INVESTIGATION OF THE WHEAT *PHI* LOCUS IN HETEROLOGOUS MODELS

RUOYU WEN

A THESIS

SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF PH.D.

SUPERVISED BY

PROFESSOR PETER SHAW AND PROFESSOR GRAHAM MOORE

CELL AND DEVELOPMENTAL BIOLOGY DEPARTMENT

JOHN INNES CENTRE

UNIVERSITY OF EAST ANGLIA

NORWICH, UNITED KINGDOM

SEP, 2011

To my mum, for her unconditional love and support.

Abstract

Ph1, the major locus in wheat to control homoeologous pairing, was recently mapped to an approximately 1 Mb region and a cluster of CDK like genes were identified as the *Ph1* candidate genes. It is hard to carry out molecular biological studies in wheat itself due to its large complicated genome and the relative lack of efficient research tools. Therefore the aim of my PhD project is to investigate the effects of *Ph1* candidate genes in heterologous model systems to first, assess and set up a model system for the further study of *Ph1* locus; and second to also get the first insights into the mechanism of *Ph1* locus.

In brief, the Ph1-CDKs were classified by phylogenetic analysis first, and it was found by the study that Ph1-CDKs belong to a specific kinase group that is related to but distinct from CDK. The effects of Ph1-CDKs were then assessed by ectopic expression in *Saccharomyces cerevisiae*, *Brachypodium distachyon* and *Arabidopsis thaliana* respectively under the drive of various promoters. It was found that Ph1-CDKs from A and D genomes can produce remarkable effects in *B. distachyon* and *A. thaliana* but not in *S. cerevisiae.* It was also found that the only transcribed full length B genome Ph1-CDKs (Ph1-B2) was surprisingly not a functional copy when being expressed in both *B. distachyon* and *A. thaliana*, although this Ph1-B2 was different from other Ph1-CDKs only at its C terminal end which is 30 aa in size.

In addition, a set of essential cytology research tools were developed in *B. distachyon* prior to the study of wheat *Ph1* locus in *Brachypodium*, which included the staging of meiosis, the identification of centromeric satellite repeat and the establishment of a high efficiency single site BAC clone screening system. The chromosome behaviours of *B. distachyon* during early stages of meiosis were described for the first time with the developed tools. It was found by the study that a novel centromere clustering event occurred at the onset of meiosis, in-between the status of premeiotic unpaired and pachytene paired centromeres. The developed research tools will provide a useful platform for the future study of *B. distachyon* meiosis. The finding of meiotic centromere cluster will provide useful information for our understanding towards meiotic chromosome behaviours.

Table of Contents

Acknowledgments

I would like to first of all, express my gratitude to my supervisors Prof. Peter Shaw and Graham Moore for offering me the opportunity to carry out my PhD study in a world class institute. This thesis can not be completed without their support and guidance. I would also like to thank my advisor Dr. Philippe Vain for his generous help and support during my study.

I want to thank everybody in Nuclear Structure Group and Moore group for helping me getting familiar with life in Britain and easing my studying in an international lab. Especially I want to thank Ali Pendle for helping me out always when there is a problem in the lab.

I want to especially thank Dr. Zheng Tao for generously sharing with me all his tricks and tips in molecular cloning, and for his passion and patience in discussion. I would also like to thank Dr. Colas Isabelle, Barbara Worland, and Adam Jarmuz for giving me the best training in FISH, Brachypodium and Yeast respectively. I also want to thank Dr. Melanie Febrer for providing me with all the BAC clones used in this study, and to thanks Dr. Luis Aragón for kindly hosting me in his group.

Finally, I want to thank all my friends in Norwich and London, who shared with me many memorable moments in trips, parties, photography and “Namek thinking”. Here I want to especially thank Dian, for her nice company, support, driving and many more during my last year of study. Without her, I may not be able to finish the writing of this thesis.

My PhD study was funded by the China Scholarship Council (CSC). I gratefully acknowledge the support.

List of Abbreviations

3D	three dimensional
A, C, G, T	Adenine, cytosine, guanine, thymine
AEP2	ATPase expression 2
BAC	bacteril artificial chromosome
BDM	Bateson-Dobzhansky-Muller
bp	base pairs
BSA	Bovine serum albumin
CCR	cereal centromeric retrotransposon
CAK	CDK-activating kinases
CCS1	cereal centromeric sequence 1
CDK	cyclin dependent kinase
CKI	CDK inhibitory subunits
CLSM	confocal laser scanning microscopy
CMS	cytoplasmic male sterility
DAPI	4',6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
DSB	double strand break
EMS	ethyl methanesulfonate
FISH	fluorescence *in situ* hybridization
FITC	fluorescein isothiocyanate
GFP	green fluorescent protein
GISH	genomic in situ hybridization
HI	hybrid incompatibility
Hmr	Hybrid male rescue
HP1	heterochromatin protein 1
HPA	histidinol phosphate aminotransferase
HPCG	homoeologous pairing control genes
hr	hour
HRI	heterochromatin related incompatibility

Kb	kilobase
KPP	Kip-related proteins
Lhr	Lethal hybrid rescue
LTR	long tandem repeat
Mb	megabase
min	minute
MEN	Mitosis Exit Network
mRNA	messenger RNA
MSH	mismatch repair genes
MTOC	microtubule-organizing center
NB-LRRs	nucleotide-binding leucine-rich repeats
NMI	nuclear-mitochondria incompatibility
OD	optical density
OdsH	Odysseus H
PBS	phosphate buffered saline
PCR	polymerase Chain Reaction
PDI	pathogen-defence incompatibility
Ph1	pairing homoeologous 1
Ph2	pairing homoeologous 2
PMC	pollen mother cell
PRDM9	PR domain containing 9
RENT	REgulator of Nucleolar silencing and Telophase
RNA	ribonucleic acid
rpm	revolutions per minute
RSDG	reciprocal silencing of duplicate genes
SC	synaptonemal complex
SDS	Sodium dodecyl sulfate
sec	second
SET	Lysine-4-methyltransferase
SNP	single nucleotide polymorphism
UV	ultra violet

List of Figures and Tables

CHAPTER I Introduction

1. Aim of research and thesis structure

Wheat is one of the earliest crops being domesticated back at Neolithic period (1). Wheat spread widely afterwards from Middle East to most parts of the world, and played a crucial role in human's transition from a hunter-gatherer society to a settled agricultural society. Nowadays, wheat is globally the number one cereal in amount consumed as human food (67.4 kg per caput at 2009/2010) and the second most produced cereal (682.6 million tons at 2009/2010) only after coarse grains. As a major food resource worldwide, the improvements of wheat yields and qualities do always draw great attention and have a profound impact on people's fighting against famine worldwide and the enhancement of global food security. Because of the hybridization-unfriendly flowers of wheat and the yet immature and unsecure molecular breeding methods, classical selection-after-crossing breeding strategies are still by far the most practical and successful method for wheat breeding. With the power of this method, a series of semi-dwarf, high-yield, disease-resistant wheat varieties were developed in the middle of the 20th century which increased the wheat yield by nearly 10 folds. These improved wheat varieties together with improved maize and rice have lead to the "green revolution" last century, which promoted the world from food deficiency to food surplus and saved billions of people from starvation.

To date, genetic resources used in classical wheat breeding are majorly limited to species within primary and secondary gene pools of wheat. *Ph1* is the major locus in wheat to control the homoeologous pairing and thus gene flow from tertiary gene pool to wheat. Genetic resources for classical wheat breeding can be further broadened by understanding the mechanism of *Ph1* locus and using it to transfer genes from tertiary gene pool of wheat. After more than a decade of efforts by colleges in our group (2-6), the *Ph1* locus was recently mapped to a approximately 1 Mb region and a cluster of CDK like genes were identified as the candidate genes that were responsible for the function of *Ph1* locus (5,6). Detailed molecular biological analysis was needed to understand the function and effect of these candidate genes. However it is very difficult to carry on molecular biological studies in wheat itself due to the large complicated wheat

genome and the lack of efficient essential research tools such as gene transformation method. As a result, the aim of my PhD project is to investigate the effect of *Ph1* candidate genes, Ph1-CDKs, in various model systems including yeast, brachypodium and Arabidopsis, to first find and set up a suitable model system for further study of Ph1-CDKs; and second to get the first insight into the mechanism of *Ph1* locus.

In the following sessions of Chapter I, wheat gene pools and challenges of gene flow from tertiary gene pool (species barriers) will be first discussed to give a broad view of my research subject. The effect of *Ph1* locus during wheat meiosis and the mapping of this locus will be summarised afterwards to give the up-to-date profile of the *Ph1* studies. In the next chapter, the classification and characterising of Ph1-CDK and related proteins will be discussed. Studies of *Ph1* locus in *Brachypodium*, Yeast and *Arabidopsis* are described in the following chapters respectively. Relevant results from the above chapters will be discussed together to suggest models explaining the origination and the mechanism of *Ph1* locus.

2. Species barriers and Ph1 locus

2.1 Wheat gene pools

Triticum, the genus that wheat belongs to, can be classified into three major groups based on their ploidy: diploid as einkorn (2n = 14), tetraploid as emmer (2n = 28), and hexaploid as dinkel (2n = 42). Bread wheat (*Triticum aestivum*) belongs to the dinkel type and durum wheat (*Triticum turgidum*) belongs to the emmer type. The genus *Aegilops* is also often considered as belonging to the *Triticum* tribe because of its similarity to other plants in *Triticum*. On the basis of molecular marker analysis, a model of wheat origination has been suggested: Diploid A genome ancestor (close to *Triticum urartu*) and B genome ancestor (close to *Aegilops* section *Sitopsis* species) formed hybrid first, and the genome of this hybrid was doubled to generate the Emmer type wheat (AABB). This is an automatic process which happened before the domestication of wheat. Then the Emmer type wheat was hybridized with D genome ancestor (close to *Aegilops tauschii*) and the genome doubled afterwards to generate modern bread wheat.

The acquisition of D genome is believed to be part of the wheat domestication process.

T. turgidum (AABB) and *A. tauschii* (D) constitute the primary gene pool of wheat, since transfers from these species can be made by direct cross with bread wheat (McFadden & Sears 1946, Gill & Raupp 1987). Bread wheat related species such as *T. monococcum* (A^m), *T. urartu* (A), *T. timopheevii* (AAGG) and polyploid *Aegilops* species constitute the secondary gene pool. Transfers from the secondary gene pool can be achieved by direct crosses and back crosses with varying levels of homologous recombination. All the rest *Triticeae* tribe species including rye and barley constitute the tertiary gene pool of wheat. These species are genetically isolated from wheat and cannot be crossed with wheat directly. To date, genetic resources used in classical wheat breeding are majorly limited to species within primary and secondary gene pools of wheat. Gene flow from tertiary gene pool to the target crop is in general prevented by species barriers, and the situation is even more complicated in polyploid crops.

2.2 Gene flow is impeded by species barriers.

Gene flow between species is in general prevented or at least hindered by species barriers. Depending on whether barrier is applied before or after the formation of zygotes, species barriers can be classified into prezygotic and postzygotic barriers. In case of prezygotic barriers, gametes from different species are prevented from forming zygotes by any of or combinations of isolations including spatial, temporal, mechanical, and behavioral barriers. The existence of different forms of prezygotic barriers eliminates crosses between remotely related species, and is considered as the major isolation force between species. For certain closely related species (species within the same genus normally or mostly at least same tribe), prezygotic barriers do not exist or can be bypassed by artificial crosses. Gametes from different species can merge and form zygote in this case, but the formed hybrids are either lethal, inviable, sterile (for intrinsic barriers) or with reduced fitness (for extrinsic barriers) because of the postzygotic barriers. Postzygotic barriers are crucial for speciation. The formation of postzygotic barriers, in most cases, is dependent on geographic separation of the same ancestral species. However, unlike geographic separation, postzygotic barriers are often the first irreversible isolation and the

established barriers will then be further consolidated by the accumulation of other postzygotic and/or prezygotic barriers and therefore draw special interests for evolution studies.

