome IV and chromosome V. Previous studies had shown that deletion in *FLM* was the major flowering QTL in Nd-1 X Col-0 RIL populations in the short day conditions (WERNER *et al.* 2005b). I have identified that earliness of Ei-6 was due to deletion in the *FLM*. Linkage analysis from Ei-6 X Jm-1 F2 also mapped earliness to *FLM*. The cross between Ei-6 X Col-0 F2 resulted in a continuous distribution hinting at cryptic genetic modifiers for *FLM*. Another strain Fr-2, clustered with the photoperiodic mutants and the cross with Col-0 segregated in a bimodal manner indicating the presence of a large effect locus conferring early flowering. Using F2 populations we detected a major QTL at the bottom of chromosome 5 conferring earliness in the En-1 at short day condition. Our QTL analysis reveals some interesting aspects on the genetic basis of thermal response. First, I show that *FLM* modulates thermosensitivity in *Arabidopsis thaliana.* Second, I show the there is a environment dependent modulation of leaf initiation rate, and show that the correlation

between leaf number and flowering time may not hold true in some natural populations of *Arabidopsis.* Finally, I identify two antagonistically interacting QTL that regulate leaf shape in *Arabidopsis thaliana.* These findings demonstrate the complex interactions between growth conditions, flowering time and growth rate in *Arabidopsis.*

Materials and Methods

S.No.	Crosses
1	L*er* x Jm-1
2	L*er* x En-1
3	L*er* x Wei-0
4	L*er* x Fr-2
5	L*er* x Ei-6
6	L*er* x Np-0
7	L*er* x RLD1
8	L*er* x Ll-2
9	Col x Jm-1
10	Col x En-1
11	Col x Wei-0
12	*Col* x Fr-2
13	Col x Ei-6
14	*Col* x Np-0
15	Col x RLD1
16	Col x Ll-2
17	Jm-1 x En-1
18	Jm-1 x Wei-0
19	Jm-1 x Ei-6
20	En-1 x Wei-0
21	Fr-2 x Np-0
22	RLD1 x Ll-2

Table: 1 *Crosses set up to test the genetic basis for early flowering in short days.*

Plant material and growth conditionsAll the *Arabidopsis* strains were obtained from the European Arabidopsis stock centre (www.arabidopsis.info). The recombinant inbred lines have been described (BALASUBRAMANIAN *et al.* 2009; DESLANDES *et al.* 1998; TORJEK *et al.* 2006). For complementation studies, 30-50 F1 plants from each strain were analyzed. All the natural strains, mutants and RILs were grown in the controlled growth rooms or growth chambers (Percival Scientific, Perry, IA, USA). In order to assess whether there is a common genetic basis for early flowering, several crosses were set up including crosses with known lab strains Col-0 and Ler (Table 1). F1 and F2 flowering time analyzed at 23°SD. A SALK T-DNA line for *FLM* was isolated (*flm-3)* and verified to be a RNA-null (Minchul Kim & D.W., unpublished).

Flowering time measurements

For flowering time measurements wild *Arabidopsis thaliana* population were grown at 23°C or 27ºC as specified under short day conditions with 65% relative humidity. We measured the days to flower (DTF) as well as the total leaf number (TLN) for scoring flowering time. DTF was scored as the days when the floral bud can be visually seen. The flowering time was measured in approximately 12 plants per strain by counting total leaf number (TLN) partitioned in to rosette leaf number (RLN) and cauline leaf number (CLN). Plants were sown in completely randomized design for most of the experiments and to minimize environmental variation flats were rotated once in three-days. The leaf serration was measured on a visual scale from 1 to 10 that covered from non-serration to complete serration. To avoid individual bias, two individuals scored the phenotypes independently and the QTLs were mapped with both data.

DNA Analyses

DNA was extracted by using Cetyl Trimethyl Ammonium Bromide (CTAB) method with standard lab protocols. One leaf was collected in 2 ml eppendorf tubes and immediately frozen in liquid nitrogen. The frozen tissues was then ground using a pestle and 500µl of CTAB [1.4M Sodium Chloride (NaCl), 0.1M Tris pH 8.0, 20 mM EDTA (Ethylene Diamine Tetra Acetic Acid), 2% CTAB, 0.2% v/v β- mercapto ethanol and 1µg/µl RNAse A] was added and incubated at 65°C for 20 minutes. The tubes were vortexed and 500µl 24:1 chloroform: isoamyl alcohol was added. Tubes were mixed well and centrifuged at 14,000 rpm for 10 minutes. The supernatant was transferred in to a new tubes and 0.7V

isopropanol was added. Samples were allowed to precipitate for 10 minutes and centrifuged at 14.000 rpm for 10 minutes. The supernatant was discarded and pellet was washed with 70% ethanol and centrifuged at 10.000 rpm for approximately 5 minutes. The supernatant was discarded and the pellet was air dried for half an hour at room temperature. DNA was resuspended in 50µL and 2µL of the 1:10 dilutions was used as template to perform polymerase chain reaction (PCR). The plant material was collected in the 96 well tubes. 3mM stainless beads were added to the each well and the entire 96 well box was frozen at -80°C. Just before extraction, the box was removed and immediately subjected to crushing using mixer mill (Retsch GmbH, Haan, Germany). The rest of the DNA extraction steps were followed as mentioned above.

Statistical analyses and QTL mapping

Statistical analysis was done using the JMP package (SAS Institute), the statistical package R (www.r-project.org), and Microsoft Excel. Broad-sense heritability *(H^2)* was calculated as between-line variance *(V_G)* divided by total variance. The total variance was partitioned into between-line variance and the residuals in a one-way ANOVA model using the genotype as a single factor of random effect and the TLN as the response. QTL mapping was performed using R-qtl (BROMAN *et al.* 2003). LOD thresholds were determined using 1000 permutations. Sensitivity to temperature was assessed through regression of the sample mean on the environmental mean as previously described (LEMPE *et al.* 2005). Temperature sensitivity was calculated for each of the RILs, and the accessions. The sensitivity measures obtained for the RILs were then used in QTL mapping as a phenotype to identify a QTL for thermo sensitivity.

Results.

Genetic analysis of early flowering accessions at short days

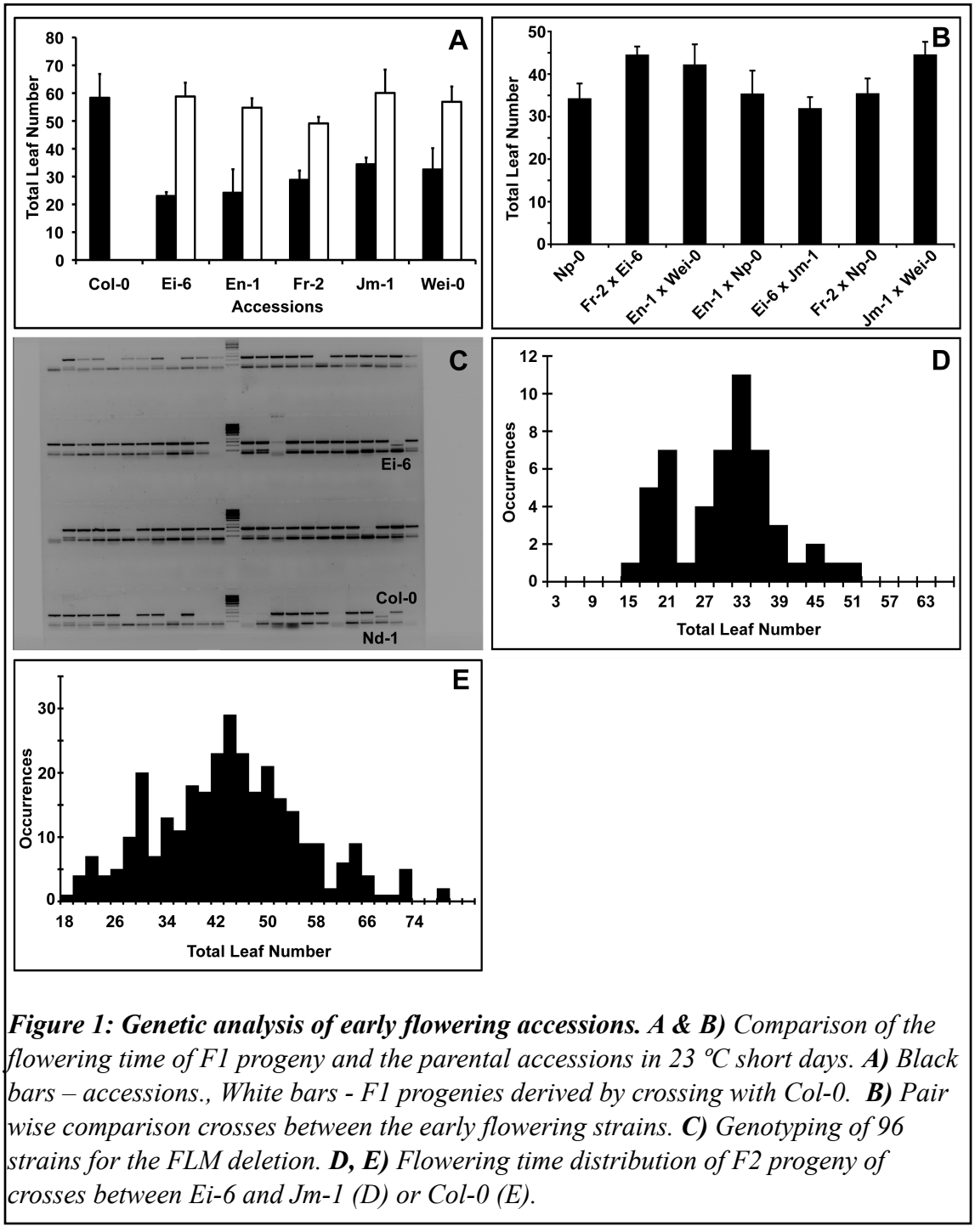

Figure 1: Genetic analysis of early flowering accessions. ***A & B)*** *Comparison of the flowering time of F1 progeny and the parental accessions in 23 ºC short days.* ***A)*** *Black bars – accessions., White bars - F1 progenies derived by crossing with Col-0.* ***B)*** *Pair wise comparison crosses between the early flowering strains.* ***C)*** *Genotyping of 96 strains for the FLM deletion.* ***D, E)*** *Flowering time distribution of F2 progeny of crosses between Ei-6 and Jm-1 (D) or Col-0 (E).*

Previously I participated in a large screen for flowering time variation in varied environmental conditions, which identified several strains to be displaying early flowering behavior under short day conditions (LEMPE *et al.* 2005). I choose 6 strains (Table 2) that flower early in short days (Fig 1A). First we asked whether they share a common genetic mechanism for their early flowering behavior. We performed pair wise complementation crosses (Table 1). Most of the F1 plants flowered similar to Col-0 suggesting that the early flowering is largely recessive (Fig 1A). In addition, crossing between two early flowering strains also resulted in late flowering F1 indicating that there is no common genetic basis for their early flowering behavior (Fig 1B). One of the known reasons for early flowering in short day conditions among natural populations is the deletion of the *FLM* locus. A naturally occurring *FLM* deletion has been found in the Nd-1 ecotype associated with early flowering at SD conditions (WERNER *et al.* 2005a). Therefore, we screened for *FLM* deletions in early flowering strains through polymerase chain reaction (PCR) genotyping in wild strains. We identified that the Ei-6, one of the strains that we identified as early flowering under short days, carries a deletion at the *FLM* locus (Fig1C). In addition F2 population obtained from Ei-6 x Jm-1 displayed 25% early flowering fraction and this phenotype was co-segregating with the *FLM* deletion indicating an association between *FLM* and early flowering in Ei-6 (Fig 1D). In contrast,

Ecotype	Country	City
Ei-6	Germany	Eifel
En-1	Germany	Enkheim
Est-1	USSR	Estland
Fr-2	Germany	Frankfurt
Jm-1	Chech Republic	Jamolice
Np-0	Germany	Nieps
Wei-0	Switzerland	Weiningen
Col-0	Lab strain	
L*er*-1	Lab strain	

Table: 2. Strains selected for analysis and their geographic origin

a cross between Ei-6 and Col-0 showed a continuous distribution, (LEMPE *et al.* 2005) revealing the presence of cryptic natural genetic modifiers for the *FLM* locus (Fig 1E).

Genetic architecture for earlier flowering under short days.

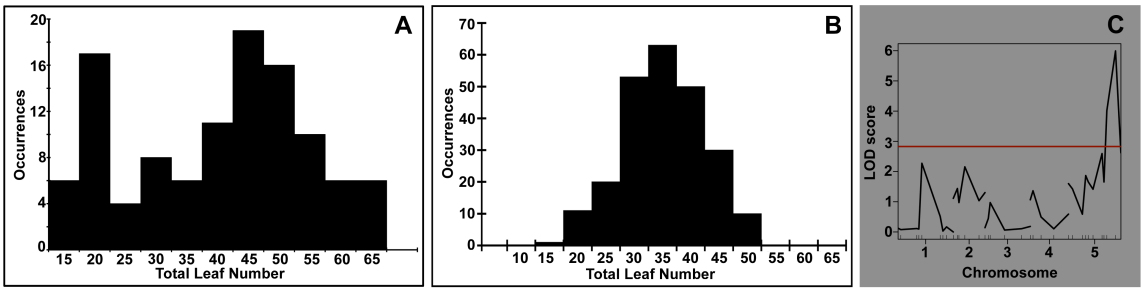

Figure 2: Genetic architecture and QTL analysis of early flowering strains. A & B) *Flowering time distribution of F2 populations derived from Fr-2 x Col-0 (A) and En-1 x Ler (B) .* ***C)*** *QTL map of flowering time in the En-1/Ler F2 population. Red line indicates LOD threshold after 1000 permutations.*

While analyzing the F2 phenotypic distributions of flowering time, (Fr-2 x Col-0 and En-1 x L*er*), we observed a bimodal distribution indicative of potential single locus of large effect in the Fr-2 x Col-0 cross and a continuous distribution in the En-1 x L*er* cross (Fig 2A and 2B). In order to map the locus responsible for earliness in En-1, we grew an F2 population consisting of 180 individuals, phenotyped for flowering time and genotyped the population with the previously developed a single nucleotide polymorphism (SNP) makers (WARTHMANN *et al.* 2007). We carried out QTL analysis, with 184 SNP markers, which identified a major QTL in the bottom of chromosome V (Fig 2C). Analysis of markers in the QTL interval suggested that the QTL is strongly linked to two SNP makers (AtMSQT_NW_241 to NW_258) and explained 14% phenotypic variance in the predicted QTL. In this QTL region there are some strong candidate genes like the *MADS-AFFECTING FLOWERING (MAF)* gene cluster that has

been shown to underlie flowering time variation in *Arabidopsis thaliana* (CAICEDO *et al.* 2009; ROSLOSKI *et al.*). We also analysed the F2 population between Fr-2 x Col-0 and the analysis of this population is presented in the next chapter. The other crosses mostly displayed continuous distribution (data not shown). These analysis suggest that the genetic basis of variation in photoperiodic flowering response, unlike the vernalisation response involves multiple genes and much more complex.

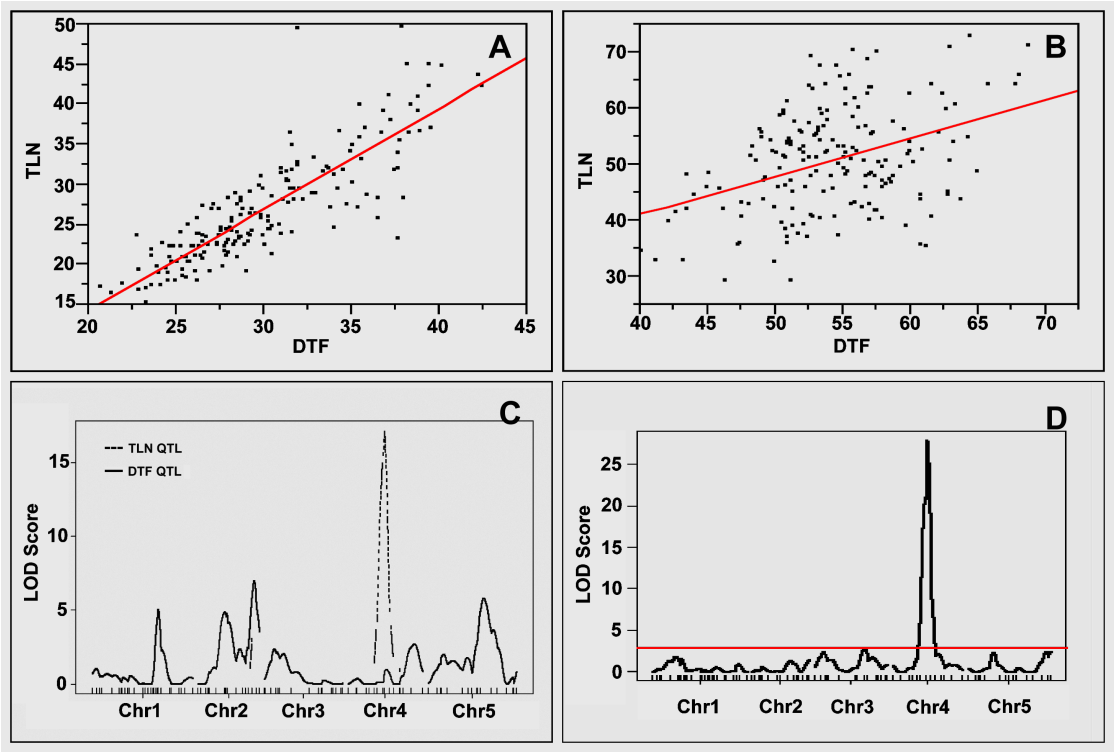

Figure 3: QTL analysis of flowering time in Est/Col-0 RILs. A & B) *Correlation between Days to flowering (DTF) and total leaf number (TLN) at 16 ºC long days (A) or 23 ºC short days (B) in the Est/Col-0 RILs. C) QTL analysis of TLN and DTF at 23ºC short days* ***D)*** *QTL analysis of Leaf Inititation Rate as measured by residuals of regressing TLN onto DTF.*

QTL analysis of flowering time variation using RILs.

Recombinant inbred lines provide an excellent opportunity to assess complex traits. Particularly, since they can be phenotyped in multiple conditions at the same time, they

also provide an opportunity to address QTL x Environment interactions. Since the analysis of the F2 populations indicated that the variation photoperiodic flowering response could be much more complex compared to the variation in vernalization response, I utilized recombinant inbred lines to assess flowering time variation. In addition, I used the RILs to assess variation under different temperature regimes and multiple phenotypes.

QTL analysis of flowering time in Est-1/Col-0 RILs

Est-1 showed a weak response to the change in photoperiod, and flowered early compared to lab strain Col-0 in short days (LEMPE *et al.* 2005). Since RILs were available for Est-1 x Col-0 lines, we took use of these recombinant lines to dissect the earliness of the Est-1. The DTF and TLN are usually correlated even under short day conditions across wild strains of *Arabidopsis thaliana* (LEMPE *et al.* 2005). However, in the Est-1 x Col-0 RILs the DTF was not correlated with TLN R^2=0.16 compared to R^2=0.72 in long days (Fig 3A and 3B) indicating that there is variability in the production of the number of leaves per day in this population under short days. The heritability of TLN (0.67) was higher than DTF (0.49) showing that this variability in leaf production is also genetically controlled. Analysis of their flowering time in different environmental conditions Est-1 and Col differ in their growth rate differ only at 23°SD (data not shown). Consistent with this, we mapped different genomic regions to be major QTLs for DTF and TLN, which in most of the other cases maps to the same region (Fig 3C). QTL analysis using the ratios (TLN/DTF) or the residuals of regressing TLN onto DTF revealed that the TLN QTL is

indeed a leaf initiation rate QTL and the LOD scores for this QTL were much higher than that of TLN (Fig 3D).

QTL analysis of flowering time in Nd-1 x Col RILs

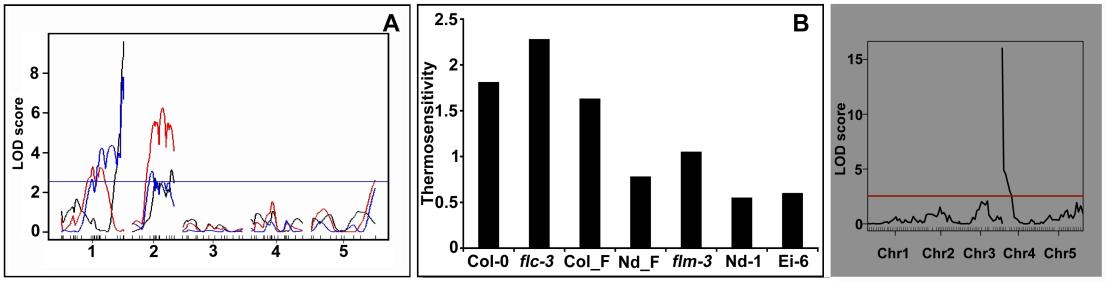

Figure 4: Effect of FLM on flowering time in 27ºC short day. A) *QTL map of Nd-1 x Col (NdC) RILs for total leaf number in 27ºC short days (red lines) and 23°C short days (black lines) and for thermo sensitivity (blue lines). The phenotype data for the 23°C map are from (Werner, 2005). A LOD threshold determined after 1000 permutations is given. Same threshold was obtained for each of the phenotypes.* ***B)*** *Thermosensitivity of different genotypes. Col_F and Nd_F represent average thermosensitivities of RILs that are homozygous for the Col (Col_F) or Nd-1 (Nd_F) allele.* ***C)*** *QTL map for flowering time in the C24/Col-0 RILs at 27ºC short days. A LOD threshold determined after 1000 permutations is given.*

Previous studies have shown higher ambient temperatures induce flowering and that even under short day conditions, early flowering could result from elevated temperatures (27ºC short days) (BALASUBRAMANIAN *et al.* 2006; BLAZQUEZ *et al.* 2003). This thermal induction was less pronounced in some of the strains including Ei-6, which carries a deletion in the *FLM* gene. Since recombinant inbred lines were available for Col-0 crossed to Nd-1, another temperature insensitive strain(BALASUBRAMANIAN *et al.* 2006), and we decided to perform QTL mapping using Nd/C RIL lines. This is particularly interesting given that previous QTL mapping experiments with the Nd-1/Col-0 RILs at lower temperatures (23ºC short days) have led to the identification of a deletion of the floral repressor *FLM* as a major cause for early flowering of Nd-1 in short days at 23°C

(WERNER *et al.* 2005a). Consistent with our expectations based on the analysis of Ei-6 strain, the effect of *FLM,* the major effect QTL in 23°C short days, was masked in 27°C short days, and the QTL was no longer detectable (Fig 4A). Consistently, there is no significant difference in the mean flowering time of plants with or without the *FLM* deletion at 27°C (Fig 4B), indicating that *FLM* modulates the sensitivity to temperature. To confirm this assumption, we calculated the thermal sensitivities of each of the RILs and asked whether a QTL for thermal response co-localizes with the *FLM* locus, which was indeed the case (Fig 4A). In contrast to *FLC*, where strains with low expression levels respond more strongly to thermal induction, lines without *FLM* respond less well to temperature compared to *FLM* wild-type strains (Fig 4B). The response to temperature is reduced, but not eliminated in the NdC lines with the *FLM* deletion, compared to lines with the wild-type allele, indicating that other factors contribute to thermal response as well. We observed a variant thermal response in *flm* mutants isolated in Ws and Col backgrounds, supporting the notion that the effect of *FLM* on thermal sensitivity depends on genetic background (not shown, (BALASUBRAMANIAN *et al.* 2006)). As expected, Ei-6 also has a reduced thermal response (Fig. 3B). However, Ei-6 flowers even faster than Nd-1 in 23°C short days. Our earlier analysis of F2 populations derived from a cross between Ei-6 and Col had pointed to a complex genetic basis of the early flowering behavior of Ei-6 (Fig 1E), consistent with the hypotheses that there are natural genetic modifiers for *FLM* effects.

QTL analysis of flowering time in C24 x Col-0 RILs

C24, which is more likely to be the same as Co-1 (Balasubramanian, personal communication) has an interesting flowering behavior and an interesting history. C24 flowers roughly at the same time as Col-0 under 23ºC short day conditions. However, when the temperature is increased, it fails to respond to the induction and this sensitivity is really low compared to Col-0. Strains with high levels of *FLC* fail to respond to thermal induction. However, there are previous reports (SANDA and AMASINO 1995), which suggested that C24 to be carrying a weak allele of *FLC*. Since, C24/Col RILs became available, we decided to assess the basis for flowering time variation in these RILs at a higher temperature. As one would predict, we found *FRI* to be the major effect QTL underlying variation in thermal response. However, in contrast to previous reports, we failed to detect a QTL at *FLC* suggesting that the C24 *FLC* may not be less active as previously suggested.

QTL analysis of leaf shape in C24 x Col-0 RILs

Interesting leaf morphological trait was observed in C24xCol RILs when (Fig 5B) grown in higher temperature. The RILs segregated for leaf serration while neither parents displayed the leaf serrations suggestive of transgression in the RILs. Mean analysis between maker and the phenotypic score of the serration revealed that, the serration QTL on chromosome II is largely promoted by the C24 allele. In contrast, the serration QTL on chromosome IV is largely promoted by Col-0 allele (Fig 4A). Therefore, in parental lines, we have one allele promoting serration and the other possibly suppressing serration.

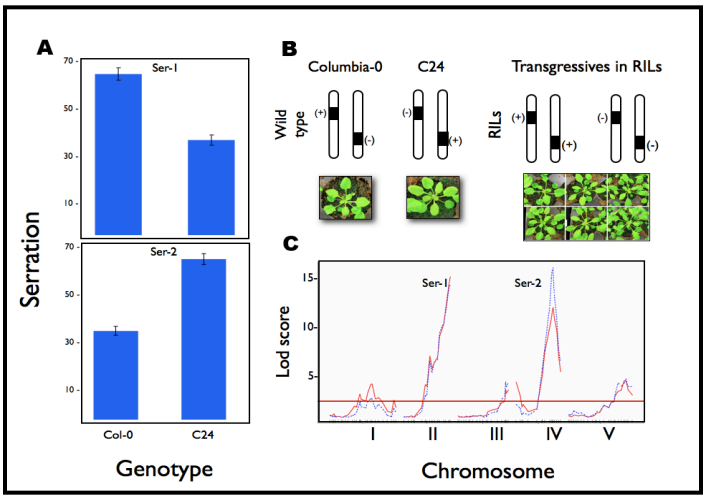

Figure 5: Genetic architecture of the leaf morphology QTLs. A) *Average phenotypic values for markers tightly linked with the Ser-1 and Ser-2 QTLs. Error bars indicate standard error.* ***B)*** *A proposed model for transgressive segregation in the C24/Col-0 RILs.* ***C)*** *Genetic architecture of the leaf morphology QTL map between C24/Col-0, there are two prominent QTLs from chromosome II and chromosome IV was responsible for serration. A LOD threshold determined after 1000 permutations is given.*

For example, in Col-0 parents, the chromosome 4 allele promotes serration and the chromosome II allele suppresses serration resulting in a smooth surface phenotype in the leaves. However in the RILs, due to the effect of two transgressive loci serrated-leaf morphology is observed (Fig 4C). Testing the same markers for flowering time generated from these lines revealed that serration is independent to flowering time (data no shown).

Discussion

Genetic basis of flowering time variation in short days is complex

Wild strains of *Arabidopsis* display extensive variation in flowering responses in multiple environmental conditions. Variation in vernalization response in *Arabidopsis* appears to be mostly modulated by *FRI/FLC,* which are the major determinants for the flowering time variation in wild *Arabidopsis* strains. *FRI/FLC* accounts for only about 17% of flowering time variation seen in SD conditions suggesting that other factors are important for this variation (LEMPE *et al.* 2005). Here, we focused on the molecular basis of

flowering time variation at SD conditions. The F1 obtained from the pair wise crosses flowering time genetic analysis of the early strains, indicates, that multiple genetic mechanisms play a role in flowering time variation in short days. Recombinant inbred lines from Est-1 and Col-0 showed a considerable variation in flowering time in short days and this variation (when measured as total leaf number) was mapped to chromosome IV and Chromosome II. The QTL analysis for flowering time in En/L*er* mapped to bottom of chromosome V, indicating multiple genetic factors controlling the flowering time. Further more the F2 analysis from crosses such as Fr-2 X Col-0 segregated for large effective loci, conferring earliness in SD conditions. Our results thus suggest that unlike the vernalization response, variation in photoperiodic response does not appear to be mediated by large effect loci (such as *FRI/FLC*). Instead, multiple loci are involved and the variation in photoperiodic response has a complex genetic basis.

Days to flowering and total leaf number may not be correlated all the time.

Flowering time is often measured in two ways; Days to flower and total leaf number. Both these traits have been shown to be highly correlated and total leaf number is often used as a proxy for flowering time measurements (ALONSO-BLANCO *et al.* 1998; KOORNNEEF 1997; KOORNNEEF *et al.* 1998). Even the wild strains display considerable correlation between flowering time and total leaf number. Consequently, when QTL analysis is done, usually the QTLs for days to flowering and total leaf number map to the same region. We have found that in the Est/Col RILs, the TLN and DTF are not correlated in an environment dependent manner. At 23ºC in short days we found the

QTLs to be mapping to different regions. Our analysis based on residuals or the derived value of leaf initiation rate demonstrated that we are looking at a growth rate QTL mapping to chromosome IV in the Est-1/Col-0 population. This notion was also supported by subsequent studies (Mendez-vigo et al. 2010). Further work showed this QTL is due to hyper activated ACD6 allele which reduces leaf initiation rate, while conferring an enhanced pathogen response. (Todesco et al. 2010). Thus, our results caution that these two traits may not be correlated all the time and modulating the growth itself could be one of the ways through which plants can vary their flowering.

FLM may be the major floral repressor in short days as opposed to *FLC*

FLM is a floral repressor similar to *FLC.* Although *FLC* is the strongest floral repressor, the flowering phenotypes of the *flc* and *flm* mutants suggest *FLM* may be more important than *FLC* under short day conditions. A mild increase in growth temperature, from 23ºC to 27°C, is equally efficient in inducing flowering of short day grown *Arabidopsis* plants as is transfer to long days (BALASUBRAMANIAN *et al.* 2006) . FLM is a floral repressor and homologue of FLC; both function in similar ways in terms of controlling the flowering time as floral repressors. However, when it comes to the activity at short days or higher temperatures they seem to differ in that the thermal induction is suppressed by *FLC* and *flc* mutants show a clear thermal induction effect*,* where as the *flm* mutants flower early in the both short and long day conditions. This indicates *FLM* modulates the sensitivity to temperature. In addition, *FLM* appears to be a major player in short days compared to *FLC.* While *FLC* accounts for only 17% of the variation observed in short days in spite of their allelic variation, *FLM* is detected as the major effect QTL in at least

two distinct strains (Ei-6 and Nd-1) and the phenotype of the *flm* mutants are much stronger than that of the *flc* mutants. Flowering time QTL analysis on NdC RIls in the short day conditions displayed a FLM as a major flowering time effecting QTL (WERNER *et al.* 2005a). This QTL was masked by higher temperature, which suggests that temperature acts in the same genetic cascade as that of *FLM* or it may act via *FLM*. Recent studies have shown that several flowering time regulators with alternatively spliced transcripts and higher temperature may modulate the splicing patterns of several genes (BALASUBRAMANIAN *et al.* 2006; QUESADA *et al.* 2005). It has been shown that *FLM* transcript undergoes alternative splicing, but the molecular basis of the alternative splicing is unknown. From our previous studies we had partially understood the abundance of one of the *FLM* splice form showed a temperature dependent differential expression. That could vastly suggestive that the gene regulation of *FLM* splice forms are controlled by temperature (BALASUBRAMANIAN *et al.* 2006).

Genetic architecture of leaf shape revealed by transgression in C24/ Col-0 recombinant inbred lines.

C24/Col-0 RILs displayed extensive leaf serration, when we were analyzing their flowering time under high temperature conditions in short days. This is a very nice example of the hidden genetic architecture controlling organ size and shape, which is revealed through transgressive variation. QTL analysis on leaf serration revealed two major loci controlling this variation. While neither parent showed serrated leaves, the RILs display extensive variation and the QTL analysis reveal both parental lines to be carrying genes that counteract each other to provide a non-serrated phenotype. In

Chromosome IV the C24 allele promotes serration and the Col-0 allele does not appear to do so. In contrast, in chromosome II, the Col-0 allele promotes serration. In addition to these loci, the QTL analysis also revealed some cryptic minor effect loci modulating leaf serration in Chromosome I and Chromosome V.

Contributions

I generated the flowering time data for various experiments presented in this chapter and analysed the data together with Suresh.

The leaf serrations were independently scored by Suresh and myself.

Suresh did genotyping of *FLM*.

The experimental strategies were compound by Suresh and Detlef Weigel.

Other contributions related to flowering time publications

I performed complementation crosses for *FLOWERING LOCUS T* in multiple *Arabidopsis* wild strains.

I have generated the Flowering time data from Est/Col QTL under 16° Long day conditions along with Janne Lempe and Suresh.

Isolated temperature sensitive and insensitive phenotypes in the *flc-3* mutants.

Chapter: 2

Allelic variation in *PHYTOCHROME C* contributes to natural variation in flowering and light responses of *Arabidopsis thaliana.*

Summary

Light affects several aspects of plant development. Light perception occurs through a set of photoreceptors. In *Arabidopsis thaliana* phytochromes and cryptochrome sense red/far red light and the blue light respectively. Studies on natural variation in flowering time and light sensitivity in *Arabidopsis thaliana* have previously identified *CRYPTOCHROME 2* and *PHYTOCHROME A* as causal genes for to be underlying varying phenotypes. In addition, a naturally occurring deletion in another phytochrome *PHYD* has also been reported. However, majority of these alleles are rare in nature and unlikely to be major players conferring adaptive response. In this chapter, I describe a detailed analysis of allelic variation in the *PHYTOCHROME C*, which is pervasive in *Arabidopsis thaliana.* I continue my characterization of the F2 population derived from a cross between Fr-2 and Col-0 described in the previous chapter and demonstrate that it carries a non-functional *PHYC* allele. Comparative Sequences analysis approach revealed allelic variation at *PHYC* locus. We show that there are two functionally distinct haplogroups for *PHYC,* which display a *FRI* dependent latitudinal cline in their distribution. Comparison with SNPs distributed throughout the genome suggested *PHYC*

to be likely under adaptive selection. In addition, *PHYC* is alternatively spliced and giving rise to two functionally different splice forms. There is natural variation in the splicing patterns of *PHYC* invoking an additional layer of complexity in the role of *PHYC* in conferring variability in phenotypes.

Introduction

In the last chapter, I discussed the genetic architecture of flowering time variation revealed under short day conditions. *FRIGIDA (FRI)* and *FLOWERING LOCUS (FLC)*, the major determinants of flowering time variation in long days accounts for only 23% of variation seen under short day conditions (LEMPE *et al.* 2005). I have also discussed the flowering time behavior of Fr-2, a strain that flowers early in short day conditions in the previous chapter and suggested that there may be a large effect locus underlying this behavior. In this chapter I explore this idea further and identify the causal locus for this phenotype. In addition to revealing the genetic basis of flowering behavior in Fr-2, which turns out to be a lesion in a photoreceptor have implications for the mechanisms of adaptation to varied environments.

Natural variation in phytochromes

There is extensive natural variation in light sensitivity of *Arabidopsis thaliana.* A dominant variant of the photoreceptor *PHYTOCHROME A(PHYA)* has been shown to underline a reduced red light response in the wild strain Lm-2 (MALOOF *et al.* 2001).

Natural Allelic variation at *PHYTOCHROME (PHYB)* locus causes the *Arabidopsis* strains respond differently to varied light conditions. (FILIAULT *et al.* 2008). QTL mapping experiments using recombinant inbred lines have detected a QTL base genetic framework for light perception and some of these QTLs co-localize with the photoreceptors (BOREVITZ *et al.* 2002; HAGENBLAD and NORDBORG 2002; WOLYN *et al.* 2004). In addition, as described in the previous chapter, a naturally occurring deletion allele of *PHYD* has also been reported (AUKERMAN *et al.* 1997). Sequence analysis of

strains reveal phytochromes in general to be highly polymorphic (CLARK *et al.* 2007b). While allelic variation in *CRY2* has been shown to be causing early flowering under short day conditions, an association with flowering has also been suggested (EL-DIN EL-ASSAL *et al.* 2001; OLSEN *et al.* 2004). However, very little is known about alleles that confer an adaptive advantage in natural populations.

Here I have explored the molecular basis of early flowering behavior in the strain Fr-2, which clustered with the photoperiodic mutants and the cross with Col-0 identified a large effective QTL in chromosome IV. I demonstrate that the molecular basis of the early flowering in Fr-2 is a non-functional *PHYC* allele. By sequencing of many strains of *Arabidopsis* strains, we have revealed two functionally distinct haplogroups of *PHYC* and allelic variation at *PHYC* is significantly associated with light responses. In addition, we show *PHYC* is alternatively spliced and the alternatively spliced form of PHYC is nonfunctional at least in the context of its ability to rescue the flowering phenotype of the *phyc* mutants. We also show that there is natural variation in the alternative splicing of *PHYC*, which adds another layer of complexity in the adaptive value for *PHYC* in natural populations of *Arabidopsis thaliana.*

Materials and Methods

Plant material, growth conditions and DNA work

Seed stocks, growth conditions, flowering time measurements and DNA extraction procedures have been described in the previous chapter. For complementation studies, 30-50 F1 plants from each strain were analyzed. The *phyc* T-DNA insertion lines were

verified through PCR and the homozygous lines were phenotyped under 23°SD conditions.

Hypocotyls length and flowering time measurements

Roughly 60-100 F2 seeds were sterilized and planted in Murashige-Schoog agar (sigma Aldrich) plates and stratified at 4°C in dark for four days and then transferred to a controlled growth chambers (Percival Scientific, Perry, IA, USA) illuminated with F17T8/TL741 bulbs (Philips Electronics, Eindhoven, Netherlands). Hypocotyls lengths were measured one week after using the software NIH image analyzer "Image J". For flowering time measurements wild *Arabidopsis thaliana* population were grown at 23°C, short day conditions and humidity was maintained at 65%. We measured the days to flower (DTF) as well as the total leaf number (TLN) as proxy for flowering time. DTF was scored as the days when the floral bud can be visually seen. The flowering time was measured more than 12 plants per strain by counting rosette leaf number (RLN) and cauline leaf number (CLN) both give rise to total leaf number (TLN). Plants were sown in a completely randomized design for most of the experiments and to avoid minimize environmental variation flats were shifted once in three-days.

Sequence analysis

For sequence analysis of *PHYC* fragments were amplified using PFU polymerase (Fermentas). Pooled products from two-four independent PCR reactions were directly sequenced on both strands (ABI sequencer). Each of the sequence files were aligned and

analyzed by SeqMan product of DNAstar, Madison, WI, and USA. Oligo nucleotide primers used for *PHYC* sequencing are given in Table 1.

Lab designation	Sequence
G-4861	CTC AGC TTC TCT CCC ACC AC
G-4862	CCC CAT AAG TGT CTG CCA GT
G-4863	CAA GTA TGG AGC AGC GTG AA
G-4864	GCA TAC CCC ATT TTC ATT GG
G-4865	TAC CGC AAG CTT CGA GAT TT
G-4866	TCG AGA GCC AAG GCT AAC AT
G-4867	CTG TGG TTT CTG GCT CCA AT
G-4868	TCC CTT TCT CAA AGG CTG AA
G-4873	GAT TGG CAG TTG AAC AAG CA
G-4874	GCA TAC CCC ATT TTC ATT GG
G-5349	CAA ATC GCA TAA ATGCAT GG
G-5350	AGT GGT GGG AGA GAA GCT GA
G-5351	CTT GTG CTC ATG AAC GGC TA
G-5352	CGT GAT GAC AAA CCA CCA AG
G-5353	CCA ATG AAA ATG GGG TAT GC
G-5354	CCT GAT GCG TCT TCT TCT CC
G-5355	GGA GAA GAA GAC GCA TCA GG
G-5356	TTC TTT CGG GAA TTT CAT CG
G-5357	GTT TGT GGC TCC CAT TTT GT
G-5358	GGA AAA GAC CGA AAC ACC AA
G-5359	GTG TCG TGA GTC GTG ACC AG
G-5360	TGG AAT CAA ACC CAA CAT CTC
G-5361	ACG CAA AGC TAC ACG GAA AC
G-5362	GAC GCC ACT GAT CCC ATA TT
G-5363	GGC TTC AGC AAA TCC TTT CA
G-5364	TCG AAC CCA GAT GAC ACA AA
G-5365	TTT TGT GTC ATC TGG GTT CG
G-5366	TGC CCG TTT AAT ACC TGC AT
G-5367	TCT CCA TCG ACG TTA AAC CA

Table. 1. Primers used for sequencing and analysis of splicing patterns. *PHYC region was amplified in four fragments using primer combinations G-5351 and G-5352; G-4863 and G-4864; G-4873 and G-4868, and G-5363 and G-5364. Other primers were used as sequencing primers and to fill in gaps in selected strains. Splicing patterns were analysed using primers G-4864 and G-4873.*

Linkage analysis

In total of 900 F2 plants, the earliest flowering 300 plants were picked and DNA was extracted. From these 300 plants a random set of 80 plants were genotyped for 18 Simple sequence length polymorphism (SSLP) markers across the genome to determine an initial linkage to chromosome V. These plants were than genotyped for 9 more markers on chromosome V to link early flowering between MSat5.22 and SO191 and to *PHYC* (Fig 1A).

Expression Studies

Microarray data was generated as a part of the AtGeneExpress project (Arabidopsis .org/info/ expression/ATGenExpress.jsp) and the data was analyzed using Gene spring (silicon genetics). The initial task was picking up the genes that are physically present between MSat 5.22 and SO191 and 2 fold difference in the

normalized expression levels between Col-0 and Fr-2 was used for picking up candidate genes. Among the 11 differentially expressed genes between Col-0 and Fr-2 we looked for genes that are specifically down regulated in Fr-2 compared to the rest of the 28 strains and found *PHYC* to be a single gene that is down regulated in Fr-2.

RNA isolation and expression studies

Total RNA was extracted from seedlings grown at 23°SD conditions. RNA extractions were performed using TRIzol reagent (Invitrogen, Karlsruhe, Germany). 1ug of total RNA was treated with RNAse free DNAse (Fermentas International Inc, Burlington, Canada) and used for the first strand cDNA synthesis as per the manufacturer's specifications (Promega). The real time PCR was performed using SYBR-Green from (Invitrogen) and detected using the Opticon Continuous Fluorescence Detection System (MJ Research, Wathman, Massachusetts, USA). All the primers used for sequencing as well as the analysis of splicing patterns are given in Table 1.

Phylogenetic analysis

Sequences were aligned using Seqman (DNA Lasergene Inc,WI , USA) and alignment was verified manually. Diversity measurements were obtained with DnaSP v4.10 (http:// www.ub.es/dnasp)(ROZAS *et al.* 2003). Sequence alignments were imported into PAUP and a heuristic search with maximum likelihood was performed using the settings for HKY model specified through Modeltest. A maximum parsimony search was performed with ACCTRAN character state optimization. In both methods, initial trees were generated through step-wise addition. TBR branch swapping option was used. 1000

bootstrap permutations were performed using the same search settings used for parsimony search with the full heuristic search option in PAUP. Parsimony and maximum likelihood resulted in similar trees. The maximum parsimony tree is shown with bootstrap values. The same split in the tree was obtained whether the coding or the non-coding regions were used.

Statistical analyses

Data were analyzed using JMP (version 5.1, and JMP 7.1, SAS Institute), Excel (Microsoft) or R (http://www.r-project.org). For association with hypocotyls length, a single factor ANOVA was performed with *PHYC* haplotype group as the factor and hypocotyls lengths as response. Previously published flowering time and hypocotyls length measurements were used for association studies (LEMPE *et al.* 2005; MALOOF *et al.* 2001). QTL analysis was done using R-QTL (http://www.rqtl.org). LOD thresholds were determined after 1000 permutations.

Results

Mapping and identification of the molecular lesion in Fr-2

Fr-2 flowers early under short day conditions and clustered together with the photoperiodic mutants in our previous analysis (LEMPE *et al.* 2005). Analysis of F2 progeny between Fr-2 and Col-0 revealed that flowering time distribution is roughly bimodal indicative of a large effect loci segregating in this population. Therefore, we grew a relatively large F2 population (900) plants in which we observed roughly 25% of the plants flower early (Chapter 1, Fig 2A). By using SSLP markers, we mapped early

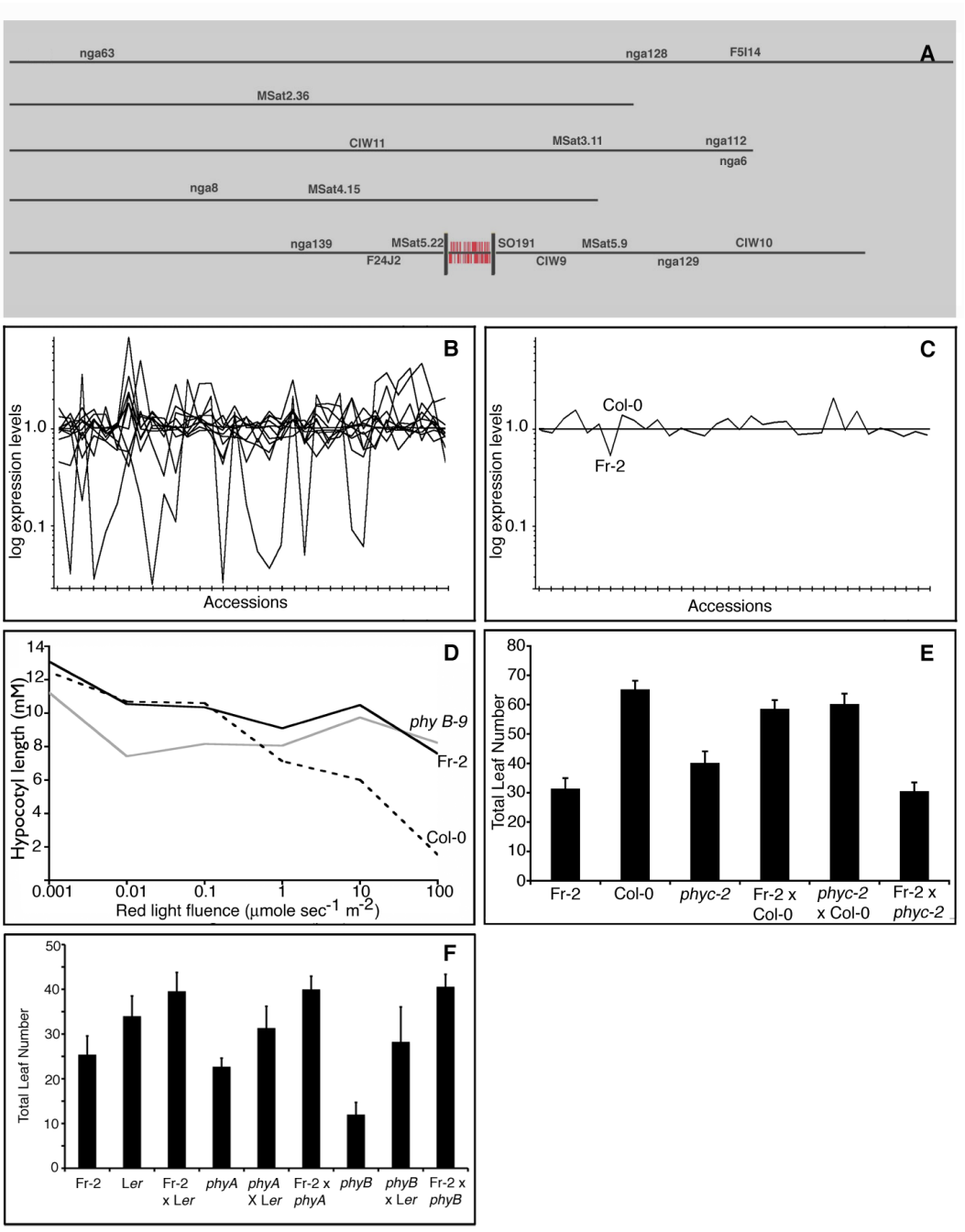

Figure 1: Identification of a defective PHYC allele in Fr-2*. **A)** Linkage mapping of early flowering in Fr-2 between markers Msat5.22 and SO191. All the markers used in the analysis are indicated. **B)** Expression profiles of 11 genes that are present in the mapping interval and are differentially expressed between Fr-2 and Col across 34 wild strains. **C)** Out of 11 genes the PHYC (At5g38540) flowering time gene is downregulated specifically in Fr-2 compared to other strains including Col-0 **D)** Red light response of Col-0, Fr-2 and phyB in Col-0 background, Fr-2 showed a similar response as phyB mutant, indicating that reduced red light sensitivity in Fr-2. **E)** Complementation test showing that Fr-2 PHYC failed to complement in phyC mutant. **F)** Flowering time data from F1 progenies from different phytochrome mutants crossed with Fr-2. Error bars indicate standard error.*

flowering between the markers MSat 5.22 and SO191 at chromosome V (Fig 1A). This is a 1Mb region containing about 300 genes within the mapped interval. We exploited the existing AtGenExpress expression profile of 34 wild strains (Fig 1B and 1C) and looked for genes that are differentially expressed between lab strain Col-0 and Fr-2 among the genes that are physically present between these markers and found 11 genes. Since early flowering phenotype was specific to Fr-2 and a recessive trait, we looked for genes that are specifically down regulated in Fr-2 compared to other accessions. We found a single gene was under expressed specifically in Fr-2 (Fig 1C).