Although postzygotic barriers could be evolved from several different routes, the most common and well studied route is by hybrid incompatible loci that was first described by William Bateson (7), Theodosius Dobzhansky (8) and Hermann Muller (9) with the so called "Bateson-Dobzhansky-Muller" (BDM) model. This model suggests that hybrid incompatibility (HI) should be the consequence of combining at least two mutated loci together, which when appears alone is neutral for the host. The reason is that the mutated allele of a single locus must be heterozygous with its original allele at the very beginning generation, if this mutation allele is incompatible with its original copy, then the very individuals that host the mutation allele in the incompatible heterozygous format will simply not survive. Hybrid incompatible loci can be derived from different species lineages (spatially separate) or can be derived one after another in just one lineage at different time points (temporally separate). When incompatible loci meet both spatially and temporally in the hybrids, epistatic interactions between them will then induce dysfunctions to the hybrids and postzygotic barriers are established. Although the BDM model was proposed more than 60 years ago, the identification and characterisation of incompatible genes was achieved recently. The reported hybrids incompatible loci up to date do support the BDM model well, and mechanisms of hybrid dysfunctions caused by these loci can be categorized into four main types: gene loss by reciprocal silencing of duplicate genes (RSDG), nuclear-mitochondria incompatibility (NMI) dysfunction, pathogen-defence incompatibility (PDI) dysfunction and heterochromatin related incompatibility (HRI). Ref

For the RSDG type, if an essential gene happened to be located at different positions in parent species, then a portion (1/16 if they are totally unlinked) of F2 hybrid population will not survive since they lack a functional copy of this essential gene after segregation. A good example for this is histidinol phosphate aminotransferase (HPA) genes in A. thaliana. Due to a recent gene duplication event, there are two copies of HPA gene exist in the A. thaliana genome — HPA1 on chromosome 5 and HPA2 on chromosome 1 (10). In most ecotypes

only one copy of these two is functional, and the other is degenerated. As a result, hybrids among ecotypes that carry different functional copy of HPA suffer a genetic incompatibility since 1/16 of their F2 progeny lack functional HPA genes. Similar cases were also reported in rice (11) and Drosophila (12), where lost of mtRPL27 or JYalpha cause the HI in certain inter/intra-specific hybrids. Silencing of duplicate genes (SDG) is a common phenomenon found across various organisms, so HI caused by SDG as explained above is long been suggested as a frequent and common speciation force (13). However, it is worth mentioning that first, the effect of this incompatibility is very much dependent on the importance of the lost gene, in other words loss of many genes won't obviously reduce the fitness of their host. Second, quite often duplicated genes are redundant rather than silenced. Also, if SDG is tightly linked with its original (in case of tandem repeat gene for example), then segregation and lost of functional copy will be unlikely to happen. Finally, even in the most extreme case, only 1/16 of F2 progeny will suffer incompatibility, which could be overcome by segregation and selection in the following generations.

For the NMI type, HI is caused by combining certain mitochondrion with alien nuclear loci brought by alien male gametes. A phenomenon called cytoplasmic male sterility (CMS) identified in a wide range of plants (14) represents well the NMI HI. In a typical case of CMS, certain maternally inherited mitochondria genes will display either a male sterility or neutral effect depending on the genotypes of pollen parent. Hybrids with pollen parents that contain incompatible nuclear locus/loci with the sterile mitochondria gene could only serve as female parents in the population. Hybrids with pollen parents that are compatible with the potentially sterile mitochondria gene, or in context of breeding can restore fertility, are capable of producing both female and male gametes. The convenient manipulation of male fertility could greatly increase the hybridization efficiency, and thus make the CMS system attractive for breeding purpose (15). NMI HI is not necessarily limited in male sterility as it is in CMS. For instance in yeast, mitochondria located gene ATPase expression 2 (AEP2) from *S. bayanus* suffer a respiratory defect and sporulation failure in the *S. cerevisiae* genetic background (16). Interestingly, the effective CMS genes reported in plants to date are also mitochondria ATPase related genes, although ATPase is only one of the three main functions of a mitochondria genome. Mitochondrion genomes

have acquired selfish genomes whose genes compete with the host nuclear genomes and go through distinct evolution route, which may explain the enrichment of HI genes within such a small genome.

For the PDI type, HI is caused by plant autoimmunity - the presence of plant pathogen responding without the infection of pathogen. In plants, inter or intra-generic hybrids often suffer a so called “hybrid necrosis” phenotype that is characterized by slow growth, wilting, yellowing, and reduced viability and fertility (17). Not only does the fact that necrotic hybrid look similar to plants pathogens response, there are also other biochemical and molecular analyses from various genus suggesting that autoimmunity is involved in hybrid necrosis (18-20). This is also directly supported by a recent study with intra-generic hybrids of *A. thaliana* (17). According to this study, about 2% of intra-generic hybrids of *A. thaliana* did suffer some degree of hybrid necrosis. Further analysis showed that a nucleotide-binding leucine-rich repeats (NB-LRRs, a feature of plant immune receptors) containing protein DM1 on chromosome 5, when epistatic interact with specific allele of DM2 on chromosome 3, is necessary and sufficient for the induction of hybrid necrosis. Pathogen resistance genes are under selective pressure and thus are quickly evolved.

For the HRI type, HI is caused by the dysfunction of dysregulation of heterochromatin or heterochromatin binding proteins. To date, only a few cases of HRI were reported, and most of them were from the studies of *Drosophila*. For some cases, BDM pairs were formed by certain heterochromatin sequence and its binding protein. HI is directly induced by dysregulation of these heterochromatin in the affected hybrids during vital cell progresses. Odysseus H (OdsH) is an X-linked gene that induces male hybrid sterility between *D. simulans* and *D. mauritiana* (21). It was found recently that OdsH is a heterochromatin binding protein that interacts with repetitive sequences on the Y chromosome. When OdsH of both parents were co-expressed, the *D. mauritiana* OdsH bound to the *D. simulans* heterochromatin and disrupted its proper condensation, which strongly possibly lead to the male hybrid sterility directly (22). In another similar case, a locus called Zygotic hybrid rescue (Zhr) causes female hybrid lethality between *D. melanogaster* and *D. simulans*. Zhr itself is pericentromeric heterochromatin sequence on X chromosome. In the affected

hybrids, *D. melanogaster* Zhr sequence fails to be properly condensed during mitotic cell divisions in the presence of *D. simulans* cytoplasm, which leads to the embryo lethality of female hybrids (23). For other cases, HI is induced by interactions of certain heterochromatin binding proteins with other functional proteins. For example, HI between female *D. simulans* and male *D. melanogaster* is contributed by interaction of Lethal hybrid rescue (Lhr) with Hybrid male rescue (Hmr) as a BDM pair. The combination of wild type *D. melanogaster* Hmr allele and *D. simulans* Lhr allele in the hybrids results in male lethality (24). Hmr encodes a DNA-binding protein and Lhr encodes a protein-binding structure. Interestingly, LHR interacts with heterochromatin protein 1 (HP1) and is co-localized with HP1 in vivo in a heterochromatic region, which strongly suggests that LHR is involved in the heterochromatin regulation. Another similar case is reported in mice. Male hybrids between certain laboratory inbred and wild mice (*Mus musculus*). is sterile due to the xxx interaction of certain laboratory mouse Hst1 allelles (termed $Hst1^{s}$) with an unidentified locus termed Hst^{ws} from wild mice (25). It was uncovered recently by the same group that PR domain containing 9 (Prdm9, or named Meisetz) was the functional gene of the Hst1 locus (26). Prdm9 is a meiosis-specific gene essential for mouse reproduction (27). PRDM9 contains an N-terminal KRAB motif, a central histone H3 Lysine-4-methyltransferase (SET) domain, and a tandem array of C2H2 zinc fingers at the C-terminal. By comparing the sequence of sterile and fertile Prdm9 alleles, it was confirmed that changes at the C-terminal zinc fingers are responsible for the HI. Interestingly, unlike other domains of PRDM9 which are highly conserved among different species, the C-terminal C2H2 array are under exceedingly fast evolution. This rapidly evolving C2H2 array is confirmed by both bioinformatic (28) and biochemical (29)methods to recognize and bind an also rapidly evolving 13 bp degenerate nucleotide motif in both mice and human. This Prdm9 targeted motif is located in heterochromatic regions and is associated with 40% of human crossover hotspot (28)and more than 70% of mice DSB hotspot (30). Accelerated evolution of Prdm9 zinc fingers are also found among other diverse metazoans (31), which suggests that Prdm9 gene might have conserved function within metazoan taxa. Although it’s not yet clear how Prdm9 is involved in HI, it is suggested by the

above compelling evidences that meiotic recombination mediated via heterochromatin plays an important role.

2.3 Gene flow is completely blocked in polyploid by HPCG

The already complex and unsolved species barrier puzzle in between diploid species is even more complicated under the situation of polyploidy. Polyploids are cells or organisms that have more than two haploid sets of chromosomes (often as multiple homologous pairs in nature) in their genome. Polyploidy is seen commonly in plants, and also in some fishes and birds. Polyploidy is much rarer in other organisms, however, accumulating evidences from genomic studies of various organisms suggest that dynamic process of polyploidization and then depolyploidization is shaping most if not all eukaryotic genomes (32-35). Since many crops including bread and pasta wheat are polyploid, the study of species barriers under the context of polyploidy is intriguing from the aspect of evolution studies as well as crop breeding. Unlike diploids whose genome have two homologous chromosome sets, polyploids often contain similar but distinct (homoeologous) chromosome sets within one genome (for allopolyploid like wheat). In a way these homoeologous chromosome sets can be treated as different species and allopolyploid as the combination of two or more diploid species. The dilemma here is: from one hand, species barrier between donor diploid species has to be compromised or altered to an extent to allow them merge in one allopolyploid genome; from the other hand, the identities of combined diploid species have to be reinforced to maintain the genome stability of the formed allopolyploid. In detail, a diploid like meiosis process were observed in most if not all the allopolyploid plants and is considered to be the universal strategy for allopolyploid plants to secure a faithful segregation of chromosomes and thus the genome stabilization (). During such a meiosis process, chromosomes only pair as bivalents of two true homologous chromosomes to allow the balanced segregation of chromosomes afterward, rather than multivalents of several homoeologous chromosomes which would lead to uneven segregation and genome instability. The presence of homoeologous pairing control genes (HPCG) which eliminate homoeologous pairing and the formation of multivalents is widely considered as the mechanism to allow a diploid like meiosis in allopolyploids. Such HPCG were reported to be

present in a wide range of allopolyploid plants, however recent progress come mainly from the study of Brassica PrBn locus and wheat *Ph1* locus.