The single gene that was specifically down regulated in Fr-2 encodes for the photoreceptor PHYC. Several observations suggested that Fr-2 could have defects at its *PHYC* locus. First, the early flowering phenotype was strongly linked to the PHYC marker. Second, *phyc* mutants particularly show an early flowering phenotype at short day conditions similar to Fr-2 and show a red light phenotype in hypocotyl length. (FRANKLIN *et al.* 2003a; MONTE *et al.* 2003). In accordance, Fr-2 also displayed red light phenotype that was similar to the *phyc* mutant flowered early under short days (Fig 1D). These observations together with our mapping efforts make *PHYC* a strong candidate gene for the early flowering response of Fr-2. As a next step we screened T-DNA insertion lines in *PHYC* and obtained the *phyc* mutants. We performed quantitative complementation experiments with Fr-2 and Col-0. While the early flowering of *phyc-2* was complemented by Col-0, Fr-2 failed to complement the early flowering of *phyc-2* mutant indicating that the molecular lesion underlying the early flowering of Fr-2 is *PHYC*. (Fig 1E)

Arabidopsis carries five discrete phytochromes A to E, that epistatically interact with each other (FRANKLIN *et al.* 2003b). In order to test the other phytochromes, we crossed Fr-2 with *phyA* and *phyB* mutants. In F1, Fr-2 alleles complemented the flowering phenotypes of *phyA* and *phyB* mutants induced in the L*er* background, indicating that *PHYA* and *PHYB* alleles of Fr-2 are functional (Fig 1F).

Fr-2 *PHYC* is highly polymorphic and shares the polymorphisms with L*er*

We sequenced the *PHYC* cDNA from Fr-2 revealed that Fr-2 carries a pre mature stop codon, in the first exon that converts K300 in to a stop codon. Theoretically, this will lead to a truncated PHYC protein, which lacks almost all the functional domains of PHYC. The GAF, PHY, PAS and the histidine kinase domains, all of which are typically required for phytochrome function will not be present in the truncated protein potentially encoded by the protein. However, it is also possible protein products may not be formed as we detected low amounts of *PHYC* mRNA in Fr-2 possibly due to nonsense mediated mRNA decay. Apart from stop codon it is highly polymorphic with 12 additional amino acid changes compared to the reference strain Col-0. However the Fr-2 *PHYC* allele is more similar to L*er* PHYC allele (COWL *et al.* 1994). This information provoked novel possibilities regarding the functional status of the L*er* allele of *PHYC*. First of all the *phyc* mutants never recovered or picked up in screens performed in the L*er* background (CHORY *et al.* 1996; CHORY *et al.* 1994; MITRA *et al.* 2004; MONTE *et al.* 2003; REED and CHORY 1994; SCHEPENS *et al.* 2004). Second, while the over expression of Col *PHYC* and L*er PHYC* differed in their phenotypes with the Col *PHYC* producing stronger

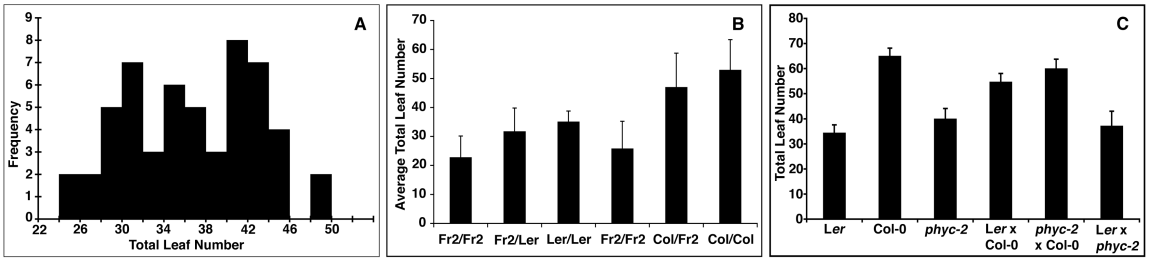

Figure 2 : Flowering behavior of populations segregating for different PHYC allele. A) *Flowering time distribution in Fr-2x Ler F2s.* ***B)*** *Average flowering time of plants with different allelic combinations at PHYC in F2 populations derived from Ler x Fr-2 and Col x Fr-2. A single copy of the Col-0 allele delays flowering much more than a single Ler allele.* ***C)*** *Genetic complementation on F1 flowering time analysis of Ler crossed with phyC mutant in Col-0 background. Error bars indicate standard error.*

phenotypes (HALLIDAY *et al.* 1997; QIN *et al.* 1997). Third, consistent with an early flowering under short day phenotype, and a co-localizing quantitative trait loci (QTL) at *PHYC* has been detected in L*er X* Col recombinant inbred lines (RIL) in fall cohorts (WEINIG *et al.* 2002). In contrast to Col-0 x Fr-2 F2 population, which displayed segregation of a large effect locus, the phenotypic distribution in Ler x Fr-2 population appeared continuous (Fig 2A). Genotyping of F2 populations derived from Fr-2 x Col or Fr-2 x L*er* revealed that the average flowering time of plants that carry Col-0 *PHYC* was much higher than that of plants with Fr-2 or L*er* alleles (Fig 2B). A single Col-0 allele delays flowering much more than a single L*er* allele (Fig 2B). Furthermore, the *PHYC* allele of L*er* failed to fully complement the *phyC* knockout in the Col-0 background. Fr-2 x L*er* F1 plants flowered early in short days, consistent with the limited activity of L*er PHYC* (Fig 2C). In addition, quantitative complementation analyses showed that phenotypic differences due to allelic variation between Fr-2 and L*er PHYC* were less than those observed between Col-0 and either Fr-2 or L*er* (Fig 3). All of this data suggests that L*er* carries a less active allele of *PHYC* compared to Col-0 and the Fr-2 allele may be less

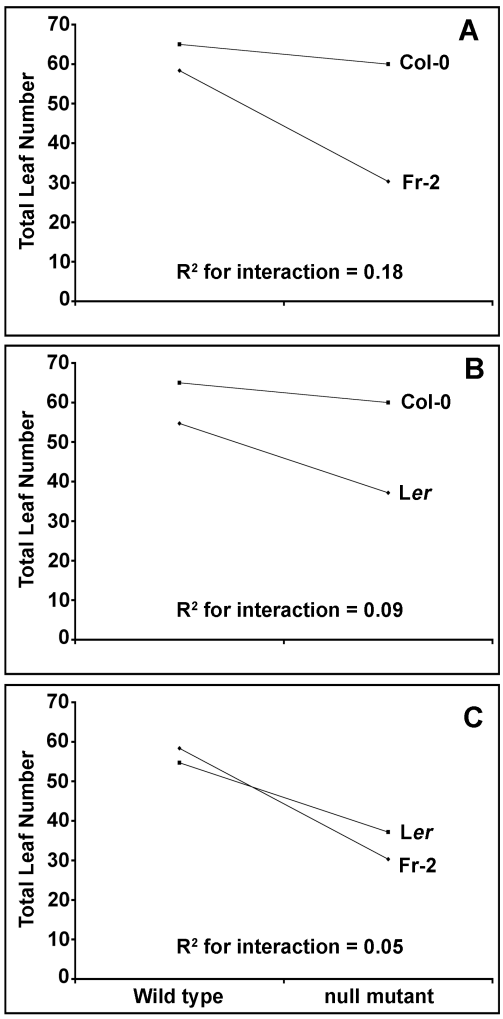

Figure 3 : Quantitative complementation analysis with different parental lines. *Col-0 was used as the wild type and the null allele phyC-2 in the Col-0 background as a tester for the crosses. An ANOVA was performed with the following model: TLN~ Line + Cross + Line x Cross. The Line x Cross interaction was significant in all three combinations (p = <0.0001). However, the proportion of total variance accounted for by the line x cross interaction varies (shown as R^2), which is consistent with effects of allelic variation between the lines. Individual pairs of comparison are depicted in different panels with Col-0/Fr-2 (A) or Col-0/Ler (B) or Ler/ Fr-2 (C)*

active even in the absence of the observed stop codon. Therefore our results suggest whether the L*er* allele itself could be a weaker allele and the Fr-2 PHYC could be on its way to becoming a pseudogene. Since the sequence analysis as well as the genetic data suggested two functionally distinct *PHYC* alleles, we studied the haplotype structure of *PHYC* in *Arabidopsis thaliana*.

Analysis of allelic variation at *PHYC* locus

We sequenced *PHYC* locus from a randomly chosen samples of 29 wild strains and with *Arabidopsis lyrata* as an out-group. An approximate 5 Kb region, consisting of the promoter region and coding region was sequenced from 26 wild strains and 3.4 Kb coding region form *A.lyrata.* Analysis of the promoter region from 26 strains revealed that both L*er* and Fr-2 contain 600bp insertion and 500 bp upstream of the ATG. Phylogenic tree was constructed using coding regions and with *A.lyrata* as an out-group. The

resulting phyogenetic tree clearly split the strains in to two classes with high bootstrapping values (Fig 4A) . First L*er* and Col fell into two different groups and consistent with our early hypothesis Fr-2 fell into the L*er* group. Between the two groups, in addition to the 600bp promoter indel there were 9 amino acids changes all of which were in complete Linkage disequilibrium (LD) (Fig 4B). Using the promoter indel as a diagnostic marker for the two haplotype groups, we determined the *PHYC* haplotype of more than 250 Eurasian strains (Table 2, Please see at the end of the chapter).

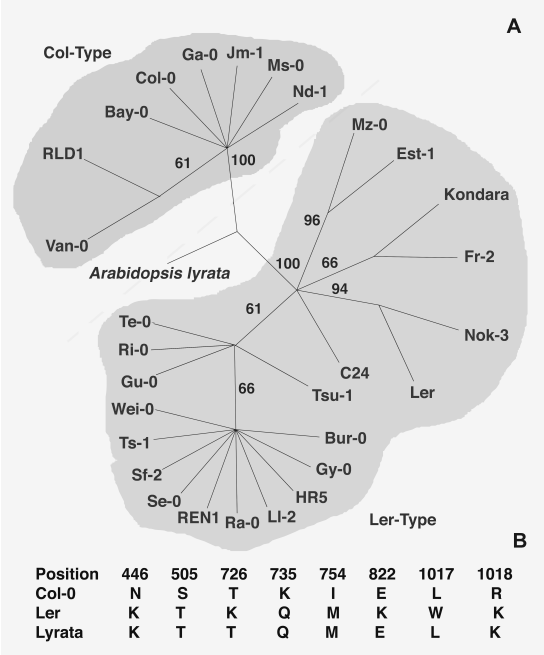

Figure 4 : PHYC haplotypes. A) *A representative unrooted phylogenetic tree generated from PHYC coding sequences is shown on top. Numbers indicate bootstrap values more than 60.* ***B)*** *The amino acid changes that delineate the two haplotype groups are given below. Unique changes compared to the outgroup A. lyrata are found in both haplotypes.*

Functional significance of allelic variation at *PHYC*

Mapping experiments have detected QTL for flowering time and hypocotyl length in many RIL populations across multiple environments. Haplotype analysis on multiple crosses has been previously used for identifying candidate genes for a QTL (WANG *et al.* 2004). In a reverse approach, we predicted that if the two *PHYC* haplotypes are distinguished by their activity, variation in flowering time and hypocotyl length should map to *PHYC* in crosses derived from parental lines with contrasting haplotypes. We examined published QTL maps from 6 different crosses for flowering time and

hypocotyl length (ALONSO-BLANCO *et al.* 1998; EL-LITHY *et al.* 2006; EL-LITHY *et al.* 2004; LOUDET *et al.* 2002; WEINIG *et al.* 2002; WERNER *et al.* 2005a; WOLYN *et al.* 2004). All populations in which the parents carried different haplotypes at *PHYC* showed

RIL set	**Trait**	***PHYC* haplotype contrast**	**QTL at *PHYC***	**Reference**
L*er* x Col-0	Flowering time in fall cohorts	Yes	Yes	(WEINIG *et al.* 2002)
Bay-0 x Shahdara	Flowering time in short days	Yes	Yes	(LOUDET *et al.* 2002)
Kas-1 x Col-0	Hypocotyl length in red light	Yes	Yes	(WOLYN *et al.* 2004)
L*er* x Cvi-0	Flowering time in short days; hypocotyl length in red light	No	No	(ALONSO-BLANCO *et al.* 1998) (BOREVITZ *et al.* 2002)
Nd-1 x Col-0	Flowering time in short days	No	No	(WERNER *et al.* 2005a)
L*er* x Shahdara	Flowering time in short days	No	No	(EL-LITHY *et al.* 2004)

Table 3. Summary of QTLs detected near PHYC.

a QTL near *PHYC*, while this was not the case for populations derived from parents that fall into the same haplotype group (Table 3). The direction of the effect of the QTL in all the crosses is consistent with the L*er*-like haplotype being less active than the Col-like haplotype. In addition, in agreement with the known function of *PHYC*, the QTL are dependent on the environment, with the light sensitivity QTL being detected under red light and the flowering time QTL in short days. We found that *PHYC* could explain 8% of the variation in white light, that it was significantly associated with hypocotyls lengths

across several conditions (Table 4). Taken together, these results strongly suggest that *PHYC* is responsible for the flowering time and hypocotyls length QTL that have been reported in this genetic region. Some of the recent QTL studies failed to detect QTLs for flowering time co-localizing with *PHYC* suggesting the presence of additional modifiers that may mask this effect (ATWELL *et al.* ; BRACHI *et al.* ; ZHAO *et al.* 2007).

Environment	**Average hypocotyl length (mm)***		***p*-value**
	L*er* haplotype group	**Col-0 haplotype group**	
White	7.11	6.09	0.001
Blue	6.42	5.78	0.02
Red	10.42	9.43	0.02
Far red	5.40	4.77	0.02
GA	8.42	7.03	0.0003
BRZ	5.62	5.08	0.05
Dark	12.45	12.00	0.21

*from ref. (Maloof et al., 2001)

Table 4 Associations of PHYC with hypocotyl lengths across range of conditions.

Population genetics and latitudinal cline

That the less active *PHYC* haplotype group is quite common suggests that variation at this locus may be adaptive in the wild. We therefore asked whether the frequency of *PHYC* haplotypes varies with latitude, since co-variation with environmental factors can be evidence of adaptation(ENDLER 1977). Latitudinal clines in light sensitivity and flowering time are known in *A. thaliana*(LEMPE *et al.* 2005; MALOOF *et al.* 2001; STINCHCOMBE *et al.* 2004). Indeed, we found that the more active Col-0 *PHYC* haplotype group was more frequent at northern latitudes ($p = 0.0001$, $n = 221$). This is particularly striking among strains that do not carry obvious lesions in *FRI* (Fig. 5A). In addition, the

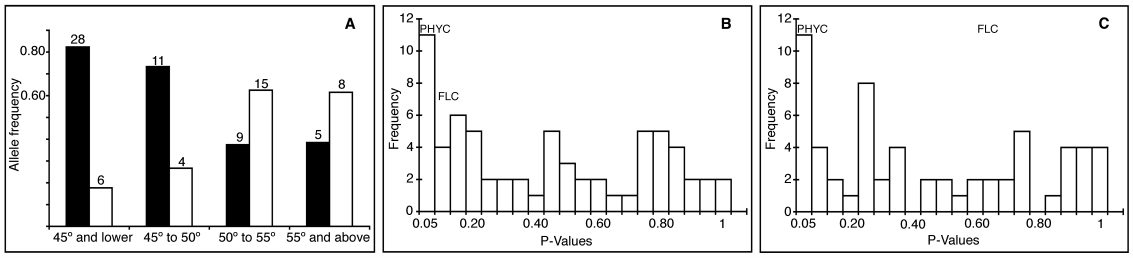

Figure 5 : Latitudinal cline of PHYC alleles. A) *Proportion of Ler-type (black) and Col-0-type (white) PHYC alleles at different latitudes among apparently FRI functional strains. The absolute numbers for each of the classes is given on top of the histograms.* ***B)*** *Distribution of p-values of a nominal logistic regression model with latitude as a factor and genotypes as response. Allele information of 65 random SNP markers with similar allele frequency as that of PHYC was available in a set of 163 strains. This information was used as a response. Note genome-wide skew towards small p-values.* ***C)*** *Distribution of p-values for interaction of a given random marker with FRI in an interaction model with latitude as the response and FRI and marker genotypes as factors with interaction.*

contribution of *PHYC* to flowering time is latitude-dependent. A *PHYC* x latitude interaction can explain 10% of the residual variation in short day flowering, after accounting for effects of *FRI,* suggesting that *PHYC* partially contributes to the late flowering of *FRI* positive strains at northern latitudes (Table 5). Since latitudinal associations could also be due to population structure, we asked how often significant interaction terms were observed in 163 strains for a set of 65 SNPs that have comparable allele frequencies and that are spaced throughout the genome (SCHMIDT *et al.* 2006). The same analysis was carried out for *FLC,* which has been reported to show a *FRI*-dependent latitudinal cline independent of population structure (CAICEDO *et al.* 2004). The *p*-values for the association of *PHYC* ranked first for latitude (Fig 5B) and fourth for the *FRI*-dependent interaction with latitude (Fig 5C). In addition, it always ranked ahead of the corresponding p-values for *FLC*. *PHYC* was also the highest-ranking marker for association with hypocotyl length after gibberellin treatment (MALOOF *et al.* 2001) (Table

Factor	DF	Sum of Squares	FValue	Probability (>F)
Latitude	1	2144.61	6.374	0.013
PHYC haplotype group	1	484.79	1.44	0.23
PHYC haplotype group x latitude interaction	1	5344.66	15.88	0.0001

Multiple R-Squared: 0.12, Adjusted R-squared: 0.10, p-value = 0.0017

Table 5 *Results of analysis of variance of latitude by PHYC haplotype group interaction on residual variation in flowering time at 23ºC in short days after accounting for FRI functionality. The significance for each of the factors is given.*

6, please see at the end of this chapter). While there is an excess of significant *p*-values among the random loci, which is consistent with the effect of population structure, the relative ranks of the *PHYC p*-values are suggestive of adaptive selection(ARANZANA *et al.* 2005) (Table 6). A similar analysis with an independent set of 69 SNP markers, even with a much smaller sample size of 56 strains yielded comparable results (Table 7).

Alternative Splicing of *PHYC*

Since all the available genetic data suggested two functionally distinct haplogroups, we cloned the *PHYC* cDNA from the L*er* and Col-0 strains and transformed the *phyc-2* mutant and analysed their flowering phenotypes. Consistent with our earlier analysis, the Col-0 *PHYC* was able to suppress the *phyc-2* early flowering phenotype much more strongly compared to the L*er PHYC* confirming the differential activity of *PHYC* alleles at least with respect to their flowering behaviour (Fig 6A). While verifying the clones, we came across two types of *PHYC* clones, with one of them retaining a partial intron. We then designed specific primers and tested for the abundance of these two respective

transcripts and found both forms of *PHYC* to be expressed at reasonable levels suggesting that *PHYC* undergoes alternative splicing (Fig 6B). Theoretically, the alternatively spliced

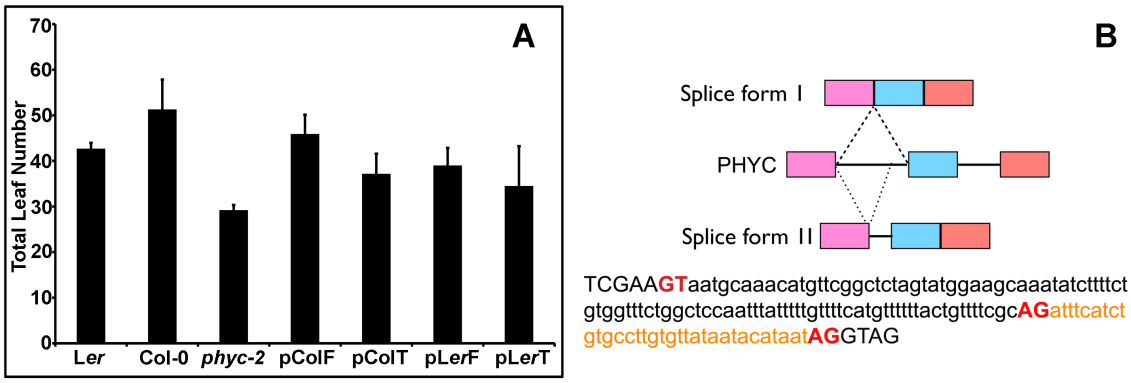

Figure 6 : Alternative splicing in PHYC locus. A) *. The schematic representation of alternative splicing at the PHYC locus. The splice donor and splice acceptor sites are presented in capitols. The red region represents the part retained in the alternatively spliced product. B) Effects of over expression of PHYC full length and truncated versions from Col-0 and Ler background. The flowering time analysis of independent T1 lines.*

form of *PHYC* would encode a protein that is likely to be non-functional since the partial intron retention leads to a premature stop codon thus resulting in a truncated protein lacking all the functional domains. To test the activity of *PHYC* splice products we have cloned full and truncated splice products individually using under 35S CaMV in the *phyc-2* mutant background. We screened the T1 phenotypes and confirmed the transgenic lines through expression analysis (data not shown). The flowering time measurements revealed that Col-0 full-length version is much strongly complemented the early flowering compared truncated version (Fig 6A). The Col-0 full and truncated versions differed much more compared to the full length and truncated version of L*er* consistent with a reduced activity of L*er PHYC.* We then compared the sequence of this intron in 96 strains of *Arabidopsis thaliana* and found the alternative splice site to be highly conserved (not shown).

Discussion

Combinatorial approach for faster detection of genes underlying quantitative traits

The F2 analysis of the Fr-2 X Col-0 segregated for large effective loci conferring earliness in SD conditions. Based on phenotypic clustering, linkage analysis, transcriptome based mapping and literature mining and subsequent cloning and sequencing we confirm *PHYC* as the underlying gene for the early flowering phenotype of Fr-2. Our study is one of the examples of using a combinatorial approach to identify the molecular lesion underlying a quantitative phenotype. Our analysis shows that once an interval is known, at least for a QTL of large effect in a pathway such as flowering time, where a genetic framework exist, the increasingly available expression profiles and expression level association can be used as an intermediate step to identify candidate loci (DYBBS *et al.* 2005).

Two functionally distinct PHYC haplotype groups

We have identified a naturally occurring putative null allele of *PHYC* that encodes a protein lacking all functional domains and shows a typical *phyc* mutant phenotype. We have shown that *Arabidopsis* contains two functionally distinct *PHYC* haplogroups with the L*er*-type being less sensitive to light as measured by the hypocotyl length. Several previous findings coupled with our analysis suggest that the L*er* allele of *PHYC* is likely to be a weak allele. Previous literature is consistent with the suggestion that L*er* carries a weak allele of *PHYC.* However, the L*er* allele might still retain some activity, which may

explain some of the phenotypes seen in heterologous over-expression studies using L*er PHYC* (HALLIDAY *et al.* 1997). Consistent with this, QTL mapping experiments have detected QTLs both for flowering time as well as light sensitivity that co-localise with the *PHYC* gene. It has been previously suggested that *PHYC* is a candidate gene for the *RED1* QTL detected in Col X Kas-1 recombinant in bred lines for long hypocotyl measured under red light conditions. (WOLYN *et al.* 2004).

Evolution of *PHYC* in *Arabidopsis thaliana*

While *PHYC* arose from a duplication event from *PHYA,* functionally it is more similar to *PHYB* (MATHEWS 2005; SHARROCK and QUAIL 1989). PHYC appears to function as a red light receptor similar to PHYB in *Arabidopsis,* but functions as a far-red light receptor similar to PHYA in rice. Consistent with this functional divergence, *PHYC* is fast evolving compared to other phytochromes and, *PHYC* has been suggested as a target gene for adaptive evolution in sorghum (ALBA *et al.* 2000; DEVOS *et al.* 2005; WHITE *et al.* 2004). It is interesting to note that *PHYC* in rice appears to functions as a far-red receptor in contrast to *Arabidopsis.* It is currently unclear exactly what changes in *PHYC* could have resulted in this altered function. Further studies are needed to elucidate the evolutionary dynamics of *PHYC* at a functional level.

Alternative splicing and its functional relevance

It has been shown that eukaryotic genome undergoes alternative splicing, but the functional relevance's are not yet understood. The recent studies revealed that 40% of *Arabidopsis* genes were undergoes alternative splicing (Filichkin et al. 2009), especially

the serine and arginine rich proteins are mainly involved in alternative splicing in *Arabidopsis* genome. (PALUSA *et al.* 2007; REDDY 2007). We have discovered the light receptor *PHYC* undergoes into alternative splicing and give s rises to two different forms. The over expression of the two individual splice forms in the *phyC* mutants, revealed that the full-length version is complemented in terms of Flowering time. The truncated version seems to have a minor effect on flowering. That could suggest that these two forms functions differently are the truncated one does not have any functional relevance in protein levels.

Photoreceptors as agents for conferring phenotypic variation

It is plausible that different environments present varied selection pressures on *PHYC* that might influence its divergence. Unlike the previously reported natural variants, *PHYC* variation appear to be more pervasive and contribute to pleiotropic effects consistent with light regulating several developmental events in the life cycle of a plant. Our analysis coupled with our work on *PHYB* and earlier works on *PHYA, PHYD* and *CRY2* suggests that natural variation at the level of light perceiving photoreceptor proteins possibly underlies differentiation to the environments in *A.thaliana*.

Contributions

I mapped and cloned the Fr-2 *PHYC* and performed the complementation assays described in this chapter.

Suresh and myself analysed flowering time data.

Mitesh , myself and Suresh performed the sequencing. Suresh and Mitesh analysed the

data and constructed the phylogeny studies.

I generated the genotypic data on *PHYC* across all the strains and Suresh carried out the analysis of clinal variability.

Oliver Bracko discovered alternative splicing at *PHYC* and Oliver and myself carried out the transgenic studies.

Todd P. Michael, Carrie Wessinger and myself generated the hypocotyl measurements, Suresh, Detlef , Joanne Chory, Julin N. Maloof and myself designed the experiments.

Table 2. *PHYC* haplotypes of wild strains

No	Accessions	Stock#	Country	Latitude	*PHYC*	*FRI**†	*FLC*†
1	Aa-0	CS900	Germany	51	Ler	Del	B
2	Ag-0	CS901	France	45	Ler	Wt	A
3	Ak-1	N939	Germany	48.6	Ler	Del	A
4	An-1	CS6603	Belgium	51.5	Ler	Del	A
5	Ang-1	N951	Belgium	50.62	Ler	Del	A
6	Bay-0	N955	Germany	49.95	Col	Del	B
7	Bch-1	N957	Germany	53.37	Col	Del	A
8	Bch3	N959	Germany	53.37	Ler	Del	B
9	Bch4	N961	Germany	53.37	Col	Del	A
10	Be-1	N967	Germany	49.5	Ler	Del	B
11	Ber-0	CS8068	Denmark	55	Ler		
12	Bl-1	CS6615	Italy	44.5	Col	Wt	B
13	Bla-1	N971	Spain	41.68	Ler	Wt	A
14	Bla-11	N985	Spain	41.68	Ler	Wt	A
15	Bla-12	N987	Spain	41.68	Col	Del	A
16	Bla-14	N989	Spain	41.68	Col	Del	
17	Bla-2	N973	Spain	41.68	Ler	Wt	A
18	Bla-3	N975	Spain	41.68	Ler	Wt	A
19	Bla-5	N6620	Spain	41.68	Col	Wt	
20	Bla-6	N6621	Spain	41.68	Ler	Wt	A
21	Blh-1	CS6645	Czech	48.83	Ler	Wt	
22	Bor-4	CS22591	Czech	49.5	Ler	Wt	
23	Br-0	N995	Czech	49.2	Ler	Wt	A
24	Bs-1	N997	Switzerland	47.55	Ler	Del	B
25	C24	CS906	Portugal	40.2	Ler	Wt	A
26	Ca-0	CS6658	Germany	50.5	Ler	Del	B
27	Can-0	N1065	Spain	28	Ler	Wt	
28	Cha-0	N1069	Switzerland	46.03	Col	Wt	
29	Chi-0	N1073	Russia	54	Col	Wt	
30	Co-1	N1085	Portugal	40.2	Ler	Wt	A
31	Co-2	N1087	Portugal	40.2	Col	Wt	
32	Co-3	N1089	Portugal	40.2	Col	Wt	A
33	Col	Lab Stock			Col	Del	A
34	Ct-1	CS6674	Italy	37.5	Ler	Del	
35	Da(1-12)	CS917	Czech	49.45	Ler	Wt	A
36	Da-0	CS6676	Germany	50	Ler	Del	B
37	Db-0	N1101	Germany	50	Col	Del	B
38	Db-1	CS6678	Germany	50	Col	Del	B
39	Db-2	N6679	Germany	50	Col	Del	B
40	Di-1	N1109	France	47.42	Col	Del	B
41	Di-2	N1111	France	47.42	Col	Del	A
42	Di-g	CS910	France	47.2	Ler	Del	A

43	Dr-0	N1115	Germany	51.05	Ler	Del	B
44	Dra-0	CS6685	Czech	49.42	Ler	Wt	A
45	Dra-1	N1119	Czech	49.42	Ler	Del	A
46	Dra-2	N1121	Czech	49.42	Col	Del	A
47	Ei-2	CS6689	Germany	50.25	Ler	Wt	A
48	Ei-5	N6691	Germany	50.25	Ler	Wt	
49	Ei-6	N1131	Germany	50.25	Ler	Del	B
50	Eil-0	N1133	Germany	51.5	Col	Del	A
51	El-0	N1135	Germany	51.5	Ler	Del	B
52	En-1	N1137	Germany	50.15	Ler	Del	B
53	En-2	N1139	Germany	50.15	Col	Del	
54	Ep-0	N1141	Germany	50.5	Ler	Del	
55	Er-0	N1143	Germany	49.58	Col	Del	A
56	Est-1	N1151	Estonia	59	Ler	Del	A
57	Et-0	N1153	France	44.63	Ler	Wt	
58	Fei-0	CS22645	Portugal	39	Ler	Del	
59	Fi-0	N1157	Germany	50.5	Col	Del	A
60	Flo-0	CS6044	Italy	43.8	Ler	Wt	A
61	Fr-2	N1169	Germany	50.2	Ler	Del	B
62	Fr-3	N1171	Germany	50.12	Ler	Del	B
63	Fr-4	N1173	Germany	50.12	Ler	Del	B
64	Fr-5	N1175	Germany	50.12	Col	Del	
65	Fr-6	N1177	Germany	50.12	Ler	Del	A
66	Fr-7	N1179	Germany	50.12	Ler	Del	B
67	Ga-0	N1181	Germany	50.3	Col	Del	A
68	Gd-1	N1185	Germany	53.55	Ler	Del	
69	Gie-0	N1193	Germany	50.5	Col	Del	B
70	Go-0	N1195	Germany	51.5	Ler	Del	
71	Go-2	N1197	Germany	51.5	Ler	Del	B
72	GOT1	N22277	Germany	51.5	Ler	Wt	
73	GOT10	N22286	Germany	51.5	Ler	Wt	
74	GOT7	CS22608	Germany	51.5	Ler	Wt	
75	Gr-1	N1199	Austria	47.06	Ler	Del	A
76	Gr-3	N1203	Austria	47.06	Col	Wt	A
77	Gr-5	N1207	Austria	47.06	Col	Wt	B
78	Gr-6	N6728	Austria	47.06	Ler	Wt	B
79	Gu-0	N1213	Germany	50.4	Ler	Del	B
80	Gu-1	N1215	Germany	50.4	Ler	Del	B
81	Gy-0	N1217	France	49	Col	Del	A
82	H55	CS932	Czech	50	Col	Del	A
83	Ha-0	N1219	Germany	52.5	Col	Del	A
84	Hau-0	N1221	Denmark	55.67	Ler	Del	
85	Hi-0	CS6736	Netherlands	52.5	Ler	Del	B
86	Hl-2	N1231	Germany	52.38	Ler	Del	B
87	Hl-3	N1233	Germany	52.38	Ler	Del	A

88	Hn-0	N1235	Germany	51.5	Ler	Del	B
89	Is-0	N1241	Germany	50.48	Ler	Del	B
90	Jl-3	CS6745	Czech	50.1	Col	Del	A
91	Jm-0	CS6748	Czech	49.04	Ler	Del	B
92	Jm-1	N1261	Czech	49.04	Col	Del	A
93	Ka-0	CS6752	Austria	46.5	Col	Del	A
94	Kas-2	CS6751	India	34	Ler	Wt	B
95	Kb-0	CS6753	Germany	50.5	Col	Del	A
96	Kl-0	N1275	Germany	50.93	Ler	Del	B
97	Kl-1	N1277	Germany	50.93	Ler	Del	B
98	Kl-5	CS1284	Germany	50.93	Ler	Del	A
99	Kn-0	N1287	Lithuania	54.9	Col	Wt	B
100	Ko-2	N1289	Denmark	55.5	Ler	Wt	
101	Kondara	N6175	Tajikistan	38.81	Ler	Wt	B
102	Kro-0	N1301	Germany	50.2	Col	Del	B
103	KZ10	N22442	Kazakhstan	50.42	Col	Wt	
104	KZ11	N22443	Kazakhstan	50.42	Col	Wt	
105	La-1	N1303	Poland	52.73	Ler	Wt	
106	Ler	Lab Stock			Ler	Del	
107	Ler-K	Lab Stock			Ler	Del	
108	Li-10	N6911	Germany	50.38	Col	Wt	B
109	Li-2	N6908	Germany	50.38	Col	Wt	
110	Li2:1	N1315	Germany	50.38	Col	Wt	A
111	Li-7	CS6878	Germany	50.38	Ler	Del	B
112	Lip-0	N1337	Poland	53.47	Col	Wt	A
113	Ll-0	N1339	Spain	41.82	Ler	Wt	A
114	Ll-1	N1341	Spain	41.82	Ler	Wt	A
115	Ll-11	N1341	Spain	41.82	Col	Wt	
116	Ll-2	N1343	Spain	41.82	Ler	Del	A
117	Lm-2	CS6784	France	48	Ler	Del	A
118	Lo-1	N1347	Germany	47.62	Col	Del	B
119	Lo-2	N1349	Germany	47.62	Ler	Del	B
120	Lov-5	CS22575	Sweden	62	Ler	Wt	
121	Lz-0	CS6788	France	46	Ler	Del	A
122	Ma-0	N1357	Germany	50.5	Ler	Del	A
123	Mh-0	N1367	Poland	53.78	Ler	Del	B
124	Mh-1	N1369	Poland	53.78	Ler	Del	B
125	Mir-0	N1379	Italy	45.65	Ler	Wt	
126	Mnz-0	N1371	Germany	50	Ler	Del	
127	Mr-0	N1373	Italy	44.5	Col	Wt	A
128	Mrk-0	N1375	Germany	49	Ler	Del	B
129	Ms-0	N1377	Russia	55.75	Col	Wt	B
130	Mz-0	N1383	Germany	50.85	Ler	Del	A
131	Na-1	CS6801	France	47.5	Ler	Del	B
132	Nc-1	N1389	France	48.6	Ler	Del	A

133	Nd-0	N1391	Germany	50.47	Col	Del	
134	Nd-1	N1680	Germany	50.47	Col	Del	A
135	No-0	N1395	Germany	51	Col	Del	A
136	Nok-0	N1399	Netherlands	52.23	Col	Wt	B
137	Nok-3	N1405	Netherlands	52.23	Ler	Wt	B
138	Np-0	N1397	Germany	52.68	Col	Del	
139	Nw-0	N1409	Germany	50.32	Ler	Del	B
140	Nw-2	N1413	Germany	50.32	Ler	Del	B
141	Nw-3	N1415	Germany	50.32	Col	Del	B
142	Ob-0	N1419	Germany	50.2	Ler	Del	B
143	Ob-1	N1421	Germany	50.2	Col	Del	
144	Ob-2	N1423	Germany	50.2	Ler	Del	B
145	Ob-3	N1425	Germany	50.2	Col	Del	
146	Old-1	N1427	Germany	53.03	Ler	Del	B
147	Old-2	N1429	Germany	53.03	Ler	Del	B
148	Or-0	N1433	Germany	50.5	Col	Wt	B
149	Ost-0	N1431	Sweden	60.5	Col	Wt	
150	Ove-0	N1435	Germany	53.33	Col	Del	B
151	Oy-0	N1437	Norway	60.5	Ler	Del	B
152	Oy-1	N6929	Norway	60.5	Col	Del	A
153	Pa-2	N1441	Italy	37.83	Ler	Wt	A
154	Pa-3	N1443	Italy	37.83	Ler	Del	B
155	Per-1	N1445	Russia	58.02	Col	Wt	B
156	Per-2	N1449	Russia	58.02	Col	Wt	
157	Per-3	N1451	Russia	58.02	Col	Wt	
158	Pf-0	N1453	Germany	48.5	Ler	Del	B
159	Pi-0	N1455	Austria	47.12	Ler	Del	B
160	Pi-2	N1457	Austria	47.12	Ler	Del	B
161	Pla-1	N1459	Spain	41.82	Ler	Wt	A
162	Pla-2	N1463	Spain	41.82	Ler	Wt	
163	Pn-0	N1469	France	48	Ler	Wt	A
164	Po-1	N1473	Germany	50.5	Col	Wt	A
165	Pr-0	N1475	Germany	50.5	Ler	Del	B
166	PUZ16	N22451	Czech	49.12	Ler	Wt	
167	Ra-0	N1481	France	46.02	Ler	Del	
168	Rak-2	CS6846	Czech	49.03	Col	Del	A
169	Rd-0	N1483	Germany	50.5	Ler	Del	B
170	REN1	N22253	France	46.02	Ler	Del	
171	REN11	N22263	France	46.02	Ler	Del	
172	RLD1	N913	Russia	56.5	Col	Del	B
173	Rou-0	N1489	France	49.5	Ler	Wt	A
174	Rsch-0	N1491	Russia	56.86	Col	Wt	B
175	Rsch-4	N1495	Russia	56.86	Col	Del	B
176	Sav-0	N1515	Czech	50	Col	Del	
177	Se-0	N1503	Spain	42.5	Ler	Wt	A

178	Sei-0	N6853	Italy	46.5	Ler	Del	B
179	Sf-1	N1513	Spain	42.05	Ler	Wt	A
180	Sf-2	N1517	Spain	42.05	Ler	Wt	
181	Sf-2e	N1675	Spain	42.05	Ler	Wt	
182	Sg-1	N1519	Germany	48.12	Ler	Del	A
183	Sg-2	N1521	Germany	48.12	Ler	Del	A
184	Shakdara	N6180	Tajikistan	37.48	Ler	Wt	B
185	Sn(5)-1	N6181	Czech	50	Col	Del	A
186	Sorbo	CS931	Tajikistan	39	Ler	Wt	B
187	Sp-0	N1531	Germany	52.5	Col	Wt	A
188	St-0	N1535	Sweden	59	Ler	Wt	A
189	Stw-0	N1539	Russia	52.96	Ler	Wt	A
190	Ta-0	N1549	Czech	49.42	Col	Wt	A
191	Tamm-2	CS22604	Finland	60	Col	Wt	
192	Te-0	N1551	Finland	60	Ler	Wt	A
193	Ts-1	N1553	Spain	41.72	Ler	Wt	
194	Ts-5	N6871	Spain	41.72	Ler	Wt	A
195	Ts-6	N1561	Spain	41.72	Ler	Wt	A
196	Ts-7	N1563	Spain	41.72	Ler	Wt	B
197	Tu-1	N1569	Italy	45	Ler	Del	B
198	Uk-1	N1575	Germany	48.03	Ler	Del	A
199	Uk-2	N1579	Germany	48.03	Ler	Del	A
200	Uk-3	N1577	Germany	48.03	Ler	Del	A
201	Uk-4	N1581	Germany	48.03	Ler	Del	A
202	ULL2-5	CS22586	Sweden	56.09	Ler	Wt	
203	Var2-6	CS22581	Sweden	55.33	Col	Wt	
204	Wa-1	N1587	Poland	52.5	Col	Wt	B
205	Wc-1	N1589	Germany	52.6	Col	Del	A
206	Wc-2	N1591	Germany	52.6	Col	Del	A
207	Wei-0	N6182	Switzerland	47.42	Ler	Del	A
208	Wei-1	N1683	Switzerland	47.42	Ler	Del	B
209	Wil-0	CS6888	Lithuania	54.68	Col		
210	Wil-1	N1595	Lithuania	54.68	Col	Wt	A
211	Wil-3	N1599	Lithuania	54.68	Col	Wt	A
212	Wl-0	N1631	Germany	48.75	Ler	Del	B
213	Ws	N6924	Ukraine	52.5	Col	Del	
214	Ws-0	CS6891	Ukraine	52.5	Ler	Wt	A
215	Ws-2	CS2360	Ukraine	52.5	Col	Del	B
216	Wt-1	N1605	Germany	52.67	Ler	Del	A
217	Wt-4	N1611	Germany	52.67	Ler	Del	A
218	Wt-5	N1613	Germany	52.67	Ler	Del	B
219	Wu-0	N1615	Germany	49.78	Ler	Del	B
220	Zu-0	N1627	Switzerland	47.42	Ler	Wt	A
221	Zu-1	N1629	Switzerland	47.42	Ler	Del	A

*"Wt" denotes no lesion known, thus putatively wild type.
†Genotype of *FRI* and *FLC* from: Lempe, J. et al. Diversity of flowering responses in wild *Arabidopsis thaliana* strains. PLoS Genet. 1, e6 (2005); Shindo, C. et al. Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of Arabidopsis. Plant Physiol. 138, 1163-73 (2005); Caicedo, A.L., Stinchcombe, J.R., Olsen, K.M., Schmitt, J. & Purugganan, M.D. Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. Proc. Natl. Acad. Sci. USA 101, 15670-15675 (2004).

Table 6. P-values obtained from for different models testing for association with latitude (Latitude), *FRI* dependent interaction with latitude (Interaction), hypocotyl length under GA treatment (GAHypo) and total leaf number in short days (TLNSD) for 67 markers across 163 strains. The rankings for each of the markers in each of the association is given (A) Model: Logistic regression model of SNP ~ Latitude. (B) Model: Latitude ~ *FRI* functionality + SNP + SNP * *FRI* functionality. The p-values for the whole model, *FRI* functionality, SNP, and for interaction with *FRI* are tabulated. Ranks are given for the interaction with *FRI.* (C) Model: Hypocotyl Length ~ SNP (D). Model: TLNSD ~ SNP. The table is sorted according to the rankings obtained for the logistic regression model for association with latitude. While the p-value ranking for flowering time (D) does not appear to be significant after background correction, the relative ranking is still consistent with a true association.

S.No	SNP_ID	Chr	Latitude	Interaction	GAHypo	TLNSD	A	B	C	D
1	PHYC	5	0.0008	0.0053	0.0007	0.0464	1	4	1	12
2	MASC05657	2	0.0015	0.003	0.5509	0.0107	2	1	45	5
3	MASC09206	1	0.0018	0.008	0.0048	0.1806	3	5	4	22
4	MASC04123	4	0.002	0.6459	0.9914	0.0015	4	41	65	2
5	MASC01545	5	0.004	0.1401	0.3519	0.4806	5	17	27	43
6	MASC03470	5	0.0043	0.6222	0.0015	0.2799	6	40	2	28
7	MASC04925	3	0.0085	0.0542	0.0232	0.6168	7	12	9	49
8	MASC02841	3	0.0157	0.9957	0.0413	0.5713	8	60	11	47
9	MASC02668	4	0.0158	0.0425	0.2932	0.0898	9	10	26	15
10	MASC04608	3	0.0208	0.0839	0.7705	0.4357	10	14	55	40
11	MASC04199	4	0.0322	0.003	0.2083	0.012	11	2	20	6
12	FLC	5	0.065	0.6914	0.165	0.4469	12	43	17	42
13	MASC04642	4	0.0691	0.7149	0.4218	0.026	13	45	34	8
14	MASC09219	3	0.0763	0.2447	0.8056	0.5349	14	25	58	46
15	MASC03128	5	0.0982	0.0138	0.0098	0.6111	15	6	6	48
16	MASC04262	3	0.1074	0.9856	0.9817	0.2048	16	59	64	24
17	MASC09216	4	0.1219	0.0405	0.5819	0.7262	17	9	47	55
18	MASC04523	3	0.1243	0.216	0.2516	0.1393	18	22	23	19
19	MASC03898	3	0.1271	0.2063	0.285	0.0656	19	19	25	14
20	MASC09214	4	0.1306		0.9575		20	65	62	65
21	MASC03344	3	0.1359	0.0484	0.216	0.4093	21	11	21	38
22	MASC03263	4	0.1501	0.039	0.4177	0.6615	22	8	33	50
23	MASC03336	4	0.1545	0.3065	0.7493	0.0025	23	31	53	3
24	MASC09224	3	0.1908				24	64	66	64
25	MASC04275	5	0.1928	0.0326	0.7308	0.9608	25	7	51	61
26	MASC03218	3	0.1978	0.3064	0.9651	0.0075	26	30	63	4
27	MASC05434	2	0.2061	0.0031	0.1286	0.1418	27	3	16	20
28	MASC05803	2	0.223	0.4262	0.7569	0.1165	28	33	54	18
29	MASC02820	4	0.2553	0.9957	0.0758	0.1925	29	61	15	23
30	MASC02675	5	0.265	0.4309	0.3951		30	34	29	67
31	MASC04350	5	0.3068	0.3032	0.5487	0.4376	31	29	44	41

32	MASC03340	1	0.3377	0.9327	0.0016	0.9389	32	57	3	59
33	MASC01582	5	0.3984	0.1158	0.4567	0.0564	33	16	38	13
34	MASC09208	5	0.4068	0.0915	0.5423	0.2976	34	15	43	29
35	MASC09222	2	0.4078	0.2145	0.6152	0.8375	35	21	48	57
36	MASC09225	4	0.4341	0.189	0.3974	0.6978	36	18	30	52
37	MASC09204	1	0.4425		0.4427	0.3189	37	62	37	32
38	MASC05029	1	0.446	0.2473	0.9011		38	26	61	63
39	MASC09209	5	0.4692	0.2837	0.4333	0.018	39	28	35	7
40	MASC04170	1	0.4773	0.4648	0.3656	0.9573	40	35	28	60
41	MASC04516	3	0.4804	0.2292	0.5042	0.0014	41	23	40	1
42	MASC04531	5	0.507	0.7203	0.0459	0.1567	42	46	12	21
43	MASC01361	5	0.5156	0.9625	0.0065	0.2725	43	58	5	27
44	MASC04983	5	0.5687				44	67	67	66
45	MASC05258	4	0.5879	0.2078	0.4156	0.0929	45	20	32	16
46	MASC05360	2	0.6034	0.8711	0.8991	0.5288	46	51	60	45
47	MASC09223	1	0.6925	0.8852	0.6706	0.3146	47	52	50	31
48	MASC03911	1	0.7062	0.6533	0.6247	0.6995	48	42	49	53
49	MASC03631	1	0.708	0.7279	0.5422	0.3653	49	47	42	34
50	MASC03754	1	0.7213	0.9171	0.199	0.4061	50	55	19	37
51	MASC09203	1	0.7375	0.5523	0.7385	0.2307	51	38	52	25
52	MASC06808	2	0.749	0.0572	0.0107	0.8759	52	13	7	58
53	MASC05386	2	0.7646	0.4901	0.0741	0.705	53	36	14	54
54	MASC03001	3	0.7705	0.2707	0.178	0.418	54	27	18	39
55	MASC03447	1	0.7952	0.927	0.0237	0.3217	55	56	10	33
56	MASC01171	3	0.7967	0.9082	0.4587	0.7354	56	54	39	56
57	MASC05962	2	0.798		0.255	0.4043	57	63	24	36
58	MASC03612	5	0.8153		0.2268	0.3117	58	66	22	30
59	MASC07090	3	0.8166	0.8059	0.5729	0.4927	59	49	46	44
60	MASC05857	2	0.8183	0.2335	0.7916	0.0456	60	24	56	11
61	MASC04591	5	0.831	0.8886	0.0133	0.9837	61	53	8	62
62	MASC02577	1	0.8586	0.7039	0.886	0.2659	62	44	59	26
63	MASC03658	1	0.8815	0.5312	0.0733	0.0407	63	37	13	10
64	MASC04209	1	0.9083	0.559	0.4395	0.3838	64	39	36	35
65	MASC09210	5	0.9267	0.852	0.7992	0.0972	65	50	57	17
66	MASC03952	5	0.9722	0.3312	0.5352	0.6953	66	32	41	51
67	MASC04819	3	0.975	0.7392	0.4083	0.0403	67	48	31	9

Table 7. P-values obtained from for two different models testing for latitudinal association with a completely independent set of markers across 56 strains. (A) Model: Latitude ~ *FRI* functionality + SNP + SNP * *FRI* functionality. The p-values for the whole model, *FRI* functionality, SNP, and for interaction with *FRI* are tabulated. Ranks are given for the interaction with *FRI.* (B) Model: Logistic regression model of SNP ~ Latitude. SNP_ID refers to the unique identifier for the SNP available at Genaissance. The table is sorted according to the rankings obtained for *FRI* dependent interaction with latitude.