Brassica napus (AACC, 2n = 38) is a newly formed allotetraploid originating from natural hybridizations between ancestors of *Brassica rapa* (AA, 2n = 20) and *Brassica oleracea* (CC, 2n = 18) (36). The genomes of *B. rapa* and *B. oleracea* are highly similar and their chromosomes do pair and form bivalents in the inter-generic hybrids (37). By contrast, although haploids (AC, 19 chromosomes) produced from variety *Darmor-bzh* of *B. napus* did show similar pairing behaviour as the inter-generic hybrids, those produced from variety *Yudal* of *B. napus* did show much lower pairing, which indicate that homoeologous pairing was restricted in variety *Darmor-bzh* by HPCG (36,38). By using segregating population of haploids, a HPCG called Pairing regulator in B. napus (PrBn) was mapped on C genome within a 10-cM interval linkage group DY15 (39). It was demonstrated later that PrBn have an effect on meiotic homoeologous recombination. In the affected *B. napus* haploids, the frequency of crossovers between homoeologous chromosomes was reduced by PrBn during meiosis, although the crossovers distribution stayed more or less the same.

2.4 Wheat and the Ph1 locus

Ph loci are a series of loci that help to restrict the chromosome pairing to only true homologues in wheat. When Ph loci are deleted, various extents of homoeologous pairing can occur (40). *Ph1* is the major suppressor of homoeologous pairing, which is located on the long arm of chromosome 5B (41). A minor suppressor of homoeologous pairing has been found on the short arm of chromosome 3D, termed Ph2 (42). In the absence of Ph2 there is an increase in recombination rates between homoeologous chromosomes, especially between interspecific hybrids (42). Another minor suppressor has been found on the short arm of chromosome 3A (43). Absence of both the 3AS and 3DS chromosome arms resulted in homoeologous recombination rates as high as those seen in the *Ph1* deficient line.

Studies on the synaptic patterns of *Ph1*b and Ph2b mutants indicate that *Ph1* and Ph2 may have very different modes of action (44). A recent study shows that mismatch repair genes (MSH), MSH7 genes are located within the deletion

region of Ph2a line. This identifies TaMSH7 as a candidate for the Ph2 gene, and suggests that the function of Ph2 locus may involve the mismatch repairing mechanism as in bacteria (45), yeast (46), and mammals (47). In contrast, no MSH gene is found to be located within the deletion region of *Ph1* line, which indicates that mismatch repairing mechanism does not contribute to the function of *Ph1* locus.

The presence of *Ph1* in wheat was first displayed by Okamoto. It was shown by him that in the hybrids (AA^mBDD) between *T. aestivum* (AABBDD) and synthetic tetraploid wheat (A^mA^mDD), meiotic pairing was increased with the absence of chromosome 5B (48). This increase of meiotic pairing without chromosome 5B was also confirmed later in hybrids (AA^mBD) between *T. aestivum* and *T. monococcum* (A^mA^m) (49) and in haploid (ABD) of *T. aestivum* (50); Since only the loss of chromosome 5B but not other chromosomes showed a strong effect of promoting homoeologous pairing (50), it was suggested then that a HPCG that eliminate homoeologous pairing is located on chromosome 5B.

Homeologous pairing is greatly increased in wheat without the presence of *Ph1* locus. Moreover, the displacement of *Ph1* locus even induces a certain degree of chromosome pairing between wheat and rye chromosomes in wheat-rye hybrids. Therefore, this *Ph1* system has been used to transfer various external traits into wheat soon after its discovery. Although physical methods like X-ray irradiation has been successfully hired to induce gene flow from tertiary gene pool to certain crops (51), such kind of methods do have limited practical value because of their extremely low efficiency . In wheat, the major practical method to transfer genes from tertiary gene pool is by using *Ph1* locus.

The *Ph1* locus based wheat breeding protocol is featured by several rounds of crossings and back crossings to introduce the exogenous traits in the absence of the *Ph1* locus, and to recover the *Ph1* locus afterwards to avoid the potential genome instability caused by missing *Ph1* locus. In practise, crossing is an extraordinary laborious process that requires the input of massive amount of time, labour and resource. The use of *Ph1* locus in breeding is therefore much restricted. However, by understanding the mechanism of *Ph1* locus and manipulating its activity before and after exogenous cross, massive efforts can be saved since no back cross would be needed to recover the *Ph1* locus and

maintain the genome stability. Besides, the deletion of *Ph1* locus still can only induce external chromosome pairing to a limited extend in wheat only. By fully understanding the mechanism of *Ph1* locus, it is possible to extend the effect of this locus and transfer more exogenous genes/traits into wheat and possibly other polyploid crops.

3. Meiosis

3.1 General description of early meiosis events

One part of our effort in understanding *Ph1* locus has been focused on studying its effect on chromosome pairing in wheat and wheat-rye hybrids for the past decade. To better explain these effects, the meiotic prophase I with which chromosome pairing is happening will be described below.

The word "meiosis" comes from the Greek meioun, meaning "to make less". As defined by Cyril Dean Darlington, “meiosis is a method of nuclear division leading to an orderly reduction of the chromosome number. It is coextensive with sexual reproduction and in the sexual cycle it compensates for fertilization.” (52).

Although different organisms have different life cycles and sexual cycles, the main steps and mechanisms of meiosis for all the sexual reproducing organisms are in general the same, as briefly summarised below. Chromosomes are first duplicated at meiotic S phase to form sister chromatids and get ready for the process of meiosis. Axial elements begin to form on chromosomes at leptotene. Homologous axial elements then link together to form the lateral elements at early zygotene. Later, central elements are formed between the paired lateral elements by a process called synapsis, by which the structure formed is called synaptonemal complex (SC). Synapsed chromosomes then begin to cross over at pachytene. After the exchange of pairing partner, the synaptonemal complex breaks down, and the two homologous chromosomes start to separate but remain linked by structures called chiasmata at diplotene. Homologous chromosomes separate during diakinesis and Meiosis Ⅰ. Finally, sister chromatids separate during Meiosis Ⅱ, and the haploid gametes or spores are produced thereafter.

The whole meiosis process comprises two consecutive divisions called Meiosis Ⅰ and Meiosis Ⅱ. Each division contains one set of Prophase, Metaphase, Anaphase and Telophase like a mitotic division does. Based on the morphology of the chromosomes, Prophase I can be classified into sub stages from leptotene, zygotene, pachytene, diplotene to diakinesis in sequence. The chromatin is stretched out very thinly at leptotene, and chromatin from different chromosomes is mixed together and can not be distinguished separately under microscope. Chromosomes start to get thicker at zygotene. Synaptonemal complex can be seen in squashed nuclei at this stage. Chromosomes continue to get thicker and shorter at pachytene, Individual chromosomes can be distinguished at this stage. Later at diplotene, bivalent containing four chromatids can already be seen. Then at diakinesis, centromeres move away from each other, and the chromosomes remain joined by chiasmata structures, waiting to be aligned at the division plate later. Apart from the morphology traits, factors including nuclear size, number of nucleoli, the position of the nucleolus/nucleoli (53) and the occurrence of the telomere cluster (54) have also been used to help stage the meiosis process.

3.2 Chromosome recognition

One unresolved problem in chromosome pairing is how homologous chromosomes recognize and pair with each other. Chromosome morphology, specific sequence distribution, and proteins bound to DNA all may contribute to chromosome homology recognition, but the molecular mechanism remains to be established

Data from yeast to plants (55) are consistent with a model in which the telomeres attach to the nuclear envelope at random, and then cluster by an active process. A simple hypothesis proposes that the polarized movement of telomeres is driven by the cytoskeleton. As the nuclear envelope functions as a microtubule-organizing center (MTOC) in plants, it is possible that nuclear-envelope-associated microtubules are involved in telomere movement. Although pairing and bouquet formation are mutually independent, the clustering of telomeres is one of several possible mechanisms that may facilitate the initial homology recognition

In all the diploid organisms studied, telomeres always start to cluster at the onset of zygotene, which has led to the suggestion that bouquet formation directly facilitates homologous chromosome pairing by bringing the ends of chromosomes together. However in wheat, telomere start to cluster at the onset of leptotene (54), the sub-stage before chromosome pairing occurs, suggest that telomere clustering in meiosis may have alternative roles in chromosome pairing.

Although the occurring timing of the telomere clustering may be different between diploid organisms and polyploid organisms, the sub-telomere region of both organisms all undergo a similar chromatin conformation change process: the elongation of the sub-telomere region before clustering (56,57). It has been shown in wheat that this chromatin conformation change process is required for chromosome pairing (58).

Unlike telomere clustering which is seen in almost all the organisms studied, pre-meiotic centromere association has only been reported in wheat (44) and budding yeast (59). By counting the centromere loci in ndj mutant and zip1 mutant separately, Tsubouchi and Roeder found that the centromere association in budding yeast is independent of telomere clustering and recombination, but depends on the synaptonemal complex component Zip1. In both wheat and budding yeast meiocytes, centromere associations are initially nonhomologous and then undergo switching until all couples involve homologues. In yeast, this transition depends on Spo11, a protein required for the initiation of meiotic recombination. In wheat, this transition is achieved by alignment of all the homoeologous chromosomes, and depends on the *Ph1* locus (60). In yeast, staining of Zip1 protein shows that synaptonemal complex is assembled from centromere. This suggests that homologous centromeres may serve as sites of synapsis initiation in yeast.

3.3 Synapsis

The process of synapsis is often included as part of the chromosome pairing process. However various evidences suggest that it’s independent from homologous pairing. For instance, in haploid yeast and plant cells, which obviously lack homologous chromosome pairs, extensive synapsis has been

documented (61,62). On the other hand, in yeast mutants that cannot assemble synaptonemal complexes, chromosomes can still pair to some extent (63-65). In polyploids, all homologs ‘pair’, but then ‘synapse’ in twos (60). Simply put, pairing is an interaction between homologous chromosomes, which is based on homology recognition. Synapsis is a process of cementing an association of two chromosomes by installation of the SCs. Depletion of sister chromatid cohesion protein induces defective SC installation and absence of chiasmata (66), which shows that SC formation and chiasmata maintenance depends on the presence of sister chromatid cohesion

The yeast Hop1 protein is a component of the transverse filaments and is involved in the formation of lateral elements. Its reported equivalent proteins include Arabidopsis ASY1, rice PAIR2, and Maize AFD1. In Arabidopsis, foci of Asy1 protein appear at the onset of meiosis and then extend or fuse through chromosome condensation to form continuous linear signals by the end of leptotene. These tracts are maintained until the SC is dismantled at diplotene and diakinesis (67). Similar immunostaining pattern are also found in Brassica oleracea, Zea mays, Rye and Yeast. In Arabidopsis, Brassica oleracea and rye, the ASY1 signal is maintained longer than Hop1 signal in Yeast. ASY1 signal is lost when the chromosomes desynapse, In contrast, the HOP1 signal disappears at pachytene when full synapsis is completed. Immunogold labelling of ASY1 in Arabidopsis revealed a discontinuous pattern along the axial/lateral elements. In rye, the tracts of Asy1 protein are discontinuous at leptotene and zygotene. These observations could suggest a role of ASY1 in recruiting the bases of chromatin loops to the developing axial/lateral elements. In maize, ASY1 is released from the chromosomes when chromosomes synapse, whereas AFD1is maintained.