S.No	SNP_ ID	Chr	Model	*FRI*	SNP	Interaction (A)	Latitude (B)	Rank A	Rank B
1	44607503	2	9E-05	9E-05	8E-04	0.00009	0.0669	1	15
2	44607971	1	9E-05	9E-05	3E-04	0.00009	0.0756	2	19
3	44607332	2	9E-05	9E-05	0.448	0.0021	0.8365	3	64
4	44606338	5	9E-05	0.0011	0.005	0.0027	0.1083	4	22
5	44606550	4	9E-05	0.0061	0.002	0.0032	0.0097	5	6
6	44606183	1	0.0001	0.0039	0.256	0.0036	0.7846	6	59
7	44607627	2	9E-05	9E-05	7E-04	0.0048	0.0675	7	16
8	44607727	2	9E-05	0.0853	0.031	0.0049	0.1595	8	28
9	44607751	4	0.0002	0.0001	0.355	0.0099	0.6177	9	48
10	PHYC	5	9E-05	0.0048	0.004	0.0137	0.0213	10	10
11	21607148	5	0.0002	0.5391	0.003	0.0255	0.0041	11	3
12	44607470	2	0.0007	0.0883	0.025	0.0336	0.0487	12	14
13	21607640	1	0.0001	0.008	0.003	0.0342	0.0221	13	11
14	44607250	5	9E-05	0.0021	0.009	0.0433	0.1002	14	21
15	44606460	5	0.0002	9E-05	0.018	0.0457	0.1407	15	26
16	44607372	1	0.0006	0.0063	0.135	0.048	0.3208	16	35
17	44606631	3	9E-05	0.0146	0.003	0.057	0.0044	17	4
18	44607685	1	9E-05	9E-05	0.001	0.0598	0.0126	18	8
19	44607364	1	0.001	0.0002	0.507	0.0639	0.1609	19	29
20	44606867	1	0.0028	0.0002	0.308	0.0902	0.762	20	58
21	44607792	4	0.0006	0.0002	0.277	0.1278	0.1403	21	25
22	21607556	2	0.0023	0.2412	0.05	0.1356	0.0403	22	12
23	44607389	3	0.0008	0.0005	0.575	0.1359	0.7544	23	57
24	44607528	1	0.0067	0.0084	0.402	0.1421	0.7198	24	56
25	44606134	1	0.0005	0.0086	0.036	0.1612	0.0713	25	18
26	44607545	4	0.0029	0.0002	0.321	0.1794	0.8851	26	66
27	21607250	3	0.0019	0.0001	0.599	0.1898	0.8422	27	65
28	44607014	3	9E-05	0.0009	6E-04	0.1948	0.0004	28	1
29	44607841	4	0.0003	0.0405	0.009	0.2167	0.0066	29	5
30	21607496	3	0.0044	0.1019	0.477	0.2335	0.3426	30	37
31	44608020	3	0.0134	0.0314	0.351	0.2398	0.2445	31	31
32	44606484	5	0.0007	0.0119	0.439	0.2566	0.0024	32	2
33	FLC	5	0.0022	0.0002	0.939	0.2993	0.4111	33	41
34	44606273	3	0.0077	0.0228	0.543	0.3061	0.6601	34	51

35	44607307	1	0.0127	0.0017	0.095	0.3287	0.6678	35	52
36	44606354	1	0.0007	0.0409	0.024	0.3377	0.0142	36	9
37	44606989	2	0.0045	0.0041	0.983	0.3435	0.8179	37	60
38	44606843	3	0.0018	0.0002	0.824	0.344	0.5148	38	45
39	44606199	2	0.0028	0.0002	0.82	0.4136	0.9311	39	68
40	44607561	4	0.001	0.0028	0.036	0.4194	0.0108	40	7
41	44607701	2	0.0015	0.0008	0.299	0.4318	0.0404	41	13
42	44606208	3	0.0087	0.0014	0.912	0.4388	0.9917	42	71
43	44607446	5	0.0037	0.0004	0.362	0.5047	0.9009	43	67
44	21607463	1	0.0087	0.0065	0.924	0.5338	0.9351	44	69
45	44607718	3	0.0035	0.0003	0.595	0.5496	0.8255	45	61
46	44607775	1	0.0043	0.0009	0.436	0.5526	0.3139	46	34
47	44607898	2	0.0035	0.0016	0.359	0.5668	0.4164	47	42
48	44606118	4	0.0106	0.0014	0.87	0.5742	0.6848	48	55
49	44607397	3	0.0021	0.0003	0.59	0.5852	0.3223	49	36
50	44608060	1	0.0039	0.0008	0.4	0.5973	0.2513	50	32
51	44607759	4	0.0041	0.001	0.565	0.618	0.6765	51	53
52	44606313	1	0.0049	0.0006	0.525	0.6253	0.56	52	46
53	44606940	2	0.0007	0.0002	0.233	0.6292	0.2321	53	30
54	44606981	1	0.0068	0.0012	0.421	0.6471	0.3441	54	38
55	21697327	4	0.0176	0.0555	0.681	0.6748	0.4299	55	44
56	44606753	1	0.0018	0.0004	0.095	0.6838	0.1227	56	23
57	44607553	4	0.0476	0.0652	0.765	0.6963	0.4169	57	43
58	44607479	2	0.0054	0.0007	0.791	0.7222	0.3566	58	39
59	44607193	5	0.0055	0.0026	0.926	0.7245	0.8339	59	63
60	44606387	3	0.0068	0.0007	0.885	0.7247	0.8307	60	62
61	44607873	1	0.004	0.0003	0.287	0.7387	0.5846	61	47
62	44606794	1	0.0054	0.0017	0.761	0.8111	0.1506	62	27
63	44606607	3	0.0014	0.0005	0.065	0.8348	0.0775	63	20
64	44606916	4	0.0055	0.0007	0.774	0.8535	0.6765	64	54
65	44607160	5	0.0037	0.0005	0.722	0.8843	0.6312	65	50
66	44607955	4	0.0047	0.0022	0.188	0.9104	0.0688	66	17
67	44606216	4	0.0072	0.0495	0.873	0.9307	0.276	67	33
68	44607299	1	0.0032	0.0064	0.355	0.9356	0.9446	68	70
69	21607631	4	0.0028	0.0003	0.348	0.9685	0.6289	69	49
70	44607405	4	0.0036	0.0047	0.214	0.9749	0.1342	70	24
71	44606102	3	0.0059	0.0587	0.98	0.9833	0.3713	71	40

Chapter: 3

A genetic defect caused by a triplet repeat expansion in *Arabidopsis thaliana*

Summary

Trinucleotide repeats are simple repetitive sequences of three nucleotides or short tandem repeats, also known as microsatellites present in across organisms. Functions of these repeats are largely unknown, but expanded repeat alleles have functional consequences to organism and leads to pathology. More than 40 different neurological diseases in humans are due to expansion in trinucleotide repeats. These hereditary diseases also display a phenomenon known as genetic anticipation, in which the nature of the illness passes through successive generations with a progressive increase in severity along with an early onset of the disease. So far these diseases have been described only in humans although repeat variability associated phenotypic variation has been described in microbes. Here, we report a triplet repeat associated genetic defect in an inbred strain of *Arabidopsis thaliana*. Bur-0 strain, from Burren, Ireland carries a dramatically expanded TTC/GAA repeat in the intron of the *ISOPROPYLMALATE ISOMERASE LARGE SUBUNIT 1 (IIL1)* gene. Our findings reveal striking parallels to the human genetic disease Friedreich ataxia (FRDA), a neuronal disease caused by a GAA/TTC triplet expansion at intronic region in *FRATAXIN (FXN)* gene. Spontaneous and induced reductions in the repeat length are sufficient for a reversal of the detrimental phenotypes in Bur-0. Whole genome wide distribution of triplets in the reference strain Col-0 and sequence of the TTC/GAA repeat of *IIL1* from 96 wild *Arabidopsis* strains revealed that the expansion is a rare

signature in the Bur-0 strain. Dynamics of the triplet repeat mutation in different environmental conditions suggests that repeat variability is indeed dependent on the environment. Further analysis of the genome has identified genes that carry varying lengths of repeats in their exons resulting in variability in poly amino acid stretches. These genes provide further candidates for analysing the relationship between repeat variability and their potential role in mediating phenotypic variation.

Introduction

Expanded trinucleotide repeats involved in human diseases

It is now known that expanded trinucleotide repeats are causal for more than forty different neuronal degenerative diseases (PEARSON *et al.* 2005). Trinucleotide repeat sequences fall in to a class of repetitive sequences known as microsatellites, which are short repetitive units of 1-6 nucleotides. Although the functional relevancies are not yet understood, studies reveal that these sequences were found in the genomes of all organisms (ARMOUR 2006). The repeat expansions can occur at all parts of the gene (Fig 1). Based on their occurrence in the gene the repeat expansion disorders are divided into two major classes namely coding region disorders and non-coding region diseases. Triplets in coding regions typically lead to poly-glutamine (CAG) and poly-alanine (GCU GCC GCA GCG) (GATCHEL and ZOGHBI 2005) tracts and in non-coding regions triplet expansion can modulate gene expression (e,g., Friedreich ataxia) (CHI and LAM 2005; POLLARD *et al.* 2004; SINDEN *et al.* 2002)

The intronic expansions are typically much larger than exonic expansions, probably due to tolerance in the intronic regions. Often the genetics of these diseases and genetic anticipation phenomenon does not fit with the simple Mendel's hypothesis (one gene one mutation) (FONDON *et al.* 2008) as the repeat length varies and the onset of the disease get worsened in the subsequent generations. The early onset of diseases appears around the thirties and individuals exhibit-increased severity in diseases with age that is associated with the progressive nature of the diseases. The expanded trinucleotide mutations are fall into a novel type of mutations also known as dynamic mutations and

the mechanism that underlies this instability is not clear (GOMES-PEREIRA *et al.* 2001).

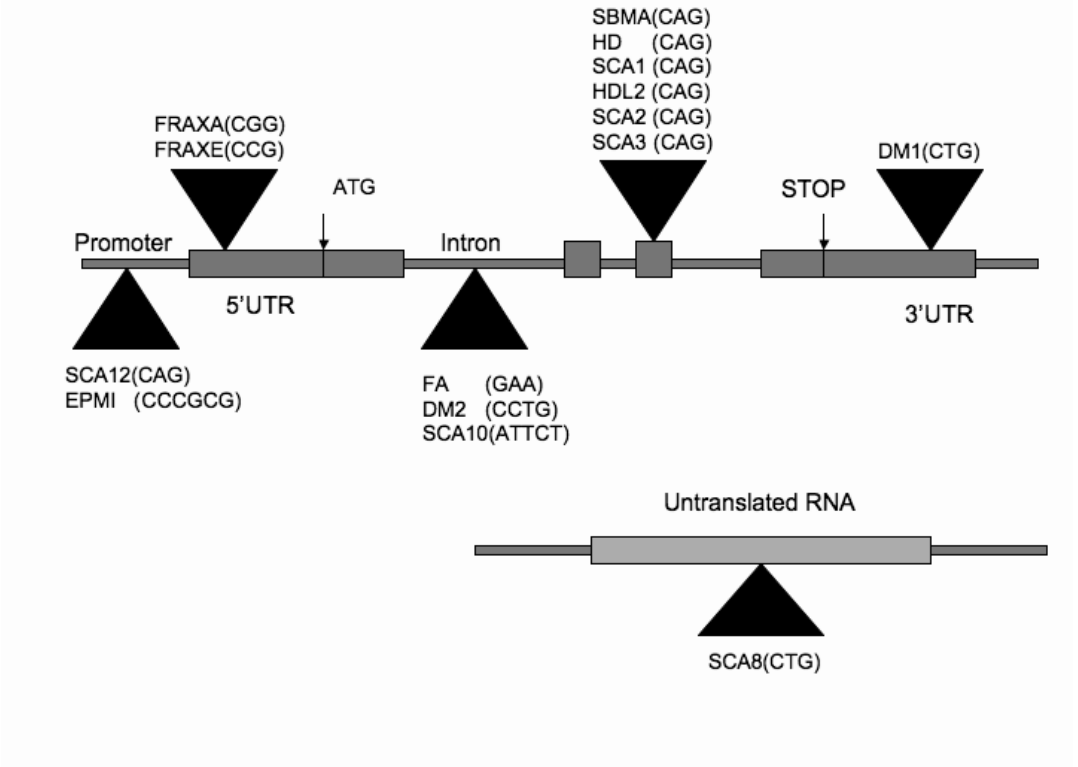

Figure. 1. A few known examples for repeat expansion disorders in humans. Physical position of the expansions with respect to the gene structure are shown . Diseases and their corresponding repeats that undergo expansion are depicted. FRAX – Fragile X syndrome, SCA – Spino cerebellar ataxia, HD – Huntington's disease, HDL – Huntington's disease like, FA – Fredereich's ataxia, DM - Dystrophica myotonica, SBMA – Spinal and bulbar muscular atrophy, EPM1 – Epilepsy, progressive myoclonic 1.

This is primarily due to the absence of appropriate models as the triplet expansion disorders have been described only in humans until recently and the efforts to engineer long stable repeat containing alleles have been had limited success in the model organisms (BIDICHANDANI *et al.* 1998; GOMES-PEREIRA *et al.* 2007; RICHARD and DUJON 1996; RICHARDS 2001). While it appears to be feasible to generate transgenic models harboring smaller expansions in the coding regions, it has proven to be challenge to

engineer large intronic expansions (KRASILNIKOVA *et al.* 2007; MIRKIN and MIRKIN 2007; MIRKIN 2007; ORR and ZOGHBI 2007).

Expansions are suggested to occur largely due to DNA polymerase slippage and unequal cross over during the replication. Expanded repeats adopt a unusual DNA structures like triplex forms or sticky DNA and interfere with replication and transcription (BIDICHANDANI *et al.* 1998; BROWN and BROWN 2004; KRASILNIKOVA *et al.* 2007 ; WELLS *et al.* 2005). It has also been demonstrated that the long repeats has negative implications on the genome structural change (GATCHEL and ZOGHBI 2005; VOLKER *et al.* 2009).

Our present knowledge of triplet repeat associated disorders have been limited to humans. While there are 10 families are triplets exists, so far only the CAG/CTG, CGG/ CCG, and GAA/TTC repeats are known to cause diseases.. Depending on the location and the nature of expansion, the mutation might cause lose of function of the gene, or gain of function of the gene, as a result of toxic protein accumulation in the cells and subsequently lead to pathology.

Non-coding trinucleotide repeats expansion disorders

Non-coding trinucleotide repeats were tending to be more dynamic compared to coding repeats as introns are more tolerated for long expanded repeats, compare to exons (BIDICHANDANI *et al.* 1998) (OHSHIMA *et al.* 1996). The longer the expansions the difficult it becomes to engineer them. However studies with partially expanded repeats reveal that long intronic repeats interfere with replication and transcription. One of the triplet repeat diseases humans identified in early 90s is Friedreich ataxia (FRDA). Here

normal healthy individual carries 30-40 repeat GAA repeats in the first intron of the *FXN* gene and the patients carry more than 1000 repeats. The resultant hyper expanded alleles has implications on less production of mature frataxin mRNA and Frataxin protein (ALPER and NARAYANAN 2003; BERCIANO *et al.* 1997; BIDICHANDANI *et al.* 1997; BIDICHANDANI *et al.* 1998; BIDICHANDANI *et al.* 2000)

The long intronic repeats acquire peculiar structural properties (HEIDENFELDER *et al.* 2003; HEIDENFELDER and TOPAL 2003). The expanded GAA/TTC repeats form hairpins loops and acquire a tendency to form non-Watson and Crick base pairs, duplexes of which form what is called as sticky DNA (SAKAMOTO *et al.* 1999; VOLKER *et al.* 2009). The sticky DNA interferes with replication and induces errors in the replication and could also lead to double strand break repair and unequal recombination (WELLS *et al.* 2005). Long GAA/TTC repeats have been shown to interfere with transcription as well in a length- and orientation dependent manner, and GAA/TTC expansions have a more severe phenotypes than TTC/GAA expansions (GRABCZYK *et al.* 2007; GRABCZYK and USDIN 2000; OHSHIMA *et al.* 1998; OHSHIMA and WELLS 1997; WELLS *et al.* 2005). GAA/TTC interferes with replication and experiments using yeast as model have shown that mutations in the replication enzymes for example *Rad27* homology to yeast *FEN*-1 and Rad6 and Rad5 and DNA helicases Sg1 lead to significance increase in the trinucleotide repeats in the wild type background and decreased in the knock outs (LEE and PARK 2002; LIU *et al.* 2006; SHISHKIN *et al.* 2009; SINGH *et al.* 2007).

The comparative genome sequence analysis revealed that the distribution of GAA/TTC triplets is more in the human genome (BIDICHANDANI *et al.* 1998). Compared

with other organisms the mammalian genomes carry a abundance of GAA/TTC tracts (CLARK *et al.* 2006; CLARK *et al.* 2004), which indicates that they were evolving in selective constrains in linage specific ways and also suggesting that the long repeats might have been under selective pressure during evolutionary trajectories and also source of inter individual variability.

Expansions of other triplet repeats can also lead to disease phenotypes. For example In Fragile X syndrome, the expansion of CGG repeats in 5'UTR of the fragile X gene FMR1 lead to the disease condition. Another example is FRAXE mental retardation caused by GCC repeat in 5'UTR, were the normal carriers 6-25 repeat the effected once carries >170 repeats. Myotonic dystrophy (MD) is another example for autosomal dominant trinucleotide repeats expansion diseases caused by massive expansion of CTG repeats located at 3'UTR region protein kinase(DMPK) gene. All these repeats share similar properties such as including the ability to initiate repeat mediated epigenetic changes resulting in a trigger for heterochromatic formation (KUMARI and USDIN 2009) .

Coding expansions disorders

Coding expansion disorders typically have poly glutamine (polyQ) and poly alanine (PolyA) tracts in the proteins. So far more than 10-12 different repeat disorders were reported due to unstable CAG repeats, which encodes for the amino acid glutamine. Huntington's Disease (HD), an autosomal dominant inherited disorder carries a long tract of poly glutamine repeats at human huntingtin (HTT) gene. Normal individuals carry 6-35, disease-borns carry >65 times and this perturbation leads to accumulation toxic

protein in myelin sheets in brain(DUYAO *et al.* 1993; GUSELLA *et al.* 1993; MYERS *et al.* 1993).

Since triplet expansion associated disorders have so far described in human, attempts have been made to genetically engineer repeat expansions, The first transgenic model for human HD is *Drosophila melanogaster*, shown that the transgenic flies could withstand longer glutamine repeats in their genome (JACKSON *et al.* 2005). This model has been recently shown to display similar genomic instability like in humans validating this approach (JUNG and BONINI 2007). Some of the other transgenic models also contributed to the transgenic studies of human *HTT* gene (PHILLIPS *et al.* 2009; SIPIONE and CATTANEO 2001). And mouse model to study the behavioral pattern of the phenotypes and higher-primate (rhesus macaque) transgenic model for HD showed a similar phenotypic effects as humans, in subsequent generations, even though most cases higher-primate progenies were unable to tolerate the long chain glutamine in HTT gene (LIM *et al.* 2008 ; YANG *et al.* 2008). .In spite of their contributions transgenic models have been of limited success.

The molecular mechanisms of the expanded mutations are poorly understood. In humans there is often extensive somatic variability in the degree of expansion of GAA/TCC repeats(BIDICHANDANI *et al.* 1999; CLARK *et al.* 2007a; PEARSON *et al.* 2005; RICHARDS 2001) Recent studies have shown that age dependent GAA triplet expansion in the brain tissues of the transgenic mouse model (CLARK *et al.* 2007a). While it is not clear the somatic mutation of triplet is maintained same as germ line mutations, at least according

to our present knowledge somatic mutations are more vulnerable for expansions at developmental stages.

So far triplet repeat expansion associated diseases were reported only in humans, here I have discovered a natural strain of *Arabidopsis thaliana* to be containing a triplet repeat expansion associated with a deleterious growth arrest phenotype. We show that this example in *Arabidopsis* shows striking parallels with the human situation, both with phenotypic and reduction in the gene expression and thus provides an excellent experimentally amenable genetic model to study fundamental aspects of triplet expansions associated genetic diseases.

Materials and methods

Plant material and growth

T-DNA insertion lines N606966, N630604, N500855, N630533, N601014, N571128, N606964, N529510 and all other seeds were obtained from European Arabidopsis stock centre. Plants were grown in controlled growth rooms or Percival chambers under a mixture of Cool white and Gro-Lux wide spectrum fluorescent lights. Temperature was maintained at 27°C and 23°C according to the genotypes, the humidity was maintained around 65%. Seeds were stratified in 0.1% agarose for 3 to 4 days at 4°C before sowing into soil.

Histology

For light microcopy, the leaves were collected from Bur-0 and Pf-0 grown from 27°C. The leaves were collected after *ill1* phenotype visually screened. The collected leaf

materials were stored in ethanol and fixed in formaldehyde acetic acid (FAA) and embedded in paraffin. Sections were 8 to 12 μthicknesses were stained with 0.1% toludine blue (Fluka, Buchs, Switzerland) and destained with diluted alcohol EtOH series.

Mapping and sequence analysis

Six hundred Bur-0 X Pf-0 F2 plants were used for mapping the *ill1* phenotype. For initial linkage we tested 10 SSLP (simple sequence length polymorphism) markers across the chromosomes. For fine mapping we used 10 more SSLP on Chromosome IV and narrow down to 16.5 Kb interval .For further mapping, 22 fragments designed between 300-500 base pairs were sequenced in recombinants and parental lines and marker heterozygosity was determined by visual inspection of sequence traces using SeqMan product of (DNAstar, Madison, WI, USA). The repeat lengths were analyzed by PCR and Pooled products from independent reactions were sequenced directly on both strands using ABI sequencer. PCR reactions were performed using proof editing enzyme PFU Taq from (Promega, Germany)

Transgenes

Bur-0 (At4g13430) cDNA was cloned using seedlings grown in 27°C, 1.84 Kb of full length cDNA was amplified was using PFU taq polymerase (MBI Fermentas). Sequence verified constructs were inserted in to CaMV 35 S promoter in a modified version of pGREEN vector (pFK210). Final construct pSS03 was transferred in to *Agrobacterium,* to induce flowering the Bur-0 plants were vernalized at 4°C for four weeks.

EMS Mutagenesis

Marco Todesco gave EMS mutagenised seeds of Bur-0 M2 seeds. For suppressor screen I screened for 30,000 M2 plants, out of 30,000 we found 156 plants lost the *ill1* phenotype. Repeat analysis of theses suppression plants we found 27 lines lost the repeat and 28 repeat maintained homozygous lines were used for second site mutations screen. The entire phenotypic assays were performed under 27°SD conditions both plant rooms and Percival.

Gene expression studies

Bur-0, Pf-0 and EMS suppressor had been grown under 27°C, shoot apices were harvested under stereomicroscopes and flash-frozen in liquid nitrogen. RNA was extracted from two biological replicates using Trizol reagent (Invitrogen Corporation, Carlsbad, CA, USA) method. Extracted RNA was quantified and hybridized into Affymetrix ATH1 arrays. Global expression profiles were analyzed using the gcRMA (IRIZARRY *et al.* 2003a; IRIZARRY *et al.* 2003b) algorithm and visualized using Gene Spring (Agilent Technologies, Santa Clara, CA, USA). Differentially expressed genes were identified by pair-wise analysis using Logit-T ($p<0.05$) (LEMON *et al.* 2003)

RNA isolation and cDNA synthesis

For RNA isolation and cDNA synthesis please refer chapter 1.

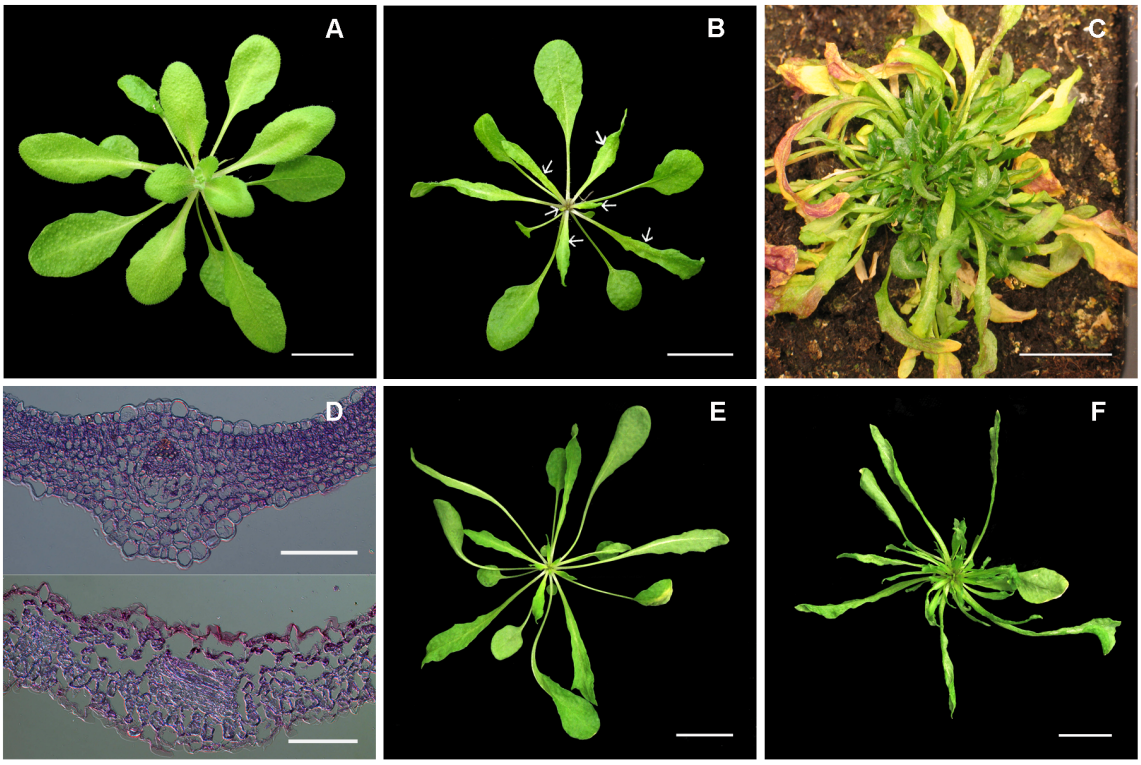

Figure 1: Cryptic Bur-o phenotype at higher temperature. *A) Four-week old Pf-0 plant.* ***B)*** *Bur-0 plant. Arrows indicate the progressively more worm-like habitus of leaves.* ***C)*** *10-week old Bur-0 plant, in which phenotype has become even more apparent.* ***D)*** *Cross section of expanded leaves of Pf-0 (top) and Bur-0. Note disorganized cellular architecture in Bur-0.* ***E and F)*** *Effect of light intensity and irradiance on the Bur-0 phenotype, 14-18 klux/41-50 W/m^2 (E) and 20-22 klux/55-6 W/m^2 (F). Normal looking leaves (arrows) are only seen under the lower intensity. Scale bars= 1.5 cm except D, which is 200*μ*.*

Genome wide analysis of triplet repeats length

We downloaded the whole genome sequence of the *Arabidopsis thaliana* Col-0 reference strain. Ramya Harilal, Korbinian Shneeberger and David Richter wrote a Perl script and pulled the triplet from complementary and reverse strands. The different combinations of triplets were classified as 10 groups and each group carried a six consecutive trinucleotide repeats (e.g. CCA, CAC, ACC, TGG, GTG, GGT). Mononucleotide triplets were combined in a single triplet group. The analyzed repeat data and TAIR sequence data was stored in MYSQL database.