Arabidopsis AtZYP1 protein, a putative orthologue of yeast Zip1, mammalian SCP1 and D.melanogaster C(3)G, is the only reported protein to date that forms the SC transversal filaments in plants. However, analysis of protein homolog shows two highly homologous protein of AtZYP1 in *O. sativa* genome (68). The direct application of AtZYP1 antibody on rye meiocytes also shows a specific labeling on the central elements region (69). These results indicate that cereal crops carry orthologues of ZYP1. AtZYP1a and AtZYP1b are functionally

redundant proteins that are present in meiocytes during prophase I only. Double immunolocalizations demonstrated a central localization of AtZYP1, bordered by ASY1 on both lateral elements of the SC (70). It was also showed that the initiation of recombination is necessary for AtZYP1 recruitment but is not sufficient for its polymerization in a central element. A T-DNA insertion mutant of AtZYP1a and AtZYP1b resulted in delayed meiosis, absence of both pairing and synapsis in most meiocytes. However, recombination was only slightly reduced in the absence of AtZYP1, which indicates that SC is not required for the process of recombination. In the absence of AtZYP1, recombination occurs between both homologous and nonhomologous chromosomes, which suggest that SC is required to ensure the specificity of homologous chromosome associations.

3.4 Recombination

Recombination is initiated by double-strand breaks (DSBs). In somatic cells, DSBs can result from stresses such as ionizing radiation or chemicals on the chromosomes. They can also be produced when DNA replication forks encounter DNA single-strand breaks or other types of lesion (66). However, in meiosis process of most if not all organisms, DSBs are generated by enzyme Spo11 at a much higher rate than those generated in somatic cells.

The Spo11 protein is a eukaryotic homologue of the archaeal DNA topoisomerase VIA subunit (topo VIA). In archaea it is involved, together with its B subunit (topo VIB), in DNA replication. However, most eukaryotes, including yeasts, insects and vertebrates, instead have a single gene for Spo11/topo VIA and no homologues for topo VIB. In these organisms, Spo11 mediates DNA double-strand breaks that initiate meiotic recombination. In contrast to what is known from other eukaryotes, Arabidopsis thaliana carries in its genome three SPO11 homologues. The homologues in Arabidopsis, AtSPO11-1, AtSPO11-2 and AtSPO11-3, all share 20–30% sequence similarity with other Spo11/topo VIA proteins (71). Previous genetic evidence suggests that AtSPO11-1 is a true orthologue of Spo11 in other eukaryotes and is required for meiotic recombination, whereas AtSPO11-3 is involved in DNA endo-reduplication.

Isolation of SPO11 homologues in A. thaliana, mouse, man, D. melanogaster, and C. elegans have shown that DNA replication is needed for DSBs (72), which suggests that there is a safety check mechanism to ensure that breakage is not induced before sister chromatids are available for repair, in case corresponding sequences on homologous chromosomes cannot be found.

In yeast, mouse and Arabidopsis, DSBs can exist without SC formation, but SC formation cannot exist without DSBs, indicating that DSBs are required for homology search and SC formation in yeast. However, in D. melanogaster and C. elegans pairing centres initiate synapsis first, and then DSBs are formed, indicating that DSBs are not needed for SC formation in these species.

In yeast, the initial DSBs are resected from 5’ to 3’ by the MRX complex, which is composed of Mre11, Rad50 and Xrs2/Nbs1. Arabidopsis homologs of Rad50 (AtRAD50) and Mre11 (AtMRE11) have been shown to form a complex, and functional analysis of the atrad50 and atmre11 mutants strongly suggest that plants have a functional homolog of the MRX complex (73,74).

The single-stranded DNA ends produced by the SPO11/MRX complex invade the homologous double-stranded DNA by the action of RAD51 and DMC1, which are the homologs of the RecA recombinase (75). BRCA2 is involved in the loading of RAD51 on single-strand DNA (76). Reducing BRCA2 expression in Arabidopsis meiocytes leads to chromosome fragmentation and univalents, a phenotype that is similar to that seen in Atrad51, Atrad51c, or Atxrcc3 mutant. The chromosome fragmentation phenotype is suppressed by the absence of AtSPO11-1 function, which demonstrates that BRCA2 acts downstream of SPO11-1 in Arabidopsis (77). In yeast, DMC1 is a meiotic specific protein, and RAD51 is involved in both mitotic and meiotic recombination (78). In contrast, the Arabidopsis DMC1 is expressed in both reproductive tissues and leaves (79). DMC1 and RAD51 both have the ability to catalyze the strand exchange reaction in vitro (80,81). In vitro experiments also display that the activity of human DMC1 requires ATP and is strongly dependent on the heterotrimeric ssDNA-binding molecule replication factor A (RPA) (82). The colocalization of DMC1 and RAD51 displayed by immunostaining in yeast suggest that they may interact to form a complex before the installation of SC (83).

Rad51 proteins were found to form discrete nuclear foci from early zygotene to pachytene, to co-localize with lateral element proteins in yeast, mouse, rat, human, and lily meiosis and are components of early recombination nodules. Rad51 has also been proposed to have a role in the homology search phase of chromosome pairing (84). RAD51 foci are localized near the DAPI positive sub-telomeric heterochromatin in early meiotic prophase nuclei of rye, which indicates that the initiation of recombination and DNA sequence homology testing is concentrated near the ends of the rye chromosomes.

After the strand invasion, DNA re-synthesis based on the invading strand and then ligates to form double-Holliday junctions (85). Resolution of the double-Holliday junction following cutting at alternative strands results in products with either crossovers or noncrossovers. X-ray cross complementing (Xrcc), a paralog of Rad51, has been shown to participates in Holliday junction resolution in vertebrate cells.

Recombination is the mechanism by which the DNA strand breaks are repaired using homologous chromatids as template. Recombination could happen during mitotic DNA duplication or during meiotic chromosome pairing where DSB occurs by accident or by program. The progress and key components of mitotic and meiotic recombination are very similar, apart from their template preference. Mitotic recombination tends to choose sister chromatid (inter-sister, IS) and meiotic recombination tends to choose homolog (inter-homolog, IH) as templates to repair the DSBs. The IH bias of meiotic recombination is established and maitained by several interdependent processes. The SPO11 catalyzed DSB is loaded and lead by a meiotic specific recombinase DMC1. Unlike its mitotic counterpart Rad51, DMC1 prefer to promote the invasion of broken ends into homologs rather than sister-chromatids. Deletion of only Dmc1 will cause the cells to arrest at pachytene since DSBs are not repaired. While deletion of Dmc1 together with Red1, Hop1 or will inducing efficient repair of DSBs via IS bias recombination. , u and a mechanism to establish and maintain mitotic DNA duplication, recombination could be used to repair strand breaks Hop1phosphorylation in response to DSBs triggers dimerization of Mek1 via the Hop1 C domain, thereby enablingMek1 to phosphorylate target proteins that prevent repair of DSBs by sister chromatid

4. Early meiosis events in wheat and the effect of Ph1 locus

4.1 Centromeres associate in non-homologous pairs in the floral tissues of wheat

By GISH labelling of the barley chromosome pair in a wheat substitution line, it was shown that homologous chromosomes associate pre-meiotically in the meiocytes as well as in the surrounding somatic cells (86). Such homologous associations only happened before the onset of meiosis in the presence of *Ph1* locus and were missed in the absence of *Ph1* locus. The pre-meiotic *Ph1* dependent association of homologous chromosome were confirmed later by GISH labelling of the rye chromosome arm pair in a 3RS wheat addition line (54). It was also found by the same study that, centromeres were already paired long before meiosis in both the presence and absence of *Ph1* locus. Since the association of homologous chromosomes only occur shortly before the onset of meiosis and was dependent on the presence of *Ph1* locus (86), the centromere pairs spotted in the absence of *Ph1* must be formed between non-homologous centromeres. Or in other words, in the absence of *Ph1*, centromeres paired randomly. Later by comparing the centromere behaviour in both diploid and polyploid wheat lines, it was found that the association of centromeres is induced by polyploidy. That is, centromeres were found to be in pairs in the floral tissues of all the polyploids including newly synthesized polyploids but not in any of the investigated diploid wheat (87).

The centromere behaviour analysis was carried out later in wheat-rye hybrids that contained only homoeologous chromosomes and no true homologous chromosomes (44). It was found by the study that, non-homologous centromere associations were reduced by the presence of *Ph1* during the progress of meiosis in the wheat-rye hybrids. By contrast, such non homologous associations of centromeres were increased in the absence of *Ph1*. It was shown by these results that, centromeres were maintained as non-homologous pairs before meiosis in both the presence and absence of *Ph1*. In the presence of *Ph1*, the random centromere associations were gradually transferred into

homologous pairs during the transit of pre-meiosis to meiosis. However, this non-homologous to homologous transition would only occur when there were true homologues presented. In the absence of *Ph1*, random centromere associations were un-affected during the pre-meiosis to meiosis transit, no matter if true homologues were present or not. Interestingly it was demonstrated that, there is a “bridge” step linking the random to homologous centromere association status. At this “bridge” step that was just before the onset of meiosis, centromeres in wheat or wheat-rye hybrids, in the presence or absence of *Ph1* locus, all clustered into 7 groups (60). An obvious logic was that centromeres found their right partners by rallying all the related centromeres together, and then sorted it out during the “meeting” before the start of meiosis. However, a few results from later studies showed that, the transition of random-to-homologous centromere association, or in fact the pairing and synapsis of the whole chromosomes, was not dependent on the centromeric regions, but rather on the telomeric regions (88,89). By using a series of asymmetrical isochromosomes to access which segments of a chromosome arm is responsible for the meiotic chromosome pairing initiation, it was shown that the pairing of chromosome was affected by the loss of homology in even relatively short distal regions of chromosome arms (88). In a more delicate study, wheat chromosome arms and centromeres of 1B were replaced separately by rye equivalent components. It was shown that the combination of homologous arms with heterologous centromeres didn’t affect the correct pairing. By contrast, the combination of heterologous arms with homologous centromeres failed to pair (89).

4.2 Ph1 locus suppresses homoeologous pairing by preventing chromatin decondensation.

It was noticed that massive chromatin conformation changes occurred around the onset of meiosis especially in heterochromatin in various species (56,90,91). A detailed analysis of wheat substitution lines that carried a homologous pair of distal rye heterochromatin fragments on chromosome 1D showed that chromatin decondensation of the telomeric rye heterochromatin occurred simultaneously before the form of telomere bouquet. The synchronised telomeric chromatin decondensation was dependent on *Ph1* locus. In the absence of *Ph1*, the two

homologous rye heterochromatin fragments were not decondensed at the same time (57). It was further demonstrated by placing and labelling pairs of identical, similar or distinct rye fragments in the wheat substitution lines in the presence of *Ph1* locus that, the synchronised chromatin decondensation was related to the telomeric sequence similarity level (92). Identical telomeric rye heterochromatin decondensed to the same extent at the same time before telomere bouquet, and then joined at the telomeric ends and "zipped up" afterwards to form the SC (Synapsis Complex) structure. In this case, they were paired in most all the meiocytes at metaphase I. Similar rye heterochromatin decondensed at the same time but to differential extent, in which case they paired in about half of the meiocytes at metaphase I. Chromatin conformation change didn't occur with distinct rye heterochromatin. In this case, they only paired in a small portion of the meiocytes at metaphase I. In another experiment labelling all 7 rye telomeric heterochromatin regions in the wheat-rye hybrids in the presence or absence of *Ph1* locus, it was shown that: in the presence of *Ph1*, chromatin conformation change didn't occur in rye telomeric heterochromatin regions of different rye chromosomes. By contrast in the absence of *Ph1*, telomeric heterochromatin regions of all 7 rye chromosomes decondensed and clustered together. The results showed that the synchronised telomeric heterochromatin conformation changes around the onset of meiosis were crucial for the pairing of homologues. In the presence of *Ph1* locus, the synchronised conformation changes, and the synapsis as a consequence, were restricted to true homologous chromosomes. In this case, homoeologous chromosomes can not pair because of the suppressed chromatin conformation changes. While in the absence of *Ph1* locus, the suppressed synchronised chromatin conformation changes among non-homologous chromosomes were released and homoeologous pairing happened as consequence.