Results

Discovery of Bur-0 phenotype

When we were analyzing multiple strains at higher temperatures for their thermal

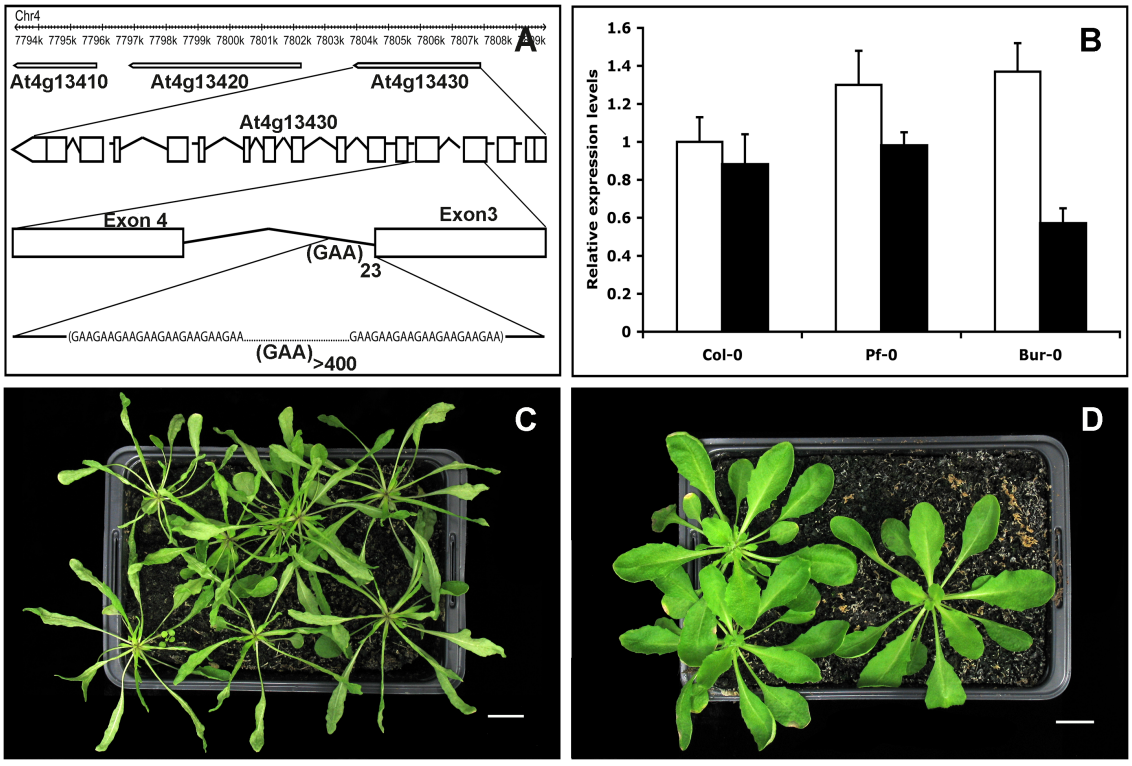

Figure 2: Identification of IIL1. A) *Final mapping interval in a Bur-0 x Pf-0 F_2 population. Entire mapping interval between the recombination breakpoints is shown. Only part of At4g13410 lies within the mapping interval. Pf-0 and the Col-0 reference strain both contain 23 repeats, while Bur-0 has more than 400 repeats.* ***B)*** *At4g13430 expression in leaves of four-week old plants grown under 23ºC (white bars) and 27ºC short days (black bars). Real time RT-PCR analysis was carried out on 10 biological replicates with two technical replicates each.* ***C)*** *Five-week old Bur-0 plant.* ***D)*** *Bur-0 plant carrying a 35S:At4g13430-Bur-0 transgene.*

response, we realized that several cryptic phenotypes are revealed under these conditions. Previous studies (BALASUBRAMANIAN *et al.* 2006) had shown mild increase in temperature to 23°C to 27°C, while most strains flowered rapidly, a strain from Burren, Ireland (Bur-0) did not flower and growth arrest was noticed compare to Col-0 reference

strain grown at 27°C (Fig 2A). In contrast to Col-0, Bur-0 plants produced blade like leaves (Fig 2B and C). This was generally observed after seventh true leaf onwards. The leaf expansion was severely reduced and cellular organization of Bur-0 is disrupted compared to the reference strains Col-0 or Pf-0 (Fig. 2D). The phenotype was strongly dependent on the environment with plants grown at higher temperatures or high light displaying stronger phenotype (Fig 2E and 2F). We checked the Bur-0 meristem under scanning electron microscopy, which did not reveal any drastic abnormalities in the organization of shoot apical meristem, suggesting that Bur-0 phenotype is a manifestation of abnormal leaf growth and leaf expansion.

Distorted phenotype of Bur-0 maps to a trinucleotide repeat expansion in *IIL1*

Since the Bur-0 phenotype was restricted to particular environment, to determine the genetic basis of the Bur-0 phenotypic defect we crossed Bur-0 with other strains, Col-0, Pf-0, Pi-0 and Wl-0. Phenotypic analyses of F1 generations from these crosses did not shown a Bur-0 phenotype indicative of simple recessive genetic inheritance. Phenotypic analyses of F2 populations from the crosses were largely varied between the crosses, suggesting the presence of strain specific genetic modifiers. However the highest proportion of Bur-0 plants, nearly 25% segregation was observed in Bur-0 x Pf-0 cross. We took advantage of this recessive inheritance and large effective locus to perform mapping using SSLP (simple sequence length polymorphism) markers. We grew 600 F2 plants and used roughly 150 plants with *iil* phenotype for the mapping to narrow down

the interval to a 16.7Kb region. The mapped region in Col-0 contains two complete genes At4g13420, At4g13430 and first seven exons on At4g13410 (Fig 2A).

Phenotypic analyses of T-DNA insertion lines in the mapping interval

From collection of sequence information, we analyzed T-DNA insertion lines, for the three genes located in the mapping interval. The insertion lines are derived from Col-0 background, we identified exonic T-DNA insertions for At4g13410 and At4g13420 does not reveal any drastic phenotypes at 27°C SD conditions. At4g13420 is expressed almost exclusively in roots, suggesting that poor candidate for a trait affecting leaf growth (SCHMID *et al.* 2005). At4g13430 the exonic insertions were not available and the intronic insertions did not display *iil* phenotype. Sequencing of the cDNA for all three genes from Bur-0 failed to identify any major effect mutations. Therefore, we surveyed the entire mapping interval through PCR analysis; with that approach we discovered approximately 1.2 kb region insertion in the At4g14330 of Bur-0 strain. Sequencing revealed that this insertion was due to dramatic expansion of trinucleotide TTC/GAA repeat in the third intron of At4g13430. Both Col-0 and Pf-0 contains only 23 repeats, while Bur-0 had more than 400 repeats at this locus (Fig. 3A).

At4g14330 encodes a protein that is likely to be the large subunit of ISOPROPYL MALATE ISOMERASE, and it has a biochemical phenotype, now on it is called as *ISOPRPYLMALATE ISOMERASE LARGE SUBUNIT 1, (IIL1*, At4g14330), an enzyme involved in leucine biosynthesis, which shares similarity with fungal and bacterial enzymes (KNILL *et al.* 2009). In Friedreich ataxia *(FRDA),* an intronic GAA/TTC triplet expansion leads to 50% reduction in accumulation of the mature frataxin *(FXN)* transcript

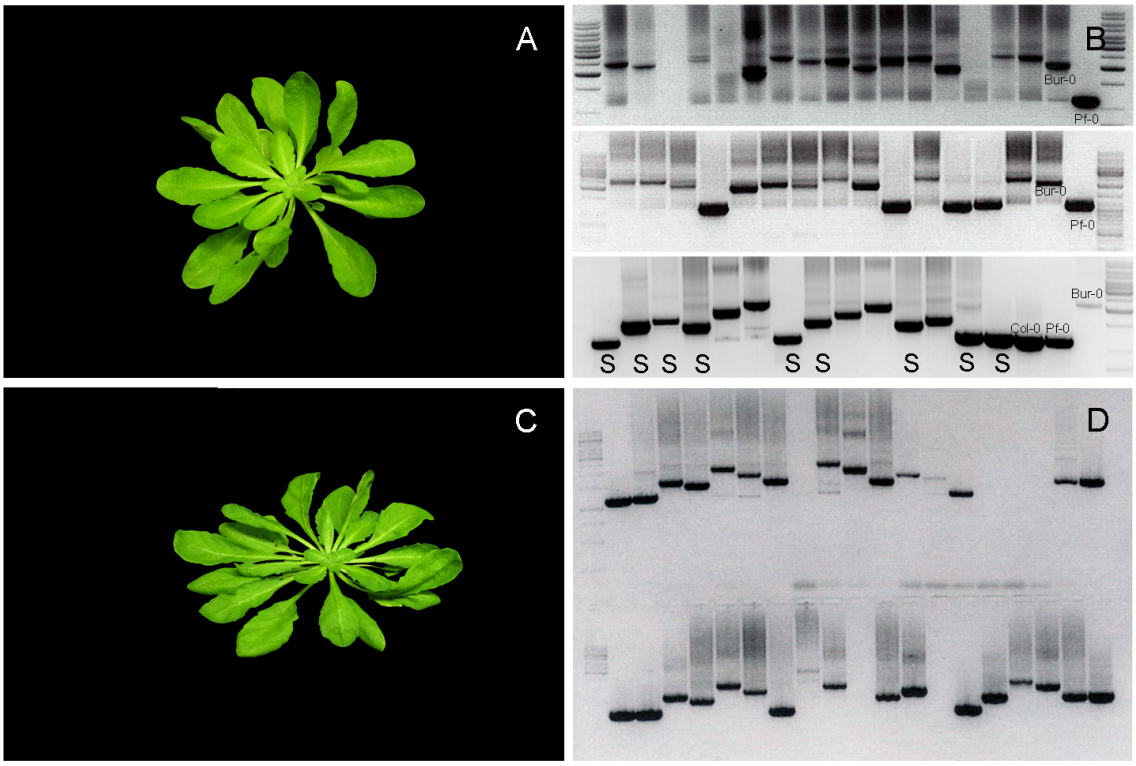

Figure 3: Triplet expansion and the iil phenotype. A) *Five week old EMS treated M_2 plants that have lost the iil phenotype screened under higher temperatures.* ***B)*** *PCR analysis of triplet repeat in EMS treated M2 plant that had retained (top) or lost (middle) the iil phenotype. Bottom lane shows PCR analysis of spontaneous revertants, with (S) indicating those in which suppression was stably transmitted to the next generation* ***C)*** *Five-week old spontaneous revertants grown under higher temperature.* ***D)*** *PCR analysis of spontaneous phenotypic revertants in transgenerations.*

(BIDICHANDANI *et al.* 1998). Long GAA/TTC repeats have been shown to interfere with transcription in a length and orientation dependent manner, and GAA/TTC expansions have a more severe phenotype than TTC/GAA (BIDICHANDANI *et al.* 1999). We compared the expression of *IIL1* in Bur-0, Col-0 and Pf-0. *IIL1* transcript was 30-50% reduced in Bur-0 compared to Col-0 and Pf-0 reference strains (Fig 3B).

If reduced *IIL1* RNA levels were the cause for the *iil* phenotype, we expected the over expression of the mature Bur-0 would rescue the growth defects in Bur-0. In order to prove and understand the hypothesis, we cloned the *IIL1* cDNA (Construct_SS03) from

Bur-0 strain, under strong viral promoter 35S CaMV. All the T1 generation over expression lines grown under higher temperature did not displayed any *iil* phenotype suggesting that higher levels of *IIL1* is sufficient to suppress the *iil* phenotype (Fig. 3C and D).

iil suppressor screens

Although the transgenic assay linked the *IIL1* gene with the *iil* phenotype, it did not prove the involvement of the triplet repeat expansion in conferring the *iil* phenotype. We sought to address this in a different way. DNA damaging agents including ethyl methyl sulfonate (EMS) has been shown to have an effect on expanded CAG repeats (HASHEM and SINDEN 2002). Therefore, we wondered whether treating Bur-0 plants with EMS could result in loss of expansion. We screened approximately 30,000 M2 progeny of EMS treated plants at 27° C in short days and screened for phenotypic suppression. Among 30,000 plants, we got 156 plants suppressed for *ill1* phenotype (Fig. 4A). Genotyping for the repeat expansion revealed that the copy number had been reduced drastically in 27 lines (Fig. 4B), suggesting that the repeat expansion is the causal polymorphism underlying the *iil* phenotype. In the next generations we tested 12 plants from 122 lines that still retained the repeat (and set seeds), for the Bur-0 phenotype and obtained 28 suppressor lines in which all 12 plants displayed the suppressor phenotype, indicating second-site suppression. (Fig 4B).

Isolation and identification of natural spontaneous suppressor (SUPS)

GO Category	Upregulated	Downregulated	% downregulated
Cell Wall	1	3	75
Chloroplast	16	32	67
Cytosol	14	3	17
Endomembrane	22	25	53
Membrane	16	5	23
Mitochondrion	12	3	20
Nucleus	36	5	12
Unknown	128	37	22
Total	245	113	31

Table 1. *Gene Ontology (GO) annotation of genes that are changed in expression in Bur-0 at 27ºC compared to Bur-0 (23ºC), Pf-0 (23ºC and 27ºC) and an EMS induced suppressor (23ºC and 27ºC). The majority of all changed genes is up regulated, with GO categories "endomembrane" and "chloroplast" being the notable exception.*

The genotyping of *IIL1* locus from Bur-0 plants indicated that the repeat is highly variable similar to the situation in humans. We reasoned that high variability may suggest that we will be able to obtain natural reduction of length alleles for the triplet expansion in *IIL1*. Therefore, we searched for a natural spontaneous suppressors of *iil.* Among 30,000 Bur-0 plants, we found 24 plants that had lost the *iil* phenotype (Fig 4C). The phenotypic suppression was confirmed in the off springs of 17 plants all of which had a reduced repeat length at least over two generational period (Fig 4D). Sequence analysis confirms that these suppressors carry a maximum of about 120 repeats at the *IIL1* locus. In some of the suppressors the repeat length is maintained at heterozygous state (Fig 4D). Analysis this suppressor in the consequent generations suggested possible penetrance, which often changed in subsequent generations. Although the repeat length varied among the suppressor, none of the stable revertants showed more than 120_n time

repeat length, supporting the assertion that triplet expansion was casual for the *iil* phenotype.

Global Gene expression studies on *iil* phenotype

While the repeat appears to be the underlying reason for the *iil* phenotype, the mechanism through which reduction in *IIL1* activity leads to the phenotype is unclear. To gain

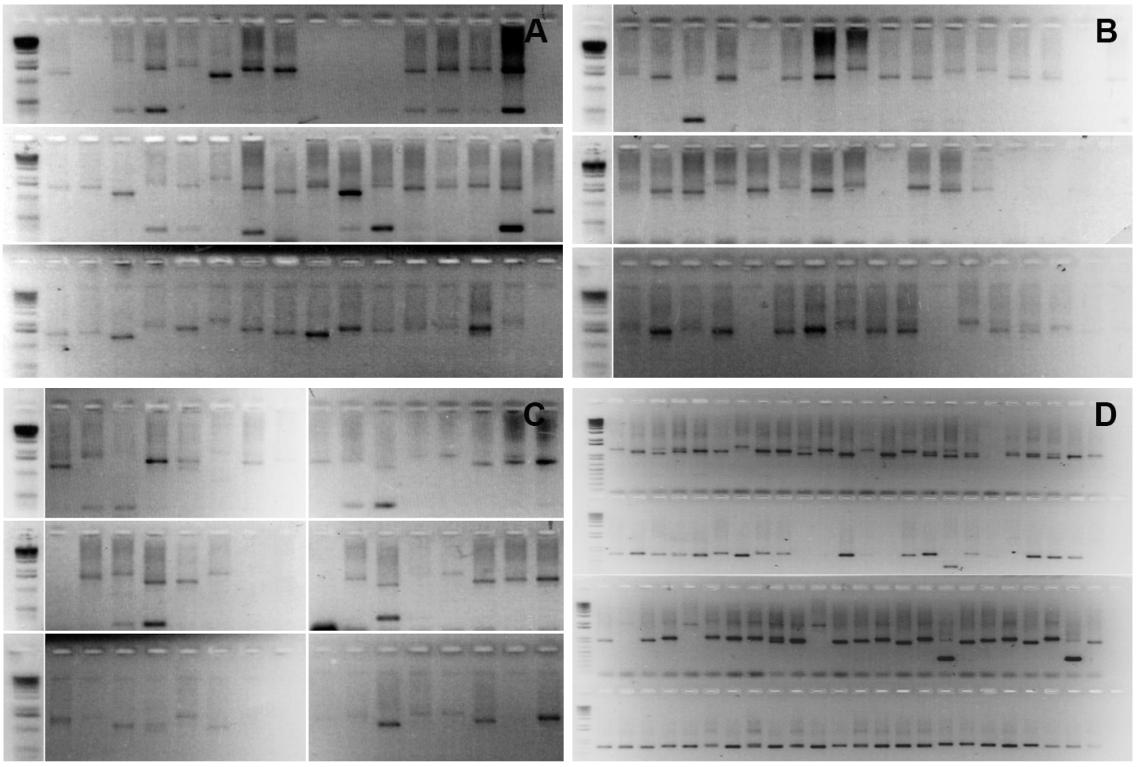

Figure 4: Somatic variability and Mutational Dynamics of repeat length. A,B and C) *PCR analysis of triplet repeat in different developmental stages (The leaf tissues were collected form week (top) week3 (middle) and week4 (bottom) of iil phenotypes grown under 27° short days (A) and 16°, 23° long days (B and C) .**D)** PCR analysis of Repeat length variability in four different spontaneous suppressors at 27° short days.*

insights in to this, we performed genome wide expression analysis by comparing global expression profiles of Bur-0 and Pf-0 along with EMS-induced suppressor grown in 23°SD and 27°SD. Analysis of the transcriptomes revealed up regulation of 70% of the

genes whose expression levels deferred in plants with and without the phenotype (Table 1). However, the organellar localization of the predicted proteins differed and many genes are annotated to encoding proteins localized to chloroplast and endomembrane were down regulated in Bur-0 strain grow under 27°SD (chi squared test<0.01 , Table 1), This suggested a problem with the chloroplasts, consistent with the likely localization of IIL1, given that amino acid biosynthesis has been shown to occur in chloroplasts (ZYBAILOV *et al.* 2008). Subsequent studies revealed that knocking down the *IIL1* gene with artificial microRNAs does lead to a chlorotic phenotype (SURESHKUMAR *et al.* 2009).

Somatic variability and Mutational Dynamics of repeat length in *iil1* plants

iil phenotype is strongly dependent on the environment and disappears in lower temperatures. Initially we grew Bur-0 in lower temperatures and different light conditions, which suggested that the observed phenotype of Bur-0 is specific to higher temperatures . Shifting Bur-0 to lower temperature like 16°C short days from 23ºC short days or shifting them to long day conditions (16ºC long days) could restore formation of normal looking leaves (data not shown). We wondered whether the triplet repeat instability itself is affected by changes in the environment. We grew Bur-0 plants at 3 different conditions (23ºC and 16ºC long days and 27ºC in short days) and assessed the variability of the expanded repeat. We assessed the repeat length in 25 plants grown at the same time under identical conditions. The repeat length was variable in all conditions and in all stages suggesting that the somatic variability occurs throughout development (Fig 5 A-C). Also, we observed in all conditions a reduction in the somatic variability at later

stage leaves (4 week old, Fig 5A-C), possibly reflecting the absence of cell division in mature leaves. Then we compared the variability in one condition with that of another. In order to assess variability, we considered the total number of bands seen, number of individuals that displayed above/below the average repeat length, number of individuals that showed reduction in repeat length etc. Based on all these parameters, we arrived at a cumulative score for repeat length variation. The genotypic data on repeat length comparison between 23°LD vs 16°LD did not show any drastic variation such as contraction or expansion as age progressed (Fig 5B and C). But there is a higher variability at 27° C short days compared to 23°C (Fig 5A). We then analyzed the suppressors, for their variability at 27ºC short days. We compared repeat length in 24 samples per suppressor (Fig. 5D). We found some suppressors to be devoid of somatic variability (Fig. 5D) suggesting a potential influence on pathways underlying mutational dynamics (Fig 5D).

Repeat length variability in populations

Since reference strain Col-0 carries only 23 GAA repeats in the 3rd exon of *IIL1* and Bur-0 carries more than 400 repeats, one wonders how would the repeat distributions be in populations. In order to address this, we sequenced the *IIL1* GAA repeat in the recently characterized 96 strains (NORDBORG *et al.* 2005). This data revealed that at least 2 strains had no repeats including loss of adjacent nucleotides and in other strains the repeat length varied with the maximum of 36 repeats (Fig 6A, Table 2, please see at the end of this chapter). However, Bur-0 was the only strain with the expanded repeat suggesting that the Bur-0 allele of *IIL1* is a rare allele in the populations of *Arabidopsis thaliana.*

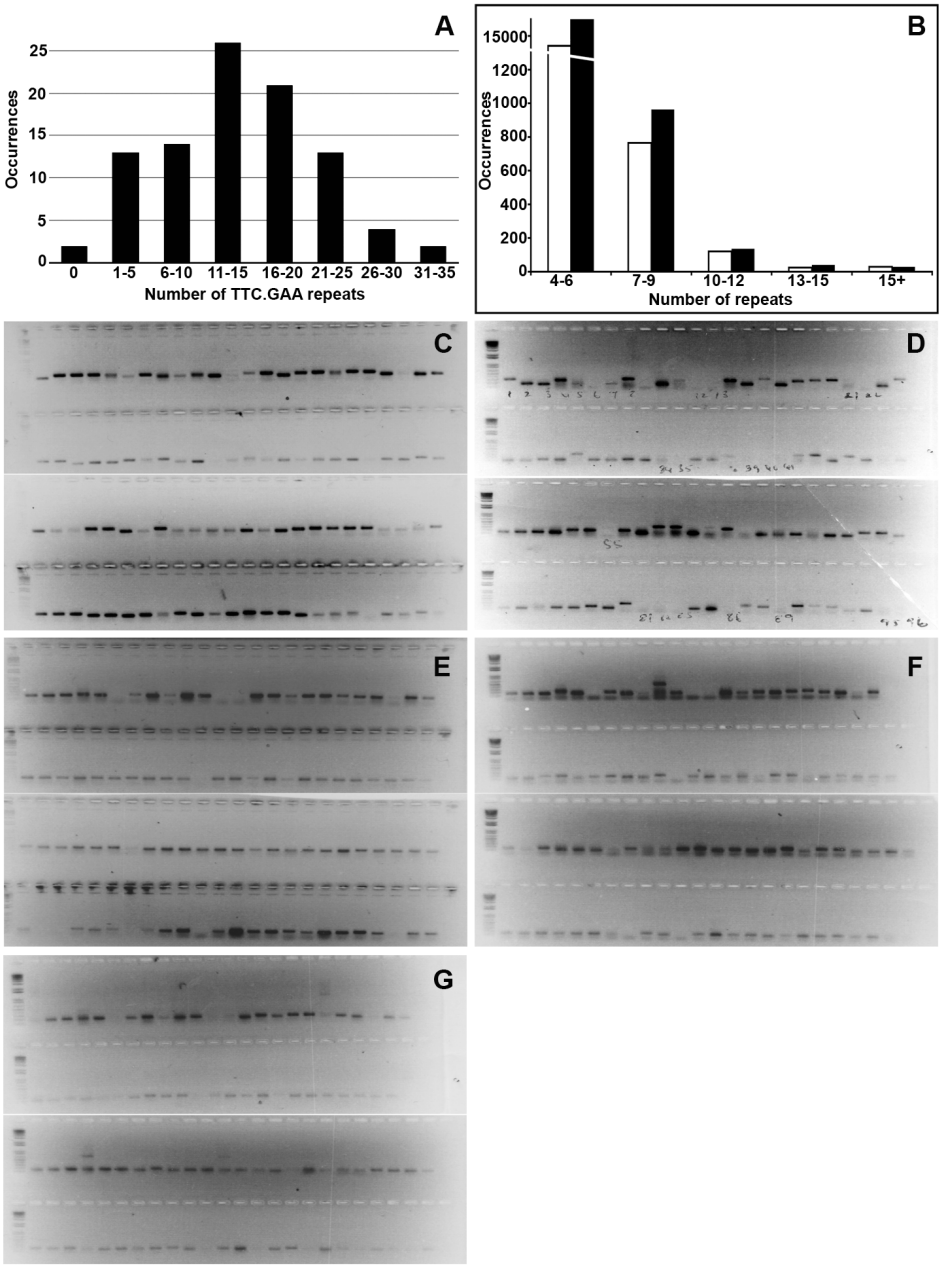

Figure 5:Triplet repeat length analysis in Arabidopsis thaliana populations. A) *Distribution of triplet repeats in A. thaliana reference genome (Col-0). White bar represents, genic regions; black bar represents, intergenic region.* ***B)*** *Distribution of IIL1 repeats copy number in 96 strains of A. thaliana.* ***C,D,E, F and G)*** *PCR analysis of triplet repeat copy number variation in exonic region of five different genes (as followed by gel figures At1g47300, At1g48400, At2g42200 , At4g11385 and At5g03710) in set of 96 natural A. thaliana strains*

If repeats are as detrimental as the *iil* phenotype indicates, one would expect to find fewer occurrences of expanded repeats in the genomes of organisms. We therefore analyzed the repeats in the *Arabidopsis thaliana* genome. Computationally the triplets can be grouped into 11 sub classes as shown in Table 3. We used a search algorithm to extract all possible non-redundant trinucleotide repeat sequences. Analysis of triplets revealed that < 1% of all triplet repeat have >6 copies (Fig 6B). Triplets with >40 copies never occur in the transcription units. 6 genes with more than 25 repeats in the coding region were expressed at very low levels in the AtGenExpress dataset. Comparison with all other triplet repeat groups TG2 (AAG, AGA, GAA, TTC, TCT, CTT) triplet group is more prevalent and densely distributed in *Arabidopsis thaliana* genome. In addition, the distribution of the TG2 group between coding and non-coding regions differed. In general, there appears to be a non-random distribution of triplet repeats in the genome of *Arabidopsis thaliana.*

Groups	Triplets
TG0	AAA , TTT, CCC, GGG
TG1	AAC, ACA, CAA, GTT, TGT, TTG
TG2	AAG, AGA, GAA, CTT, TCT, TTC
TG3	AAT, ATA, TAA, ATT, TTA, TAT
TG4	ACC, CAC, CCA, GGT, GTG, TGG
TG5	ACG, CGA, GAC, CGT, GTC, TCG
TG6	ACT, CTA, TAC, AGT, GTA, TAG
TG7	AGC, GCA, CAG, GCT, CTG, TGC
TG8	AGG, GAG, GGA, CCT, CTC, TCC
TG9	ATC, TCA, CAT, GAT, ATG, TGA
TG10	CCG, CGC, GCC, CGG, GGC, GCG

Table. 3. Sequences of triplets that form individual triplet groups.

Other potential candidates for repeat expansion

Gene	Repeat Length	Triplet Seq/Amino Acid	TG Group	Predicted Protein	SNP Number	Number ofStrains showed polymorphism
At1g47300	28	GAA/Glu	2	F-Box family	15	20
At1g48400	41	GAT/Asp	9	F-Box family	8	16
At2g42200	14	AAC/Asn	1	Squamosa binding Protein	6	12
At3g22930	22	CAA/Gln	1	Calmodulin	4	18
At4g17750	20	AAC/Asn	1	Transcription Factor	5	20
At4g02810	25	GAA/Glu	2	Unknown Protein	4	15
At4g11385	22	AAG/Lys	2	Unknown Protein	2	12
At5g49120	15	AAC/Asn	1	Senescence-Protein	16	9
At5g40340	14	GAA/Glu	2	PWWP domain	14	16
At5g03710	62	GAA/Glu	2	Unknown Protein	6	7
At5g37420	16	GAT/Asp	9	Transcription Factor	8	20

Table. 4 Candidate genes harbouring variations in triplet repeats. *The shaded genes have been tested by PCR for repeat variability.*

We combined our triplet data with that of the predicted polymorphisms in the wild strains (http://polymorph.weigel.org) and looked for genes that are both polymorphic and carried the repeats. We identified 11 candidate genes (Table 4), from there I had checked 5 genes for repeat length variation in the large scale using 96 Arabidopsis strains. Mostly these candidates were grouped to TG2, TG1 and TG9. The out come of the results explains indeed there is a natural variation in the repeat length and some of the candidate strains displayed a highest level of repeat length compare to other strains. These strains could be the strong candidates for the repeat length variation.

Conclusions

The triplet expansions underlie several human genetic disorders the genetic basis of their mutational dynamics that underlies genetic anticipation is not yet understood; this is largely due to the absence of biological models. Genetically engineering repeats have been proven to be difficult. Variability in copy number in normal individuals is common to expansion disorders in humans and this variation often underlies genetic anticipation. We have found a striking degree of somatic variation with in the plant at various developmental conditions. In addition, we found this variability is environment dependent with plants grown in different temperatures displaying varied variabilities. Specifically our data suggests higher variability with increasing temperatures. In the absence of other natural models combined with the difficulties in generating transgenic models, the discovery of a triplet repeat mediated genetic defect in *Arabidopsis* opens up numerous possibilities and provides an excellent model to study fundamental aspects of triplet expansion associated genetic diseases. Our studies have also revealed other potential candidate genes that may harbor variation in repeats among *Arabidopsis* populations and may help reveal the hidden genetic variation, which is key to understand the evolutionary implications of triplet repeats.

Contributions

I performed the mapping and identified of triplet repeat expansion in Bur-0. I have performed the cloning and phenotypic assay of the transgenic lines.

The M2-EMS pools were gifted by Marco and I performed the suppressor screen.

Suresh and myself identified the spontaneous suppressors in Bur-0 background.

Marco and myself performed the micro array experiment.

I have carried out the genotyping and sequencing the triplet repeats in 96 natural strains. Suresh, Ramya Harilal, Korbinian Schneeberger and David Richter performed genome wide triplet repeat analyses. I have analysed the candidate genes for polymorphisms in the exonic region.

Table. 2. Repeat number variability in MAGNUS 96 strains at *IIL1* locus.

S.No	accession	number	GAANumber
1	RRS-7	CS22564	4
2	RRS-10	CS22565	12
3	Knox-10	CS22566	12
4	Knox-18	CS22567	12
5	Rmx-A02	CS22568	24
6	Rmx-A180	CS22569	23
7	Pna-17	CS22570	12
8	Pna-10	CS22571	12
9	Eden-1	CS22572	10
10	Eden-2	CS22573	10
11	Lov-1	CS22574	10
12	Lov-5	CS22575	10
13	Fab-2	CS22576	4
14	Fab-4	CS22577	4
15	Bil-5	CS22578	26
16	Bil-7	CS22579	26
17	Var2-1	CS22580	17
18	Var2-6	CS22581	17
19	Spr1-2	CS22582	9
20	Spr1-6	CS22583	4
21	Omo2-1	CS22584	12
22	Omo2-3	CS22585	18
23	Ull2-5	CS22586	19
24	Ull2-3	CS22587	22
25	Zdr-1	CS22588	21
26	Zdr-6	CS22589	18
27	Bor-1	CS22590	4
28	Bor-4	CS22591	12
29	Pu2-7	CS22592	4
30	Pu2-23	CS22593	14
31	Lp2-2	CS22594	4
32	Lp2-6	CS22595	10
33	HR-5	CS22596	4
34	HR-10	CS22597	9
35	NFA-8	CS22598	27
36	NFA-10	CS22599	28
37	Sq-1	CS22600	4
38	Sq-8	CS22601	18
39	CIBC-5	CS22602	34
40	CIBC-17	CS22603	15
41	Tamm-2	CS22604	15
42	Tamm-27	CS22605	15
43	Kz-1	CS22606	12
44	Kz-9	CS22607	9

45	Got-7	CS22608	19
46	Got-22	CS22609	19
47	Ren-1	CS22610	12
48	Ren-11	CS22611	12
49	Uod-1	CS22612	0
50	Uod-7	CS22613	0
51	Cvi-0	CS22614	4
52	Lz-0	CS22615	12
53	Ei-2	CS22616	17
54	Gu-0	CS22617	22
55	Ler-1	CS22618	18
56	Nd-1	CS22619	12
57	C24	CS22620	18
58	CS22491	CS22621	19
59	Wei-0	CS22622	15
60	Ws-0	CS22623	10
61	Yo-0	CS22624	15
62	Col-0	CS22625	23
63	An-1	CS22626	16
64	Van-0	CS22627	17
65	Br-0	CS22628	9
66	Est-1	CS22629	18
67	Ag-0	CS22630	4
68	Gy-0	CS22631	17
69	Ra-0	CS22632	12
70	Bay-0	CS22633	23
71	Ga-0	CS22634	21
72	Mrk-0	CS22635	12
73	Mz-0	CS22636	14
74	Wt-5	CS22637	23
75	Kas-1	CS22638	13
76	Ct-1	CS22639	17
77	Mr-0	CS22640	9
78	Tsu-1	CS22641	17
79	Mt-0	CS22642	24
80	Nok-3	CS22643	16
81	Wa-1	CS22644	12
82	Fei-0	CS22645	9
83	Se-0	CS22646	15
84	Ts-1	CS22647	24
85	Ts-5	CS22648	12
86	Pro-0	CS22649	9
87	LL-0	CS22650	4
88	Kondara	CS22651	11
89	Shahdara	CS22652	33
90	Sorbo	CS22653	9
91	Kin-0	CS22654	12

92	Ms-0	CS22655	5
93	Bur-0	CS22656	400+
94	Edi-0	CS22657	9
95	Oy-0	CS22658	23
96	Ws-2	CS22659	18

REFERENCES

ALBA, R., P. M. KELMENSON, M. M. CORDONNIER-PRATT and L. H. PRATT, 2000 The phytochrome gene family in tomato and the rapid differential evolution of this family in angiosperms. Mol Biol Evol **17:** 362-373.

ALONSO-BLANCO, C., M. G. AARTS, L. BENTSINK, J. J. KEURENTJES, M. REYMOND *et al.*, 2009 What has natural variation taught us about plant development, physiology, and adaptation? Plant Cell **21:** 1877-1896.

ALONSO-BLANCO, C., S. E. EL-ASSAL, G. COUPLAND and M. KOORNNEEF, 1998 Analysis of natural allelic variation at flowering time loci in the Landsberg *erecta* and Cape Verde Islands ecotypes of *Arabidopsis thaliana.* Genetics **149:** 749-764.

ALONSO-BLANCO, C., B. MENDEZ-VIGO and M. KOORNNEEF, 2005 From phenotypic to molecular polymorphisms involved in naturally occurring variation of plant development. Int J Dev Biol **49:** 717-732.

ALPER, G., and V. NARAYANAN, 2003 Friedreich's ataxia. Pediatr Neurol **28:** 335-341.

ARANZANA, M. J., S. KIM, K. ZHAO, E. BAKKER, M. HORTON *et al.*, 2005 Genome-wide association mapping in *Arabidopsis* identifies previously known flowering time and pathogen resistance genes. PLoS Genet **1:** e60.

ARMOUR, J. A., 2006 Tandemly repeated DNA: why should anyone care? Mutat Res **598:** 6-14.

ATWELL, S., Y. S. HUANG, B. J. VILHJALMSSON, G. WILLEMS, M. HORTON *et al.*, Genome-wide association study of 107 phenotypes in *Arabidopsis thaliana* inbred lines. Nature **465:** 627-631.

AUKERMAN, M. J., M. HIRSCHFELD, L. WESTER, M. WEAVER, T. CLACK *et al.*, 1997 A deletion in the *PHYD* gene of the *Arabidopsis* Wassilewskija ecotype defines a role for phytochrome D in red/far-red light sensing. Plant Cell **9:** 1317-1326.

AXEISSON, T., O. SHAVORSKAYA and U. LAGERCRANTZ, 2001 Multiple flowering time QTLs within several Brassica species could be the result of duplicated copies of one ancestral gene. Genome **44:** 856-864.

BALASUBRAMANIAN, S., C. SCHWARTZ, A. SINGH, N. WARTHMANN, M. C. KIM *et al.*, 2009 QTL mapping in New *Arabidopsis thaliana* advanced intercross - recombinant inbred lines. PLoS ONE **4**(2): e4318.

BALASUBRAMANIAN, S., S. SURESHKUMAR, J. LEMPE and D. WEIGEL, 2006 Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature. PLoS Genet **2:** e106.

BENTSINK, L., J. JOWETT, C. J. HANHART and M. KOORNNEEF, 2006 Cloning of DOG1, a quantitative trait locus controlling seed dormancy in *Arabidopsis*. Proc Natl Acad Sci U S A **103:** 17042-17047.

BERCIANO, J., O. COMBARROS, M. DE CASTRO and F. PALAU, 1997 Intronic GAA triplet repeat expansion in Friedreich ataxia presenting with pure sensory ataxia. J Neurol **244:** 390-391.

BIDICHANDANI, S. I., T. ASHIZAWA and P. I. PATEL, 1997 Atypical Friedreich ataxia caused by compound heterozygosity for a novel missense mutation and the GAA triplet-repeat expansion. Am J Hum Genet **60:** 1251-1256.

BIDICHANDANI, S. I., T. ASHIZAWA and P. I. PATEL, 1998 The GAA triplet-repeat expansion in Friedreich ataxia interferes with transcription and may be associated with an unusual DNA structure. Am J Hum Genet **62:** 111-121.

BIDICHANDANI, S. I., C. A. GARCIA, P. I. PATEL and M. M. DIMACHKIE, 2000 Very late-onset Friedreich ataxia despite large GAA triplet repeat expansions. Arch Neurol **57:** 246-251.

BIDICHANDANI, S. I., S. M. PURANDARE, E. E. TAYLOR, G. GUMIN, H. MACHKHAS *et al.*, 1999 Somatic sequence variation at the Friedreich ataxia locus includes complete contraction of the expanded GAA triplet repeat, significant length variation in serially passaged lymphoblasts and enhanced mutagenesis in the flanking sequence. Hum Mol Genet **8:** 2425-2436.

BLAZQUEZ, M. A., J. H. AHN and D. WEIGEL, 2003 A thermosensory pathway controlling flowering time in *Arabidopsis thaliana.* Nat Genet **33:** 168-171.

BOMBLIES, K., J. LEMPE, P. EPPLE, N. WARTHMANN, C. LANZ *et al.*, 2007 Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. PLoS Biol **5:** e236.

BOREVITZ, J. O., J. N. MALOOF, J. LUTES, T. DABI, J. L. REDFERN *et al.*, 2002 Quantitative trait loci controlling light and hormone response in two accessions of Arabidopsis thaliana. Genetics **160:** 683-696.

BRACHI, B., N. FAURE, M. HORTON, E. FLAHAUW, A. VAZQUEZ *et al.*, Linkage and association mapping of *Arabidopsis thaliana* flowering time in nature. PLoS Genet **6:** e1000940.

BROCK, M. T., J. R. STINCHCOMBE and C. WEINIG, 2009 Indirect effects of FRIGIDA: floral trait (co)variances are altered by seasonally variable abiotic factors associated with flowering time. J Evol Biol **22:** 1826-1838.

BROMAN, K. W., H. WU, S. SEN and G. A. CHURCHILL, 2003 R/qtl: QTL mapping in experimental crosses. Bioinformatics **19:** 889-890.

BROWN, L. Y., and S. A. BROWN, 2004 Alanine tracts: the expanding story of human illness and trinucleotide repeats. Trends Genet **20:** 51-58.

CAICEDO, A. L., C. RICHARDS, I. M. EHRENREICH and M. D. PURUGGANAN, 2009 Complex rearrangements lead to novel chimeric gene fusion polymorphisms at the *Arabidopsis thaliana* MAF2-5 flowering time gene cluster. Mol Biol Evol **26:** 699-711.

CAICEDO, A. L., J. R. STINCHCOMBE, K. M. OLSEN, J. SCHMITT and M. D. PURUGGANAN, 2004 Epistatic interaction between *Arabidopsis FRI* and *FLC* flowering time genes generates a latitudinal cline in a life history trait. Proc Natl Acad Sci U S A **101:** 15670-15675.

CHI, L. M., and S. L. LAM, 2005 Structural roles of CTG repeats in slippage expansion during DNA replication. Nucleic Acids Res **33:** 1604-1617.

CHORY, J., M. CHATTERJEE, R. K. COOK, T. ELICH, C. FANKHAUSER *et al.*, 1996 From seed germination to flowering, light controls plant development via the pigment phytochrome. Proc Natl Acad Sci U S A **93:** 12066-12071.

CHORY, J., T. ELICH, H. M. LI, A. PEPPER, D. POOLE *et al.*, 1994 Genes controlling *Arabidopsis* photomorphogenesis. Biochem Soc Symp **60:** 257-263.

CLARK, R. M., S. S. BHASKAR, M. MIYAHARA, G. L. DALGLIESH and S. I. BIDICHANDANI, 2006 Expansion of GAA trinucleotide repeats in mammals. Genomics **87:** 57-67.

CLARK, R. M., G. L. DALGLIESH, D. ENDRES, M. GOMEZ, J. TAYLOR *et al.*, 2004 Expansion of GAA triplet repeats in the human genome: unique origin of the FRDA mutation at the center of an Alu. Genomics **83:** 373-383.

CLARK, R. M., I. DE BIASE, A. P. MALYKHINA, S. AL-MAHDAWI, M. POOK *et al.*, 2007a The GAA triplet-repeat is unstable in the context of the human FXN locus and displays age-dependent expansions in cerebellum and DRG in a transgenic mouse model. Hum Genet **120:** 633-640.

CLARK, R. M., G. SCHWEIKERT, C. TOOMAJIAN, S. OSSOWSKI, G. ZELLER *et al.*, 2007b Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. Science **317:** 338-342.

COWL, J. S., N. HARTLEY, D. X. XIE, G. C. WHITELAM, G. P. MURPHY *et al.*, 1994 The *PHYC* gene of *Arabidopsis*. Absence of the third intron found in *PHYA* and *PHYB*. Plant Physiol **106:** 813-814.

DESLANDES, L., F. PILEUR, L. LIAUBET, S. CAMUT, C. CAN *et al.*, 1998 Genetic characterization of *RRS1*, a recessive locus in *Arabidopsis thaliana* that confers resistance to the bacterial soilborne pathogen *Ralstonia solanacearum.* Mol Plant Microbe Interact **11:** 659-667.

DEVOS, K. M., J. BEALES, Y. OGIHARA and A. N. DOUST, 2005 Comparative sequence analysis of the phytochrome C gene and its upstream region in allohexaploid

wheat reveals new data on the evolution of its three constituent genomes. Plant Mol Biol **58:** 625-641.

DUYAO, M., C. AMBROSE, R. MYERS, A. NOVELLETTO, F. PERSICHETTI *et al.*, 1993 Trinucleotide repeat length instability and age of onset in Huntington's disease. Nat Genet **4:** 387-392.

DYBBS, M., J. NGAI and J. M. KAPLAN, 2005 Using microarrays to facilitate positional cloning: identification of tomosyn as an inhibitor of neurosecretion. PLoS Genet **1:** 6-16.

EL-DIN EL-ASSAL, S., C. ALONSO-BLANCO, A. J. PEETERS, V. RAZ and M. KOORNNEEF, 2001 A QTL for flowering time in *Arabidopsis* reveals a novel allele of *CRY2*. Nat Genet **29:** 435-440.

EL-LITHY, M. E., L. BENTSINK, C. J. HANHART, G. J. RUYS, D. ROVITO *et al.*, 2006 New *Arabidopsis* recombinant inbred line populations genotyped using SNPWave and their use for mapping flowering-time quantitative trait loci. Genetics **172:** 1867-1876.

EL-LITHY, M. E., E. J. CLERKX, G. J. RUYS, M. KOORNNEEF and D. VREUGDENHIL, 2004 Quantitative trait locus analysis of growth-related traits in a new *Arabidopsis* recombinant inbred population. Plant Physiol **135:** 444-458.

ENDLER, J., 1977 *Geographic variation, speciation and the clines*. Princeton University Press, Princeton (New Jersey).

FILIAULT, D. L., C. A. WESSINGER, J. R. DINNENY, J. LUTES, J. O. BOREVITZ *et al.*, 2008 Amino acid polymorphisms in *Arabidopsis* phytochrome B cause differential responses to light. Proc Natl Acad Sci U S A **105:** 3157-3162.

FILICHKIN, S. A., H. D. PRIEST, S. A. GIVAN, R. SHEN, D. W. BRYANT *et al.*, 2010 Genome-wide mapping of alternative splicing in *Arabidopsis thaliana*. Genome Res **20:** 45-58.

FONDON, J. W., 3RD, E. A. HAMMOCK, A. J. HANNAN and D. G. KING, 2008 Simple sequence repeats: genetic modulators of brain function and behavior. Trends Neurosci, **31:** 328-334.

FRANKLIN, K. A., S. J. DAVIS, W. M. STODDART, R. D. VIERSTRA and G. C. WHITELAM, 2003a Mutant analyses define multiple roles for phytochrome C in *Arabidopsis* photomorphogenesis. Plant Cell **15:** 1981-1989.

FRANKLIN, K. A., U. PRAEKELT, W. M. STODDART, O. E. BILLINGHAM, K. J. HALLIDAY *et al.*, 2003b Phytochromes B, D, and E act redundantly to control multiple physiological responses in *Arabidopsis*. Plant Physiol **131:** 1340-1346.

FRARY, A., T. C. NESBITT, S. GRANDILLO, E. KNAAP, B. CONG *et al.*, 2000 fw2.2: a quantitative trait locus key to the evolution of tomato fruit size. Science **289:** 85-88.

GATCHEL, J. R., and H. Y. ZOGHBI, 2005 Diseases of unstable repeat expansion: mechanisms and common principles. Nat Rev Genet **6:** 743-755.

GAZZANI, S., A. R. GENDALL, C. LISTER and C. DEAN, 2003 Analysis of the molecular basis of flowering time variation in *Arabidopsis* accessions. Plant Physiol **132:** 1107-1114.

GENDALL, A. R., Y. Y. LEVY, A. WILSON and C. DEAN, 2001 The *VERNALIZATION 2* gene mediates the epigenetic regulation of vernalization in *Arabidopsis*. Cell **107:** 525-535.

GOMES-PEREIRA, M., L. FOIRY, A. NICOLE, A. HUGUET, C. JUNIEN *et al.*, 2007 CTG trinucleotide repeat "big jumps": large expansions, small mice. PLoS Genet **3:** e52.

GOMES-PEREIRA, M., M. T. FORTUNE and D. G. MONCKTON, 2001 Mouse tissue culture models of unstable triplet repeats: in vitro selection for larger alleles, mutational expansion bias and tissue specificity, but no association with cell division rates. Hum Mol Genet **10:** 845-854.

GRABCZYK, E., M. MANCUSO and M. C. SAMMARCO, 2007 A persistent RNA.DNA hybrid formed by transcription of the Friedreich ataxia triplet repeat in live bacteria, and by T7 RNAP in vitro. Nucleic Acids Res **35:** 5351-5359.

GRABCZYK, E., and K. USDIN, 2000 The GAA*TTC triplet repeat expanded in Friedreich's ataxia impedes transcription elongation by T7 RNA polymerase in a length and supercoil dependent manner. Nucleic Acids Res **28:** 2815-2822.

GUSELLA, J. F., M. E. MACDONALD, C. M. AMBROSE and M. P. DUYAO, 1993 Molecular genetics of Huntington's disease. Arch Neurol **50:** 1157-1163.

HAGENBLAD, J., and M. NORDBORG, 2002 Sequence variation and haplotype structure surrounding the flowering time locus *FRI* in *Arabidopsis thaliana*. Genetics **161:** 289-298.

HALLIDAY, K. J., M. G. SALTER, E. THINGNAES and G. C. WHITELAM, 2003 Phytochrome control of flowering is temperature sensitive and correlates with expression of the floral integrator *FT*. Plant J **33:** 875-885.

HALLIDAY, K. J., B. THOMAS and G. C. WHITELAM, 1997 Expression of heterologous phytochromes A, B or C in transgenic tobacco plants alters vegetative development and flowering time. Plant J **12:** 1079-1090.

HASHEM, V. I., and R. R. SINDEN, 2002 Chemotherapeutically induced deletion of expanded triplet repeats. Mutat Res **508:** 107-119.

HEIDENFELDER, B. L., A. M. MAKHOV and M. D. TOPAL, 2003 Hairpin formation in Friedreich's ataxia triplet repeat expansion. J Biol Chem **278:** 2425-2431.

HEIDENFELDER, B. L., and M. D. TOPAL, 2003 Effects of sequence on repeat expansion during DNA replication. Nucleic Acids Res **31:** 7159-7164.

HOFFMANN, M. H., 2002 Biogeography of *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae). J. Biogeography **29:** 125-134.

IRIZARRY, R. A., B. M. BOLSTAD, F. COLLIN, L. M. COPE, B. HOBBS *et al.*, 2003a Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res **31:** e15.

IRIZARRY, R. A., S. L. OOI, Z. WU and J. D. BOEKE, 2003b Use of mixture models in a microarray-based screening procedure for detecting differentially represented yeast mutants. Stat Appl Genet Mol Biol **2:** Article1.

JACKSON, S. M., A. J. WHITWORTH, J. C. GREENE, R. T. LIBBY, S. L. BACCAM *et al.*, 2005 A SCA7 CAG/CTG repeat expansion is stable in Drosophila melanogaster despite modulation of genomic context and gene dosage. Gene **347:** 35-41.

JIANG, D., X. GU and Y. HE, 2009 Establishment of the winter-annual growth habit via *FRIGIDA*-mediated histone methylation at *FLOWERING LOCUS C* in *Arabidopsis*. Plant Cell **21:** 1733-1746.

JOHANSON, U., J. WEST, C. LISTER, S. MICHAELS, R. AMASINO *et al.*, 2000 Molecular analysis of *FRIGIDA*, a major determinant of natural variation in *Arabidopsis* flowering time. Science **290:** 344-347.