5. Ph1 mapping

5.1 Ph1 locus has been mapped to a region less than 2.5 Mb

Half a century has past since the discovery of the *Ph1* locus, however cloning of *Ph1* gene/genes only progressed slowly before. Since natural phenotypic variation does not exist and also the attempt to create single gene mutation by

ethyl methanesulfonate (EMS) treatment had failed (93), segregation populations can not be produced and therefore traditional positional cloning method is not feasible for *Ph1* cloning. So far, the only practical protocol of *Ph1* mapping is achieved by variant radioactive irradiation treatments which can generate *Ph1* mutants by inducing DNA deletions within the certain region of chromosome 5B. The first attempt of this kind is done by Sears in TA cv Chinese spring. After hybridising X-ray irradiated wheat pollen with normal wheat and selection in the following generations, a line called *Ph1*b had been identified and this *Ph1*b deletion line has been served as the most commonly used *Ph1* mutant line afterwards until now (94). Shortly later, similar *Ph1* mutant line, named *Ph1*c, generated also by X-ray irradiation had been abtained with durum wheat (95). It was reported later that *Ph1*b and *Ph1*c carry deletions overlapping within the long arm of chromosome 5B, which is estimated as about 70Mb in size.

Still hundreds of genes are contained within this 70Mb region. To further delimit *Ph1* locus to a smaller workable region, a two step strategy has been applies in our group: the first is to develop a marker system for the *Ph1* region, with the assistance of well annotated model cereal genomes based on cereal genome collinearity. The second is to generate wheat deletion library using alternative irradiation that can produce smaller deletions than what X-ray do, and then identify new *Ph1* deletions lines using the marker system developed in the first step. Following this strategy, a small region on rice chromosome 9 approximately 4 Mb in size was identified as equivalent to the 70 Mb *Ph1* region (2). Fast-neutron irradiation generated wheat lines was scored by a PCR based method (3). Five overlapping deletion lines within the *Ph1* region were identified and confirmed as *Ph1* deletion lines by showing an increased homoeologous pairing (4). The deletions of these five lines are clustered within a 250 kb region on rice chromosome 9. Break points of these lines have further delimited the *Ph1* to a 140 kb region in rice, which is equivalent to a 2.5 Mb region in wheat, and has been confirmed by another group recently with rice BACs and a set of *Ph1* deletion lines (96). Rice markers from this region were used to identify equivalent Brachypodium BACs and markers. These Brachypodium makers were then used to screen wheat BAC libraries. Positive BACs within the *Ph1* region were identified and assembled to give contigs that cover the *Ph1* region. A set of BACs were chosen to be sequenced to give the full sequence of the 2.5

Mb *Ph1* region, and approximately 36 genes were found within this region (5). High intense gamma irradiation was applied to generate lines that carry smaller deletions within this region. One such line which carried a 500 kb deletion was identified and survived. Since this line did show normal pairing pattern like wild type wheat, 8 genes contained within its 500 kb deletion were eliminated to be the possible *Ph1* candidates.

5.2 CDK like genes were suggested as Ph1 candidates.

For the rest of the genes within the defined *Ph1* locus, analysis of the gene transcription profiles showed that all the other genes were expressed mainly from A and/or D genome in the presence of *Ph1* except for genes from the CDK (Cyclin Dependent Kinase) like gene cluster region, whose expression were mainly from B genome (5,6). For the ease of identification, this CDK like gene cluster were termed as Ph1-CDKs; and individual Ph1-CDK was termed as *Ph1* followed by genome letter and cluster position (e.g. Ph1-A3, Ph1-B2 etc.) there after in this thesis. Since the *Ph1* locus is presented on chromosome 5B, its candidate genes are unlikely expressed from chromosome 5A or 5D rather than 5B itself, which indicate that the Ph1-CDKs region is where the *Ph1* candidate genes locate. This is also consistent with the above fact that the deletion of some of the genes outside the Ph1-CDKs region did not introduce a *Ph1* effect. This Ph1-CDKs region contains the Ph1-CDKs, the storage protein gene clusters and the *hyp* genes. Since *hyp3* was not found in the equivalent region of the related grass *T. timopheevii* (AAGG) which has *Ph1* activity, and *hyp1* and *hyp2* are known to encode storage-like proteins in wheat (97), the Ph1-CDKs located within this region were identified as the *Ph1* candidate genes. The transcription contribution of each CDK-like gene was measured by sequencing RT–PCR products amplified with conserved primers. In details, in the presence of *Ph1* locus, five Ph1-CDKs, namely Ph1-B2, B3, B4, B6 and B7, were expressed from the 5B region, together with Ph1-A3 expressed from the 5A region. While in the absence of *Ph1* locus, mainly Ph1-D2 and Ph1-A3 variants were expressed from the 5D and 5A region respectively. Thus the transcription of 5A and 5D Ph1-CDKs were being suppressed by the presence of the 5B copies. When the 5B *Ph1* region was deleted, the transcriptions of Ph1-CDKs were activated on the other genomes to compensate the loss of 5B copies (6).

5.3 Characters and regulations of CDK kinases

CDK is a group of kinases that are well conserved among all the eukaryotes and are the engine of the cell cycle events. CDK kinases belong to the CMGC super family of serine/threonine kinases (STK). Like other STK, CDK kinases contain all the conserved kinase subdomains (for details, see Chapter II), and have conserved kinase structure. The unique character of CDK is its association with cyclins. Therefore the cycline binding motif PSTAIRE, which is located at the beginning of the subdomain III, is considered as the signature of CDK.

The first CDK identified is the *Schizosaccharomyces pombe* Cdc2 protein that is essential for the progress of cell cycle. In yeast, there is only one CDK kinase in each genome (Cdc2 in *S. pombe* and Cdc28 in *S. cerevisiae*). This single CDK kinase is participated in the regulation of different stages of cell cycle by the association of mutiple stage-specific cyclins . More copies of Cdc2 related kinases were discovered in genomes of higher eukaryotes (CDK1-11 in mammals and CDKA-G in higher plants (98)) by phylogenetic analysis. Most of the CDK kinases in higher eukaryotes were known to be participated in the regulation of cell cycle . However, among several copies of CDK kinases in higher eukaryotes, normally only a single CDK plays the key role and is absolutely required for the progress of cell cycle (for instance, Cdk1 in mammals (99) and CDKA in higher plants (100)). Such Cdc2 related, cyclin associated and cell cycle core component proteins are also found in all the other eukaryotes studied, which are defined as core CDK kinases to be distinguished from CDK that plays only auxiliary role or even non-cell cycle related role (CDK2-11 in mammals and CDKB-G for instance).

There are several common mechanisms to increase or decrease the activities of CDK kinases. The first and most important one is the binding of cyclins to activate CDK. The binding of CDK and cyclin is mainly provided by the interactions between the PSTAIRE like helices in CDK and helices in the cyclin box of cyclins. For some CDK-cyclin pairs, CDK kinases and cyclins interact independently with high affinity. For some other CDK-Cyclin pairs, the proper interactions between CDK kinases and cyclins depend on additional components (e.g. binding of Cdk7-cyclin H promoted by Mat1) or modifications at certain sites (tight human Cdk1-cyclin A binding requires phosphorylation at

the activating threonine residue). The binding of cyclins can change the conformation of CDK kinases, and thus to expose the otherwise blocked catalytic cleft to allow the easy access of substrates. It was reported that, by this way the binding of cyclin A to human Cdk2 increase the activity of Cdk2 by several orders of magnitude independently (101).

The phosphorylation of CDK kinases at a conserved threonine (normally the corresponding sites of human Cdk2 Thr160 on T-loop) is another common mechanism to increase the activities of CDK kinases. Most CDK kinases require this phosphorylation event in addition to cyclin binding to achieve full kinase activity. It was reported that the phosphorylation of Thr160 can change the shape of the substrate binding site of Cdk2 in the Cdk2-cyclin A complex, and thus to improve the kinase-substrate binding and induce an additional 80-300 folds of activity increase (102). The activation threonine is phosphorylated by kinases named CDK-activating kinases (CAK). Interestingly, in higher eukaryotes like mammals and higher plants, CAK themselves are also considered as CDK kinases (Cdk7 in mammals; CDKD and CDKF in high plants). Plant CDKD kinases are closely related to the mammal Cdk7. Like Cdk7, CDKD kinases interact with cyclin (cyclin H) and directly phosphorylate the core CDK (CDKA kinases) (103). CDKF kinases are plant specific CAF. It was found that, rather than phosphorylate and activate the core CDK, the function of CDKF kinases is to phosphorylate and stabilize CDKD kinases (104,105). They can be somehow treated as the activating kinases of CAK.

The most common mechanism to suppress the CDK activity is done by the binding of CDK inhibitory subunits (CKI). Three CKIs (Pho81, Far1, and Sic1) have been identified in *S. cerevisiae*. Each of these CKIs binds to different CDK-cyclin complex and regulates different aspects of cell progress (106). In mammals, CKIs have been divided into the INK4 and the Kip/Cip class (107). Members of the INK4 class bind and inhibit CDK4 and CDK6 by impairing cyclin binding (in case of monomeric CDK) or causing the dissociation of the CDK-cyclin complex, and are structurally similar to the Pho81 of *S. cerevisiae*. In contrast, members of the Kip/Cip class bind and inhibit a broader range of CDKs and function in dimeric as well as heterotrimeric complexes with CDKs and cyclins. In plants, the so far identified CKIs are close related to the mammalian

Kip/Cip class inhibitors and are named as Kip-related proteins (KRP) (108). Biochemical evidences showed that plant KRP proteins bind A-type, but not B-type CDK, and the binding is directed by D-type or A-type cyclins (108). CKI inhibitors often function as switchers to allow the buildup and quick release of active CDK-cyclin complex.

P-loop phosphorylation is another mechanism to suppress the activity of CDK kinases. In mammals, the P-loop phosphorylation is mainly controlled by kinase Wee1 and phosphatases of Cdc25 family. Prior to mitosis, Cdc2-cyclin B complexes are held in the inactive state by P-loop phosphorylation done by Wee1. The P-loop will be abruptly dephosphorylated by Cdc25 at the end of G2 to release Cdc2 activation and mitosis. In case of DNA damage, the P-loop phosphrylation will be maintained to block the initiation of mitosis and allow the process of DNA repair. Homologues of Wee1 were identified in plants as well, and it was showed that plant Wee1 also inhibits mitosis in cells that suffer DNA damage (109). However, it was reported that plant Wee1 do not block mitosis by phosphorylate the P-loop of plant core CDK (110).