JUNG, J., and N. BONINI, 2007 CREB-binding protein modulates repeat instability in a *Drosophila* model for polyQ disease. Science **315:** 1857-1859.

KNILL, T., M. REICHELT, C. PAETZ, J. GERSHENZON and S. BINDER, 2009 Arabidopsis thaliana encodes a bacterial-type heterodimeric isopropylmalate isomerase involved in both Leu biosynthesis and the Met chain elongation pathway of glucosinolate formation. Plant Mol Biol **71:** 227-239.

KOMEDA, Y., 2004 Genetic regulation of time to flower in *Arabidopsis thaliana*. Annu Rev Plant Biol **55:** 521-535.

KOORNNEEF, M., 1997 Plant development: timing when to flower. Curr Biol **7:** R651-652.

KOORNNEEF, M., C. ALONSO-BLANCO, A. J. PEETERS and W. SOPPE, 1998 Genetic Control of Flowering Time in *Arabidopsis*. Annu Rev Plant Physiol Plant Mol Biol **49:** 345-370.

KOORNNEEF, M., C. ALONSO-BLANCO and D. VREUGDENHIL, 2004 Naturally occurring genetic variation in *Arabidopsis thaliana*. Annu Rev Plant Biol **55:** 141-172.

KRASILNIKOVA, M. M., M. L. KIREEVA, V. PETROVIC, N. KNIJNIKOVA, M. KASHLEV *et al.*, 2007 Effects of Friedreich's ataxia (GAA)n*(TTC)n repeats on RNA synthesis and stability. Nucleic Acids Res **35:** 1075-1084.

KUMAR, S. V., and P. A. WIGGE, 2010 H2A.Z-containing nucleosomes mediate the thermosensory response in *Arabidopsis*. Cell **140:** 136-147.

KUMARI, D., and K. USDIN, 2009 Chromatin remodeling in the noncoding repeat expansion diseases. J Biol Chem **284:** 7413-7417.

LE CORRE, V., F. ROUX and X. REBOUD, 2002 DNA polymorphism at the *FRIGIDA* gene in *Arabidopsis thaliana*: extensive nonsynonymous variation is consistent with local selection for flowering time. Mol Biol Evol **19:** 1261-1271.

LE ROUZIC, A., P. B. SIEGEL and O. CARLBORG, 2007 Phenotypic evolution from genetic polymorphisms in a radial network architecture. BMC Biol **5:** 50.

LEE, J. H., S. J. YOO, S. H. PARK, I. HWANG, J. S. LEE *et al.*, 2007 Role of SVP in the control of flowering time by ambient temperature in *Arabidopsis*. Genes Dev **21:** 397-402.

LEE, S., and M. S. PARK, 2002 Human *FEN-1* can process the 5'-flap DNA of CTG/CAG triplet repeat derived from human genetic diseases by length and sequence dependent manner. Exp Mol Med **34:** 313-317.

LEMON, W., S. LIYANARACHCHI and M. YOU, 2003 A high performance test of differential gene expression for oligonucleotide arrays. Genome Biology **4:** R67.

LEMPE, J., S. BALASUBRAMANIAN, S. SURESHKUMAR, A. SINGH, M. SCHMID *et al.*, 2005 Diversity of flowering responses in wild *Arabidopsis thaliana* strains. PLoS Genet **1:** 109-118.

LEVY, Y. Y., S. MESNAGE, J. S. MYLNE, A. R. GENDALL and C. DEAN, 2002 Multiple roles of *Arabidopsis VRN1* in vernalization and flowering time control. Science **297:** 243-246.

LIM, J., J. CRESPO-BARRETO, P. JAFAR-NEJAD, A. B. BOWMAN, R. RICHMAN *et al.*, 2008 Opposing effects of polyglutamine expansion on native protein complexes contribute to SCA1. Nature **452:** 713-718.

LIU, R., J. QIU, L. D. FINGER, L. ZHENG and B. SHEN, 2006 The DNA-protein interaction modes of *FEN-1* with gap substrates and their implication in preventing duplication mutations. Nucleic Acids Res **34:** 1772-1784.

LOUDET, O., S. CHAILLOU, C. CAMILLERI, D. BOUCHEZ and F. DANIEL-VEDELE, 2002 Bay-0 x Shahdara recombinant inbred line population: a powerful tool for the genetic dissection of complex traits in *Arabidopsis*. Theor Appl Genet **104:** 1173-1184.

MACKAY, T. F., and C. H. LANGLEY, 1990 Molecular and phenotypic variation in the achaete-scute region of *Drosophila melanogaster*. Nature **348:** 64-66.

MALOOF, J. N., J. O. BOREVITZ, T. DABI, J. LUTES, R. B. NEHRING *et al.*, 2001 Natural variation in light sensitivity of *Arabidopsis*. Nat Genet **29:** 441-446.

MATHEWS, S., 2005 Phytochrome evolution in green and nongreen plants. J Hered **96:** 197-204.

MENDEZ-VIGO, B., M. T. DE ANDRES, M. RAMIRO, J. M. MARTINEZ-ZAPATER and C. ALONSO-BLANCO, 2010 Temporal analysis of natural variation for the rate of leaf production and its relationship with flowering initiation in *Arabidopsis thaliana*. J Exp Bot **61:** 1611-1623.

MICHAEL, T. P., P. A. SALOME, H. J. YU, T. R. SPENCER, E. L. SHARP *et al.*, 2003 Enhanced fitness conferred by naturally occurring variation in the circadian clock. Science **302:** 1049-1053.

MICHAELS, S. D., and R. M. AMASINO, 2001 Loss of *FLOWERING LOCUS C* activity eliminates the late-flowering phenotype of *FRIGIDA* and autonomous pathway mutations but not responsiveness to vernalization. Plant Cell **13:** 935-941.

MIRKIN, E. V., and S. M. MIRKIN, 2007 Replication fork stalling at natural impediments. Microbiol Mol Biol Rev **71:** 13-35.

MIRKIN, S. M., 2007 Expandable DNA repeats and human disease. Nature **447:** 932-940.

MITRA, R. M., C. A. GLEASON, A. EDWARDS, J. HADFIELD, J. A. DOWNIE *et al.*, 2004 A Ca2+/calmodulin-dependent protein kinase required for symbiotic nodule development: Gene identification by transcript-based cloning. Proc Natl Acad Sci U S A **101:** 4701-4705.

MONTE, E., J. M. ALONSO, J. R. ECKER, Y. ZHANG, X. LI *et al.*, 2003 Isolation and characterization of *phyC* mutants in *Arabidopsis* reveals complex crosstalk between phytochrome signaling pathways. Plant Cell **15:** 1962-1980.

MOUCHEL, C. F., G. C. BRIGGS and C. S. HARDTKE, 2004 Natural genetic variation in Arabidopsis identifies *BREVIS RADIX*, a novel regulator of cell proliferation and elongation in the root. Genes Dev **18:** 700-714.

MOURADOV, A., F. CREMER and G. COUPLAND, 2002 Control of flowering time: interacting pathways as a basis for diversity. Plant Cell **14 Suppl:** S111-130.

MYERS, R. H., M. E. MACDONALD, W. J. KOROSHETZ, M. P. DUYAO, C. M. AMBROSE *et al.*, 1993 De novo expansion of a (CAG)n repeat in sporadic Huntington's disease. Nat Genet **5:** 168-173.

NORDBORG, M., T. T. HU, Y. ISHINO, J. JHAVERI, C. TOOMAJIAN *et al.*, 2005 The pattern of polymorphism in *Arabidopsis thaliana*. PLoS Biol **3:** e196.

OHSHIMA, K., S. KANG, J. E. LARSON and R. D. WELLS, 1996 Cloning, characterization, and properties of seven triplet repeat DNA sequences. J Biol Chem **271:** 16773-16783.

OHSHIMA, K., L. MONTERMINI, R. D. WELLS and M. PANDOLFO, 1998 Inhibitory effects of expanded GAA.TTC triplet repeats from intron I of the Friedreich ataxia gene on transcription and replication in vivo. J Biol Chem **273:** 14588-14595.

OHSHIMA, K., and R. D. WELLS, 1997 Hairpin formation during DNA synthesis primer realignment in vitro in triplet repeat sequences from human hereditary disease genes. J Biol Chem **272:** 16798-16806.

OLSEN, K. M., S. S. HALLDORSDOTTIR, J. R. STINCHCOMBE, C. WEINIG, J. SCHMITT *et al.*, 2004 Linkage disequilibrium mapping of *Arabidopsis CRY2* flowering time alleles. Genetics **167:** 1361-1369.

ORR, H. A., 1998 The Population Genetics of Adaptation: The Distribution of Factors Fixed during Adaptive Evolution. Evolution **52:** 935-949.

ORR, H. T., and H. Y. ZOGHBI, 2007 Trinucleotide repeat disorders. Annu Rev Neurosci **30:** 575-621.

PALUSA, S. G., G. S. ALI and A. S. REDDY, 2007 Alternative splicing of pre-mRNAs of *Arabidopsis* serine/arginine-rich proteins: regulation by hormones and stresses. Plant J **49:** 1091-1107.

PEARSON, C. E., K. NICHOL EDAMURA and J. D. CLEARY, 2005 Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet **6:** 729-742.

PHILLIPS, W., A. MICHELL, H. PRUESS and R. A. BARKER, 2009 Animal models of neurodegenerative diseases. Methods Mol Biol **549:** 137-155.

POLLARD, L. M., R. SHARMA, M. GOMEZ, S. SHAH, M. B. DELATYCKI *et al.*, 2004 Replication-mediated instability of the GAA triplet repeat mutation in Friedreich ataxia. Nucleic Acids Res **32:** 5962-5971.

PUTTERILL, J., 2001 Flowering in time: genes controlling photoperiodic flowering in *Arabidopsis*. Philos Trans R Soc Lond B Biol Sci **356:** 1761-1767.

PUTTERILL, J., R. LAURIE and R. MACKNIGHT, 2004 It's time to flower: the genetic control of flowering time. Bioessays **26:** 363-373.

QIN, M., R. KUHN, S. MORAN and P. H. QUAIL, 1997 Overexpressed phytochrome C has similar photosensory specificity to phytochrome B but a distinctive capacity to enhance primary leaf expansion. Plant J **12:** 1163-1172.

QUESADA, V., C. DEAN and G. G. SIMPSON, 2005 Regulated RNA processing in the control of *Arabidopsis* flowering. Int J Dev Biol **49:** 773-780.

REDDY, A. S., 2007 Alternative splicing of pre-messenger RNAs in plants in the genomic era. Annu Rev Plant Biol **58:** 267-294.

REED, J. W., and J. CHORY, 1994 Mutational analyses of light-controlled seedling development in *Arabidopsis*. Semin Cell Biol **5:** 327-334.

RICHARD, G. F., and B. DUJON, 1996 Distribution and variability of trinucleotide repeats in the genome of the yeast *Saccharomyces cerevisiae*. Gene **174:** 165-174.

RICHARDS, R. I., 2001 Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. Hum Mol Genet **10:** 2187-2194.

ROSLOSKI, S. M., S. S. JALI, S. BALASUBRAMANIAN, D. WEIGEL and V. GRBIC, 2010, Natural Diversity in Flowering Responses of *Arabidopsis thaliana* Caused by Variation in a Tandem Gene Array. Genetics **186:** 263-276.

ROZAS, J., J. C. SANCHEZ-DELBARRIO, X. MESSEGUER and R. ROZAS, 2003 DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics **19:** 2496-2497.

RUS, A., I. BAXTER, B. MUTHUKUMAR, J. GUSTIN, B. LAHNER *et al.*, 2006 Natural variants of AtHKT1 enhance Na+ accumulation in two wild populations of *Arabidopsis*. PLoS Genet **2:** e210.

SAKAMOTO, N., P. D. CHASTAIN, P. PARNIEWSKI, K. OHSHIMA, M. PANDOLFO *et al.*, 1999 Sticky DNA: self-association properties of long GAA.TTC repeats in R.R.Y triplex structures from Friedreich's ataxia. Mol Cell **3:** 465-475.

SANDA, S., and R. AMASINO, 1995 Genetic and physiological analysis of flowering time in the C24 line of *Arabidopsis thaliana.* . Weeds World **2:** 2-8.

SCARCELLI, N., J. M. CHEVERUD, B. A. SCHAAL and P. X. KOVER, 2007 Antagonistic pleiotropic effects reduce the potential adaptive value of the *FRIGIDA* locus. Proc Natl Acad Sci U S A **104:** 16986-16991.

SCARCELLI, N., and P. X. KOVER, 2009 Standing genetic variation in *FRIGIDA* mediates experimental evolution of flowering time in *Arabidopsis*. Mol Ecol **18:** 2039-2049.

SCHEPENS, I., P. DUEK and C. FANKHAUSER, 2004 Phytochrome-mediated light signalling in *Arabidopsis*. Curr Opin Plant Biol **7:** 564-569.

SCHLUTER, D., K. B. MARCHINKO, R. D. BARRETT and S. M. ROGERS, Natural selection and the genetics of adaptation in threespine stickleback. Philos Trans R Soc Lond B Biol Sci **365:** 2479-2486.

SCHMID, M., T. S. DAVISON, S. R. HENZ, U. J. PAPE, M. DEMAR *et al.*, 2005 A gene expression map of *Arabidopsis thaliana* development. Nat Genet **37:** 501-506.

SCHMIDT, A., N. BAUMANN, A. SCHWARZKOPF, W. B. FROMMER and M. DESIMONE, 2006 Comparative studies on Ureide Permeases in *Arabidopsis thaliana* and analysis of two alternative splice variants of *AtUPS5*. Planta **224:** 1329-1340.

SHARROCK, R. A., and P. H. QUAIL, 1989 Novel phytochrome sequences in *Arabidopsis thaliana*: structure, evolution, and differential expression of a plant regulatory photoreceptor family. Genes Dev **3:** 1745-1757.

SHINDO, C., M. J. ARANZANA, C. LISTER, C. BAXTER, C. NICHOLLS *et al.*, 2005 Role of *FRIGIDA* and *FLOWERING LOCUS C* in determining variation in flowering time of *Arabidopsis*. Plant Physiol **138:** 1163-1173.

SHINDO, C., G. BERNASCONI and C. S. HARDTKE, 2007 Natural genetic variation in *Arabidopsis*: tools, traits and prospects for evolutionary ecology. Ann Bot (Lond) **99:** 1043-1054.

SHISHKIN, A. A., I. VOINEAGU, R. MATERA, N. CHERNG, B. T. CHERNET *et al.*, 2009 Large-scale expansions of Friedreich's ataxia GAA repeats in yeast. Mol Cell **35:** 82-92.

SIMPSON, G. G., and C. DEAN, 2002 *Arabidopsis*, the Rosetta stone of flowering time? Science **296:** 285-289.

SINDEN, R. R., V. N. POTAMAN, E. A. OUSSATCHEVA, C. E. PEARSON, Y. L. LYUBCHENKO *et al.*, 2002 Triplet repeat DNA structures and human genetic disease: dynamic mutations from dynamic DNA. J Biosci **27:** 53-65.

SINGH, P., L. ZHENG, V. CHAVEZ, J. QIU and B. SHEN, 2007 Concerted action of exonuclease and Gap-dependent endonuclease activities of FEN-1 contributes to the resolution of triplet repeat sequences (CTG)n- and (GAA)n-derived secondary structures formed during maturation of Okazaki fragments. J Biol Chem **282:** 3465-3477.

SIPIONE, S., and E. CATTANEO, 2001 Modeling Huntington's disease in cells, flies, and mice. Mol Neurobiol **23:** 21-51.

STINCHCOMBE, J. R., C. WEINIG, M. UNGERER, K. M. OLSEN, C. MAYS *et al.*, 2004 A latitudinal cline in flowering time in *Arabidopsis thaliana* modulated by the flowering time gene *FRIGIDA*. Proc Natl Acad Sci U S A **101:** 4712-4717.

SUNG, S., and R. M. AMASINO, 2004 Vernalization and epigenetics: how plants remember winter. Curr Opin Plant Biol **7:** 4-10.

SURESHKUMAR, S., M. TODESCO, K. SCHNEEBERGER, R. HARILAL, S. BALASUBRAMANIAN *et al.*, 2009 A Genetic Defect Caused by a Triplet Repeat Expansion in *Arabidopsis thaliana*. Science **323:** 1060-1063.

SWARUP, K., C. ALONSO-BLANCO, J. R. LYNN, S. D. MICHAELS, R. M. AMASINO *et al.*, 1999 Natural allelic variation identifies new genes in the *Arabidopsis* circadian system. Plant J **20:** 67-77.

TODESCO, M., S. BALASUBRAMANIAN, T. HU, B. TRAW, M. HORTON *et al.*, 2010 Natural allelic variation underlying a major fitness tradeoff in *Arabidopsis thaliana*. Nature **465:** 632-636.

TORJEK, O., H. WITUCKA-WALL, R. C. MEYER, M. VON KORFF, B. KUSTERER *et al.*, 2006 Segregation distortion in *Arabidopsis* C24/Col-0 and Col-0/C24 recombinant inbred line populations is due to reduced fertility caused by epistatic interaction of two loci. Theor Appl Genet **113:** 1551-1561.

VOLKER, J., G. E. PLUM, H. H. KLUMP and K. J. BRESLAUER, 2009 DNA repair and DNA triplet repeat expansion: the impact of abasic lesions on triplet repeat DNA energetics. J Am Chem Soc **131:** 9354-9360.

WANG, X., R. KORSTANJE, D. HIGGINS and B. PAIGEN, 2004 Haplotype analysis in multiple crosses to identify a QTL gene. Genome Res **14:** 1767-1772.

WARTHMANN, N., J. FITZ and D. WEIGEL, 2007 MSQT for choosing SNP assays from multiple DNA alignments. Bioinformatics **23:** 2784-2787.

WEINIG, C., M. C. UNGERER, L. A. DORN, N. C. KANE, Y. TOYONAGA *et al.*, 2002 Novel loci control variation in reproductive timing in *Arabidopsis thaliana* in natural environments. Genetics **162:** 1875-1884.

WELLS, R. D., R. DERE, M. L. HEBERT, M. NAPIERALA and L. S. SON, 2005 Advances in mechanisms of genetic instability related to hereditary neurological diseases. Nucleic Acids Res **33:** 3785-3798.

WERNER, J. D., J. O. BOREVITZ, N. H. UHLENHAUT, J. R. ECKER, J. CHORY *et al.*, 2005a *FRIGIDA*-independent variation in flowering time of natural *Arabidopsis thaliana* accessions. Genetics **170:** 1197-1207.

WERNER, J. D., J. O. BOREVITZ, N. WARTHMANN, G. T. TRAINER, J. R. ECKER *et al.*, 2005b Quantitative trait locus mapping and DNA array hybridization identify an *FLM* deletion as a cause for natural flowering-time variation. Proc Natl Acad Sci U S A **102:** 2460-2465.

WHITE, G. M., M. T. HAMBLIN and S. KRESOVICH, 2004 Molecular evolution of the phytochrome gene family in sorghum: changing rates of synonymous and replacement evolution. Mol Biol Evol **21:** 716-723.

WOLYN, D. J., J. O. BOREVITZ, O. LOUDET, C. SCHWARTZ, J. MALOOF *et al.*, 2004 Light-response quantitative trait loci identified with composite interval and eXtreme array mapping in *Arabidopsis thaliana.* Genetics **167:** 907-917.

YANG, S. H., P. H. CHENG, H. BANTA, K. PIOTROWSKA-NITSCHE, J. J. YANG *et al.*, 2008 Towards a transgenic model of Huntington's disease in a non-human primate. Nature **453:** 921-924.

ZHAO, K., M. J. ARANZANA, S. KIM, C. LISTER, C. SHINDO *et al.*, 2007 An *Arabidopsis* example of association mapping in structured samples. PLoS Genet **3:** e4.

ZYBAILOV, B., H. RUTSCHOW, G. FRISO, A. RUDELLA, O. EMANUELSSON *et al.*, 2008 Sorting signals, N-terminal modifications and abundance of the chloroplast proteome. PLoS ONE **3:** e1994.

Publications

The following publications resulted from the work presented in this thesis.

1. Todesco, M*., Balasubramanian, S*., Hu, TT., Traw, BM., Horton, M., Epple, P., Kuhns, C., **Sureshkumar, S**., Schwartz, CS., Lanz, C., Laitinen, RAE., Huang, Y., Chory, J., Lipka, V., Borevitz, JO., Dangl, JL., Bergelson, J., Nordborg, M and Weigel, D (2010) Natural allelic variation underlying a major fitness tradeoff in *Arabidopsis thaliana.* **Nature,** 465: 632-636

2. **Sureshkumar, S***., Todesco, M*., Schneeberger, K., Harilal, R., Balasubramanian, S[$] and Weigel, D[$] (2009) A genetic defect caused by a triplet repeat expansion in *Arabidopsis thaliana.* **Science,** 323 : 1060-1063 [* denotes equal contribution]

3. Schwartz, C*., Balasubramanian, S*., Warthmann, N., Michael, TP., Lempe, J., **Sureshkumar, S**., Kobayashi, Y., Maloof, JN., Borevitz, JO., Chory, J and Weigel, D (2009) Cis-regulatory changes at *FLOWERING LOCUS T* mediate natural variation in flowering responses of *Arabidopsis thaliana.* **Genetics,** 183:723-732. [* denotes equal contribution]

4. Balasubramanian, S., **Sureshkumar, S**., Lempe, J and Weigel, D (2006) Potent induction of *Arabidopsis thaliana* flowering by elevated growth temperature**. PLoS Genetics,** 2(7): e106.

5. Balasubramanian, S*., **Sureshkumar, S*.,** Agrawal, M., Micheals, T., Weissinger, C., Maloof, JN, Clark, R., Warthmann, N., Chory, J and Weigel D (2006) Allelic Variation at *PHYTOCHROME C* underlies natural variation in flowering time and light sensitivity in *Arabidopsis thaliana* **Nature Genetics,** 38: 711-715**.** [* denotes equal contribution]

6. Lempe, J., Balasubramanian, S., **Sureshkumar, S.,** Singh, A.,Schmid, M and Weigel, D (2005) Diversity of flowering responses in wild *Arabidopsis thaliana* strains. **PLoS Genetics,** 1(1): e6

Acknowledgements

I am deeply indebted to my mentor and my supervisor **Dr. Detlef Weigel** allowed me to start my scientific carrier in his lab. His adequate scientific knowledge and his valuable suggestions added more value to my journey of PhD. It has been known that being in research journey no eloped from hurdles and down" even in the toughest moments, his immense advice and support always been remembered and cherished.

I am profoundly grateful to **Dr. Klaus Harter** from, university of Tubingen, ZMBP for his, comments and valuable discussions.

I wish to record my thanks to **Dr. Sureshkumar Balasubaramanian**, happened to my husband. During the course of time he had taught me various aspects of the genetics and trouble-shooting methods, which made me independent researcher. His contribution and his enormous efforts are unaccountable; with out him I wouldn't have completed the mission of my PhD.

I would like to acknowledge all my collaborators and particularly **Oliver Bracko, Marco Todesco, Mitesh Agrawal,, Ramya Harilal, Janne Lempe and David Richter** for their contributions.

I would like to thank **Dr. Sureshkumar Balasubramanian** and **Dr. Greg Gibson**, University of Queensland, for their financial support, at final stages of my research.

I would like to thank **Frank, Josip** and **Christa lanz** for their Technical assistance and timely placing orders made the project to move continuously.

I would like to thank **Dr. Daniel Ortiz Barrientos** University of Queensland, for his critical discussions on manuscript. I would like to thank all **Weigel** and **SKB/OB** lab members for there help both in and out side of the lab.

I thank my Swiss and German friends, **JJ Pittet** and his wife **Elizabeth** and **Frau Kuhn** and **Herr Kuhn** for their valuable friendship and moral support. I thank my **Indian friends** residing in **Germany** and **Australia** entertaining our family in various ways.

I express my vows to **Suresh's family** for their love and support. I thank my **sister** and **brother** for their moral support.

I express my thanks to Tubingen and Australian **childcare centres** offered the place for my children, at right time. Without them I would not had finished my studies on time. I am very grateful to **Max-Planck Institute**, **University of Tubingen** and **University of Queensland (Australia)**, provided a right scientific atmosphere.

I dedicate this work to **my parents** and special thanks to my **Father**, who gave me a right foundation of education and deep affirmation during both good and bad moments. With out them I would not had reached the goal.

CURRICULUM VITAE

Name	**SRIDEVI SURESHKUMAR**
Date of Birth	05.05.1976
Present address	19 Ruskin street, Taringa, QLD 4068, Australia Email: s.sureshkumar@uq.edu.au

ACADEMIC RECORD

May2010- present	Senior Research Technician, The University of Queensland Australia
Aug 2008 - May 2010	Visiting Scholar, School of Biological Sciences, The University of Queensland Australia
May 2007 - Aug 2008	Carrier break (child birth)
Sep 2006 - May 2007	Phd Student, Weigel lab, Max-Planck Institute for Developmental Biology, Tuebingen
Mar 2004 - Sep 2006	Volunteer, Weigel lab, Max-Planck Institute for Developmental Biology, Tuebingen
Thesis title	Genetic and environmental modulation of phenotypic variation in *Arabidopsis thaliana*
Supervisors	Prof. Dr. Detlef Weigel Prof. Dr. Klaus Harter