In all eukaryotes, cell cycle starts at G1 phase, followed by S phase, G2 phase and finally Mitosis. DNA is duplicated at S phase, and duplicated chromatin is divided into two daughter cells by mitosis. After mitosis, daughter cells enter G1 phase and cell cycle starts over again. Core CDK kinases are presented throughout the cell cycle at fairly constant levels. The progress of cell cycle is pushed by the rapid and irreversible switches of the core CDK activities during the transitions of different cell cycle stages, which are results of combined regulatory mechanisms mentioned above. Take *S. cerevisiae* as an example: Gln3 is a constantly transcribed cyclin (111). At late G1, Cdc28-Gln3 complex induce the transcription of both G1 cyclins (Gln2 and Gln1) and S cyclins (Glb5 and Glb6) via transcriptional factor SBF and MBF respectively (112,113). Cdc28-S cyclin complexes fire the replication of DNA (114) and also increase the level of M cyclin (Glb2) by phosphorylating Cdh1 (115), the activator of APC-dependent proteolysis. The transcription of Glb2 is controlled by the Mcm1-SFF transcription factor complex that can be activated by the Cdc28-Glb2 complex (113). The positive feedback loop of Glb2 leads to the rapid increase of Glb2 level, which is essential for the initiation of mitosis. At the end of mitosis, a

signaling network named Mitosis Exit Network (MEN) release Cdc14 from the REgulator of Nucleolar silencing and Telophase (RENT) complex (116). Cdc14 is a dual-specificity phosphatase that dephosphorylates and activates a set of proteins (e.g. Swi5, Cdh1 and Sic1) that work together to cease the activity of core CDK and enable the cell cycle to exit from mitosis (117). Cell cycle are regulated in a similar way as in *S. cerevisiae*, although with more core CDK kinases and more layers of positive and negative regulations (for details, see (106)).

CHAPTER II Bioinformatic analysis of Ph1-CDK genes

1. Background

In the previous mapping of the *Ph1* locus, it was found that the *Ph1* candidate genes, Ph1-CDKs, were only present in the genome of *Brachypodium sylvaticum* but not in the genome of rice (5). The putative brachypodium orthologues (simplified as orthologues for convenience here after) of Ph1-CDks were identified in the newly sequenced *Brachypodium distachyon* genome as bradi4g33220 and bradi4g33230 based on sequence similarity and genome synteny. They were therefore considered as the true orthologues of Ph1-CDKs. Putative genes with close sequence similarity level were also found in other regions of *B. distachyon, Sorghum bicolor* and *Zea mays* genomes, although not in the *Ph1* syntenic regions. Interestingly, such close Ph1-CDK orthologs homologues were not found in the genome of *Oryza sativa.* Ph1-CDK was reported to be most closely related to CDK2 in human and CDKG in *A. thaliana* respectively. However, it was noticed that when using human CDK2 or *Arabidopsis thaliana* CDKG as the query sequences to BLAST against the newly sequenced *B. distachyon* genome sequence, brachypodium orthologs of Ph1-CDK ranked very low on the list with comparatively low similarity scores, which indicates that Ph1-CDK may not be the true orthologues of CDK in *B. distachyon.*

Genomic information for plant genomes has developed rapidly during the recent years. Many genomes have been sequenced and annotated to assist the study of grasses. In this chapter, the previously identified *Ph1* candidate genes were analysed with updated genomic information and bioinformatic tools to give a better understanding about the classification and characters of the *Ph1* candidate genes, Ph1-CDKs.

The classification of STK and more specifically CDKs was reported previously in both *A. thaliana* and *O. sativa* (118,119). To better understand the relationship of Ph1-CDKs with other kinases, a phylogenetic analysis was performed with Ph1-CDKs and other plant kinases. It was found by this analysis that Ph1-CDKs

and their homologues belong to a special undescribed kinase group (termed as PH kinase group in this thesis) that is closely related to but distinct from CDKG kinases. Proteins in the PH kinase group do not have large CDKG-specific N-terminal domains and more than half of the PH genes are contained within gene clusters, which is an uncommon character for kinases. The PH kinase group is present in all the grass genomes investigated as well as in the *M. guttatus* genome, but is not present in most other investigated plant genomes including several rosid dicot species. PH kinase group can be divided into four major sub-groups (termed as PHA, PHB, PHC and PHD respectively) plus the PH gene cluster in *M. guttatus* (termed as PHMg). Proteins of PH kinase group were then compared in detail with other kinases to identify unique features of PH kinases at important kinase motifs and residues. It was found from the comparison that the protein sequences of different PH sub-groups share different numbers of kinase features with CDKG. In brief, PHMg proteins share nearly all the kinase features with CDKG, a few proteins in the PHA and PHD sub-groups share a moderate number of kinase features with CDKG, and proteins in PHB and PHC sub-groups share the least number of kinase features with CDKG.

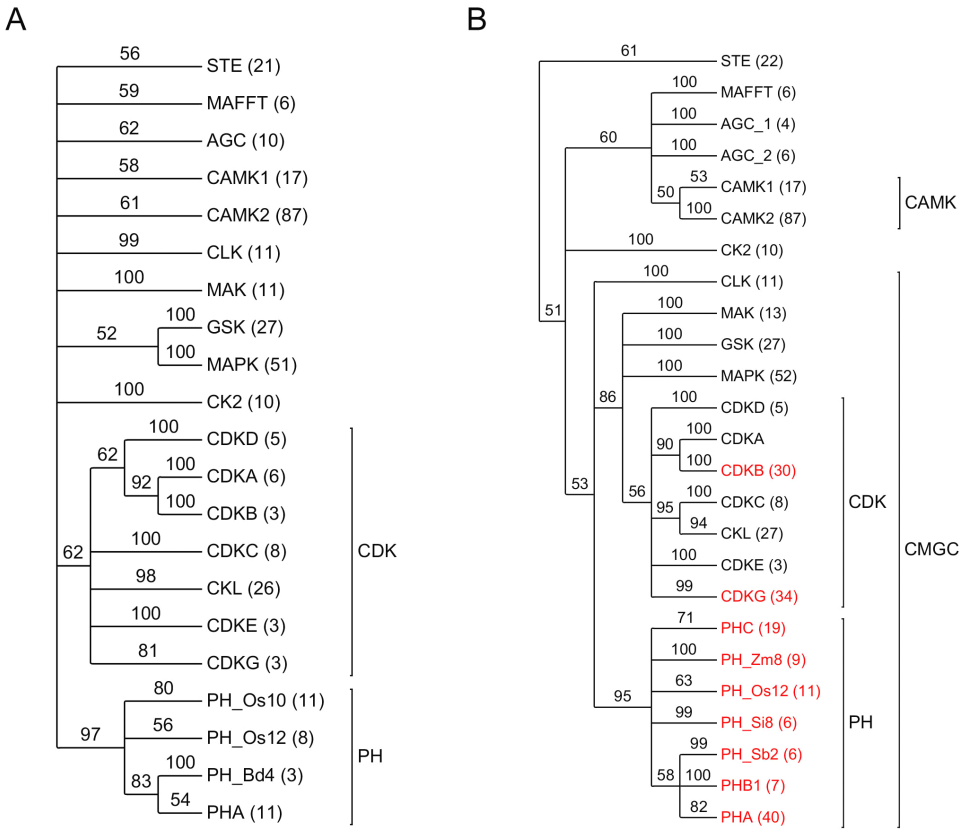
2. Results

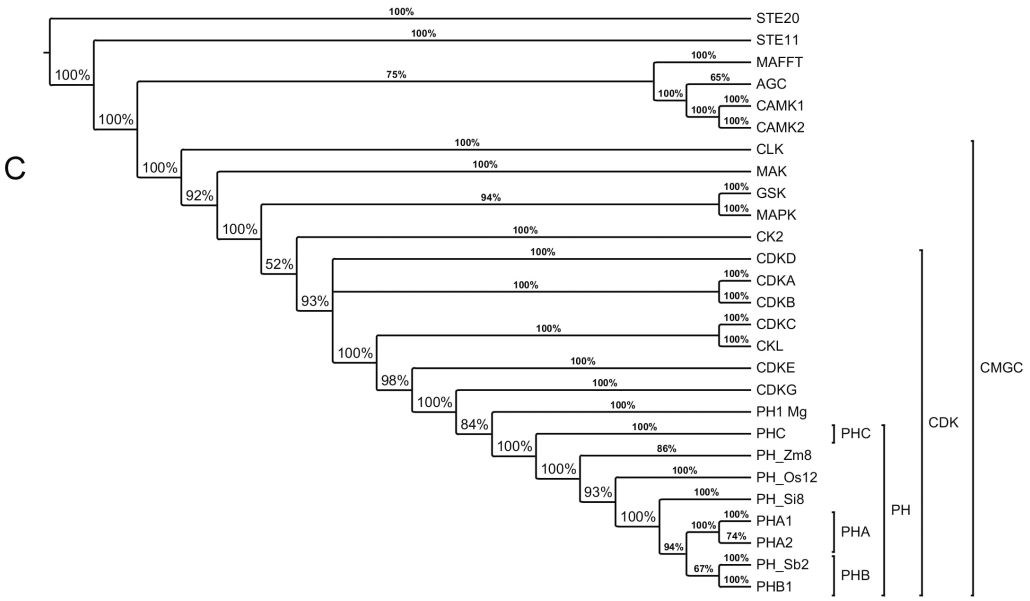
2.1 Ph-CDK orthologues belong to a special kinase group that is closely related to CDKG

The protein sequences of five putative Ph1-CDK orthologs were aligned together with Ph1-D2 and Ph1-A3. The consensus sequence of the alignment was used as template to search for Ph1-CDK related sequences in *A. thaliana, O. sativa* and *B. distachyon* genomes by protein BLAST. A total of 359 moderate size (200-700 aa) protein sequences were aligned together and Neighbor joining phylogenetic trees were constructed from the full length alignment with bootstrap analysis. Brachypodium Ph1-CDK orthologs were included in an kinase group (PH kinase group), which has not been described before, whose *O. sativa* members were annotated as “CMGC Other” by Dardick *et al* (119); the PH kinase group was separated from other kinase families in the phylogenetic trees with the support of a high bootstrap value of 97% (Fig. 1A).

Figure 1 Phylogenetic trees of Ph1-CDK related proteins.

The total sequence number of each node is displayed in the bracket after the node name. The number above the braches shows the confidence value of their corresponding braches. The branch confidences were estimated by bootstrap analysis in **A.** and **B.** Chi2 method was used in **C.** instead to estimate branch confidences. 100 in A. B. and 100% in C. indicate the maximum confidence support. **A.** The Neighbor joining phylogenetic tree of Ph1-CDK related proteins from *B. distachyon, O. sativa* and *A. thaliana.* The tree was constructed from full length protein multi-alignment. PH kinase group is separated from other kinase families on this phylogenetic tree. **B.** The Neighbor joining phylogenetic tree of Ph1-CDK related proteins from 23 more plants genomes. The tree was constructed from only the conserved domains of the protein multi-alignment. The layout of the phylogenetic tree is similar as in **B.** PH kinase group is still separated from other kinase families on this phylogenetic tree. PH, CDKG and CDKB kinase groups were labelled in red to emphasis the increased member numbers due to the adding of new proteins sequences.


A
STE (21)
MAFFT (6)
AGC (10)
CAMK1 (17)
CAMK2 (87)
CLK (11)
MAK (11)
GSK (27)
MAPK (51)
CK2 (10)
CDKD (5)
CDKA (6)
CDKB (3)
CDKC (8)
CKL (26)
CDKE (3)
CDKG (3)
CDK
PH_Os10 (11)
PH_Os12 (8)
PH_Bd4 (3)
PHA (11)
PH
B
STE (22)
MAFFT (6)
AGC_1 (4)
AGC_2 (6)
CAMK1 (17)
CAMK2 (87)
CAMK
CK2 (10)
CLK (11)
MAK (13)
GSK (27)
MAPK (52)
CDKD (5)
CDKA
CDKB (30)
CDKC (8)
CKL (27)
CDKE (3)
CDKG (34)
CDK
PHC (19)
PH_Zm8 (9)
PH_Os12 (11)
PH_Si8 (6)
PH_Sb2 (6)
PHB1 (7)
PHA (40)
PH
CMGC


C
STE20
STE11
MAFFT
AGC
CAMK1
CAMK2
CLK
MAK
GSK
MAPK
CK2
CDKD
CDKA
CDKB
CDKC
CKL
CDKE
CDKG
PH1 Mg
PHC
PH_Zm8
PH_Os12
PH_Si8
PHA1
PHA2
PH_Sb2
PHB1
PHC
PHA
PHB
PH
CDK
CMGC

Interestingly all the members of the PH group were from *O. sativa* or *B. distachyon* genome but not from *A. thaliana* genome. To test if the PH group is limited to certain species, the template used previously was used again to search for Ph1-CDK related sequences in the other 23 plants genomes available on the web server of http://www.phytozome.net/. Presumably, members of PH kinase group would have a protein similarity level higher than at least *A. thaliana* CDKG, a threshold just below the *A. thaliana*

CDKG was therefore applied in the sequence retrieving to keep only the most closely related protein sequences in each genome. Another 125 moderate size (200-700 aa) Ph1-CDK related sequences were collected from the 23 species. They were combined with the previous 359 Ph1-CDK related sequences from *A. thaliana, O. sativa* and *B. distachyon* to go through a second round of alignment and phylogenetic tree constructions. Both Neighbor joining and Maximum likelihood phylogenetic trees were constructed from the conserved domains of the multiple alignment. In both trees, almost all the newly added 125 sequences went to the PH kinase group, CDKB group or CDKG group (labelled as red at Fig. 1B), with only a few exeptions that went to other CMGC kinase families like MAPK or MAK. The newly added members of CDKB and CDKG kinases were from various plant genomes in both trees. However all the new members of PH kinase group were mostly (Maximum likelihood phylogenetic tree) or solely (Neighbor joining tree) from genomes of *Z. mays*, *S. bicolor* and *S. italica*, which like *B. distachyon* and *O. sativa* are grass species. In the second Neighbor joining tree, the grouping of most CMGC families (include PH and apart from CK2) were loosely supported (confidence value of 53%) and the PH group is still distinct from other members of the CMGC family with a high branch confidence value of 95%. However, in the more accurate Maximum likelihood phylogenetic tree, the PH group was included in the CDK family and is close related with CDKG group with three *M. guttatus* proteins located in between CDKG and PH.

2.2 PH kinase group genes tend to form gene clusters

It was noticed in the protein BLAST step that hits from assembled genomes often showed consecutive or close protein ID number (e.g. Bradi4g33220 and Bradi4g33230), which indicated that they are located close to each other in the genomes. A further detailed check of genome positions of all the PH group

members confirmed that gene clustering is a general character for PH kinase group genes (Fig. 2). 57 out of 110 grass PH group genes formed a total of 16 gene clusters. Since the genome sequences of *Z. mays* and *S. italica* are not fully assembled at the time of study, PH gene clusters of *Z. mays* and *S. italica* may further merge into bigger clusters. Three *M. guttatus* kinases located in between CDKG kinases and PH group also form a gene cluster. The clusters are named by the species name abbreviation followed by the chromosome/scaffold number. In a typical PH gene cluster unit, there are often two genes very close together or even forming a tandem gene array, often with a third one located about 100 kb away. In PHBd1, PHOs10 and PHOs12 clusters, 2 or 3 cluster units clustered further to form big clusters (Fig. 2). Genes within the same cluster are also close homologues that share almost identical protein sequences, although it is common that one or more members of a gene cluster contain insertions or deletions. Based on the protein sequence similarity, 16 gene clusters can be divided into 4 big groups, termed as PHA, PHB, PHC and PHD. PHA is the largest group that contains 8 gene clusters from 4 genomes. PHA group can be further divided into two sub-groups PHA1 and PHA2. However the difference between PHA1 and PHA2 is comparatively minor since the grouping of PHA1 and PHA2 was well supported by a high branch confidence value of 100% (Fig. 2). PHB is the group within which the Ph1-CDKs and their orthologs are located.

Apart from the wheat Ph1-CDKs, which are known to form one gene cluster per genome, there are three gene clusters in the PHB groups, one from each genome. All the three gene clusters are small clusters that contain only one cluster unit. PHB can also be further divided into two sub-groups PHB1 and PHB2. Cluster PHBd4, which contains the *B. distachyon* orthologs of wheat Ph1-CDKs, is close related to PHZm6 and these two clusters form the PHB1 sub-group. The other cluster PHSb2 is loosely grouped with PHB1 clusters (branch confidence of 67%), and is therefore grouped separately as PHB2. The groups of PHA and PHB are closely related. The grouping of PHA and PHB are well supported by a high branch confidence of 94%. PHC is a well defined group that contains only one big cluster from *O. sativa* (PHOs10) and one small

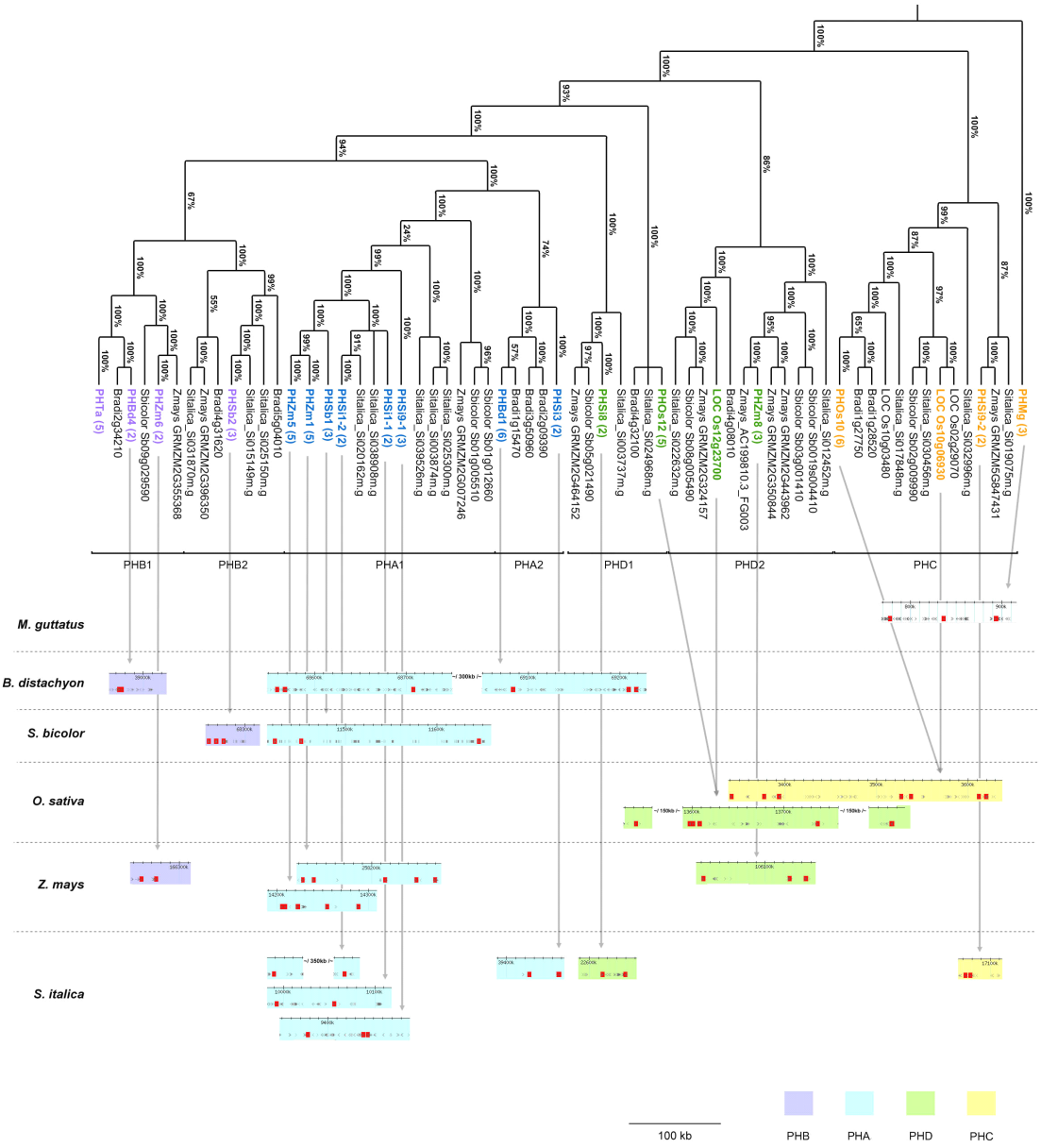


Figure 2 Gene clusters in the PH kinase group.

17 PH gene clusters from 6 plant species were displayed. Grass PH gene clusters can be divided into four sub-groups based on the Maximum likelihood phylogenetic tree. See context at section 2.2 for detailed description of these gene clusters.

cluster from *S. italica* (PHSi9-2). The grouping of PHC is well supported by a high branch confidence of 100%, and the PHC group is comparative distant from PHA and PHB group. The last group, PHD, contains three clusters, each from one genome. PHD group is in fact a very loosely defined group. Cluster PHSi8 is close to PHOs12 on the phylogenetic tree but they are not positioned in the same clade. PHOs12 cluster have only one gene that are loosely grouped with PHZm8 cluster. The PHD group is located in between PHC and PHA, PHB. It was noticed that in fully assembled genomes of *B. distachyon, S. bicolor* and *O. sativa,* there are two PH gene clusters contained within each genome. Both *B. distachyon* and *S. bicolor* genomes contain one PH gene cluster in both PHB and PHA groups, and *O. sativa* genome contains one PH gene cluster in both PHC and PHD groups. There are also a few PH genes from *B. distachyon* and *S. bicolor* that are located within PHC and PHD groups, but there are no *O. sativa* PH genes contained in the PHA and PHB groups. In fact, *O. sativa* genome is the only genome within the five that does not have genes located in the PHA and PHB group. *B. distachyon* genome as a whole is more closely related to the *O. sativa* genome than to other three *Panicoideae* genomes. It is therefore interesting to investigate why the PHA and PHB genes are missing in the genome of *O. sativa.*

2.3 PH group is exclusive to grass and M. guttatus genomes

CDKG is a newly discovered CDK kinase group that is involved with early meiosis processes (personal communication with Dr. Zhen T., as will be discussed later in chapter V). There are two well conserved copies of CDKG in the genome of *A. thaliana.* At5g63370 (CDKG;1) is included in the analysis here, and At1g67580 (CDKG;2), which was excluded from the multi-alignment since it is larger than 700 aa. CDKG kinases are characterized by a large CDKG specific N-terminal domain that is about 300 aa in size. As mentioned before, the Maximum likelihood phylogenetic tree reported in this study was constructed based on the kinase conserved domain only and the large CDKG N-terminal domains were removed from the alignment before the Maximum likelihood phylogenetic tree construction. It is therefore possible for the CDKG clade to contain some proteins that lack the CDKG specific N-terminal domain, which as a result is structurally closer to PH kinase. However it was shown by the full

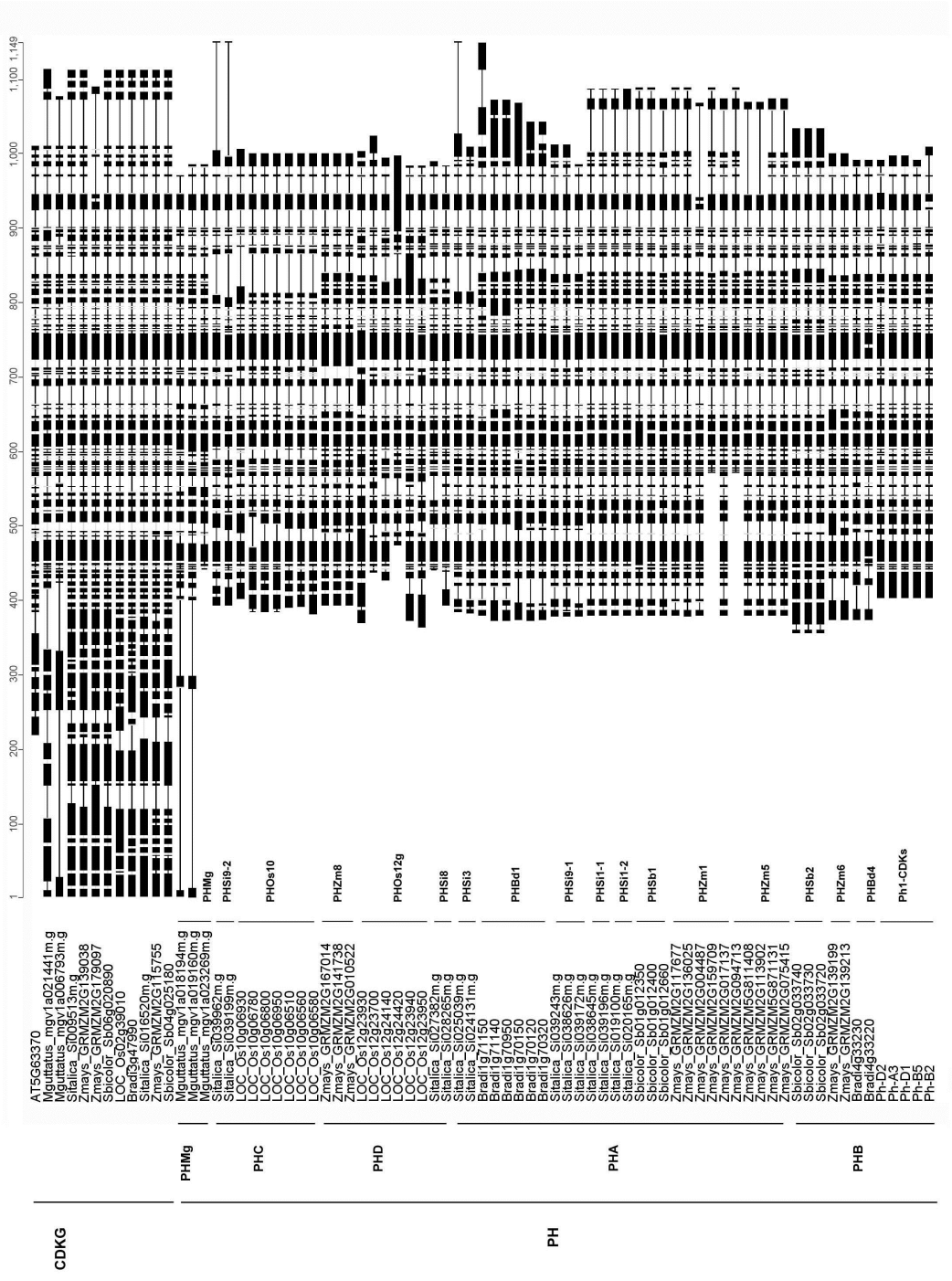


Figure 3 The overview PH proteins and CDKG proteins.

PH proteins do not contain the CDKG specific N-terminal domains. Species that contain PH proteins in their genomes do also have CDKG proteins.

length multi-alignment that all the genes in the CDKG clade have CDKG specific N-terminal domains including those from grass genomes and genome of *Mimulus guttatus* (Fig. 3). Three genes in the PHMg gene cluster, in contrast, do not contain this extra N-terminal domain. Since these three genes also form gene cluster like many other grass PH group kinases do, they were considered as included in or at least close related to the PH group. As mentioned before, the PH kinase group was limited to 5 grass genomes with genes of PHMg gene cluster staying in between CDKG and PH. It was therefore concluded that the other 19 plant genomes investigated do not have members of the PH kinase group, even when the full length protein sequences were considered. Also the analysis confirmed the form of gene clusters, the lack of extra N-terminal domains and CDKG related protein sequences as the features of PH family proteins. It is not clear yet how the PH kinase group originated; this could be answered by searching newly sequenced plant genomes with the identified PH kinase group features in the future.

2.4 Comparison of PH kinase group with other kinase families

Although PH kinase group is located within the CDKG subtree in the Maximum likelihood phylogenetic tree, the lack of CDKG specific N-terminal domain in the PH kinases and the fact that in both Neighbor joining trees PH kinase group is clearly distinct from CDKG still suggested that PH kinases are a separate kinase group that is closely related but distinct from CDKG. It is therefore of interest to compare PH kinases as a group with other defined kinase families. Such a comparison can also provide a broad view of important common motifs of kinases, which helps to understand the differences between PH kinases and CDKG kinases.

The building of Hidden Markov Models (HMM) profile of one protein family is a convenient way of summarizing the well conserved motifs and residues among a large number of proteins. In this study, full length alignments of the PH kinase group, the PHB1 sub-group together with three CMGC kinase families and three non-CMGC families were extracted separately from the second alignment of 484 protein sequences and HMM logos of each kinase family were generated from the alignments to show the invariant residues and degenerate motifs of each kinase family after sequences containing obvious abnormal insertions and

deletions were deleted from the alignment (Fig. 4). STK share a generally well conserved protein structure that is composed of 12 conserved sub-domains that are defined by various invariant residues and/or degenerate motifs. The comparison of the PH kinase group with other defined kinase families will be discussed by sub-domains in order. To avoid the loss of sub-group specific information that might be buried in the HMM profiles of whole protein groups, the representative protein sequences of each PH gene cluster and representative CDKG protein sequences from PH gene containing genomes were extracted from the multi-alignment to be compared (Fig. 5). When differences exist between the PH kinase group and CDK kinase family in important motifs, these differences are examined in the original multi-alignment as well.

Sub-domain I is the start of the N-terminal of the kinase domain. The glycine-rich loop of sub-domain I acts as a flexible flap that helps to anchor ATP to the protein, and is therefore termed as “phosphate anchor”. The four amino acids in the backbones of G-[FY]G form hydrogen-bonds with ATP’s beta- and gamma-phosphates. As shown at Fig. 4, the sub-domain I of all the investigated kinase families are quite conserved with one possibly interesting difference that the conserved amino acid [FY] in the G-G-[FY]G motif is missing in HMM logo of the PH group, which may have an effect on ATP binding. As shown by the multi-alignment, this tyrosine is missing in most of the PH group kinases, but is maintained in four wheat Ph1-CDK genes (Fig. 5, red frame 1).

Sub-domain II is not well conserved among different kinase families and contains only one invariant residue, lysine, which interacts with the phosphate groups of ATP. The invariant lysine is present in all the investigated kinase families but the rest of the sub-domain II is quite diverse even among CMGC kinase families.

Sub-domain III consists of a large alpha helix (helix C) in the N-terminal lobe. The CDK family signature PITSLRE, which is responsible for the recognition and binding of specific cyclins, is located at the beginning of the sub-domain III.

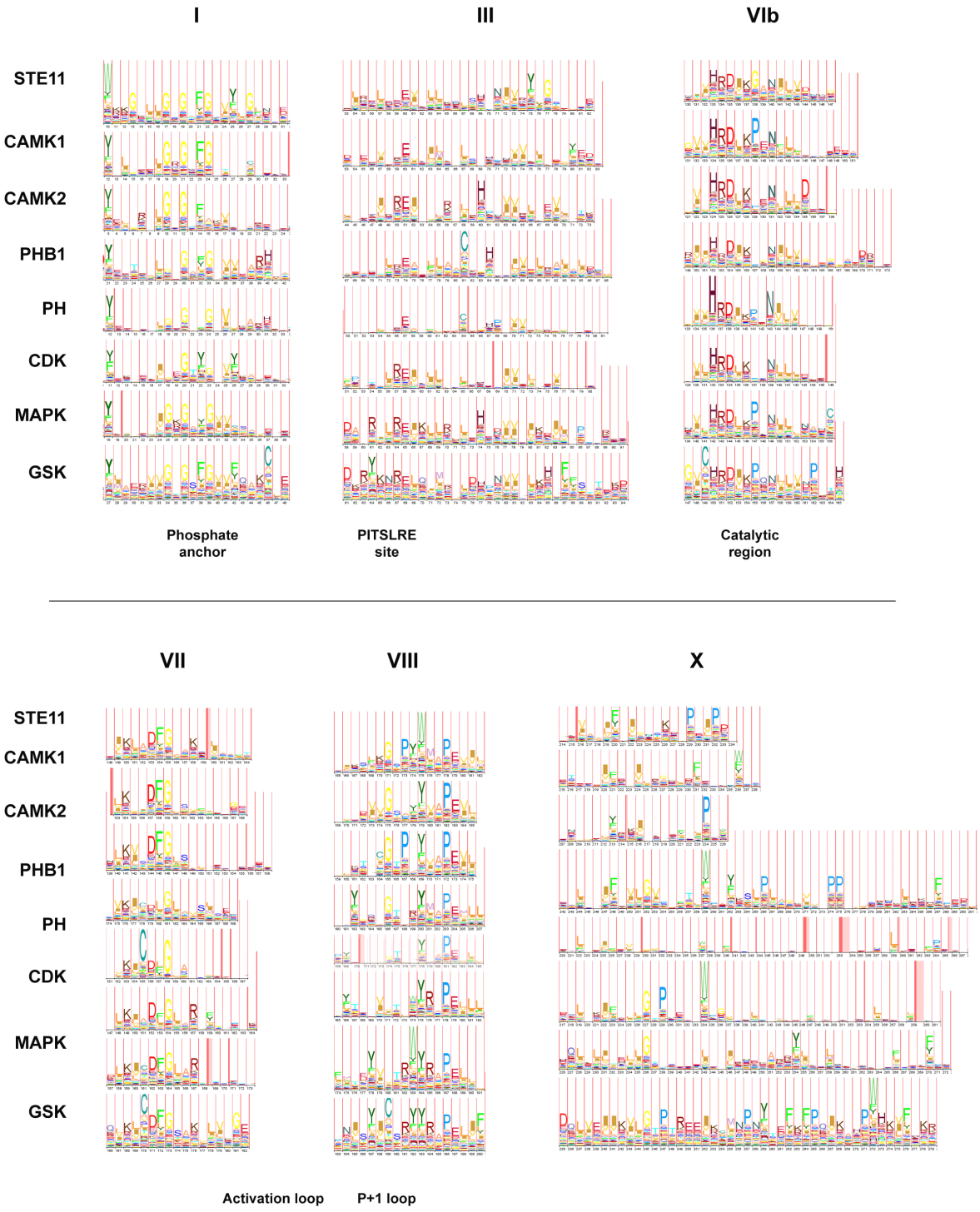
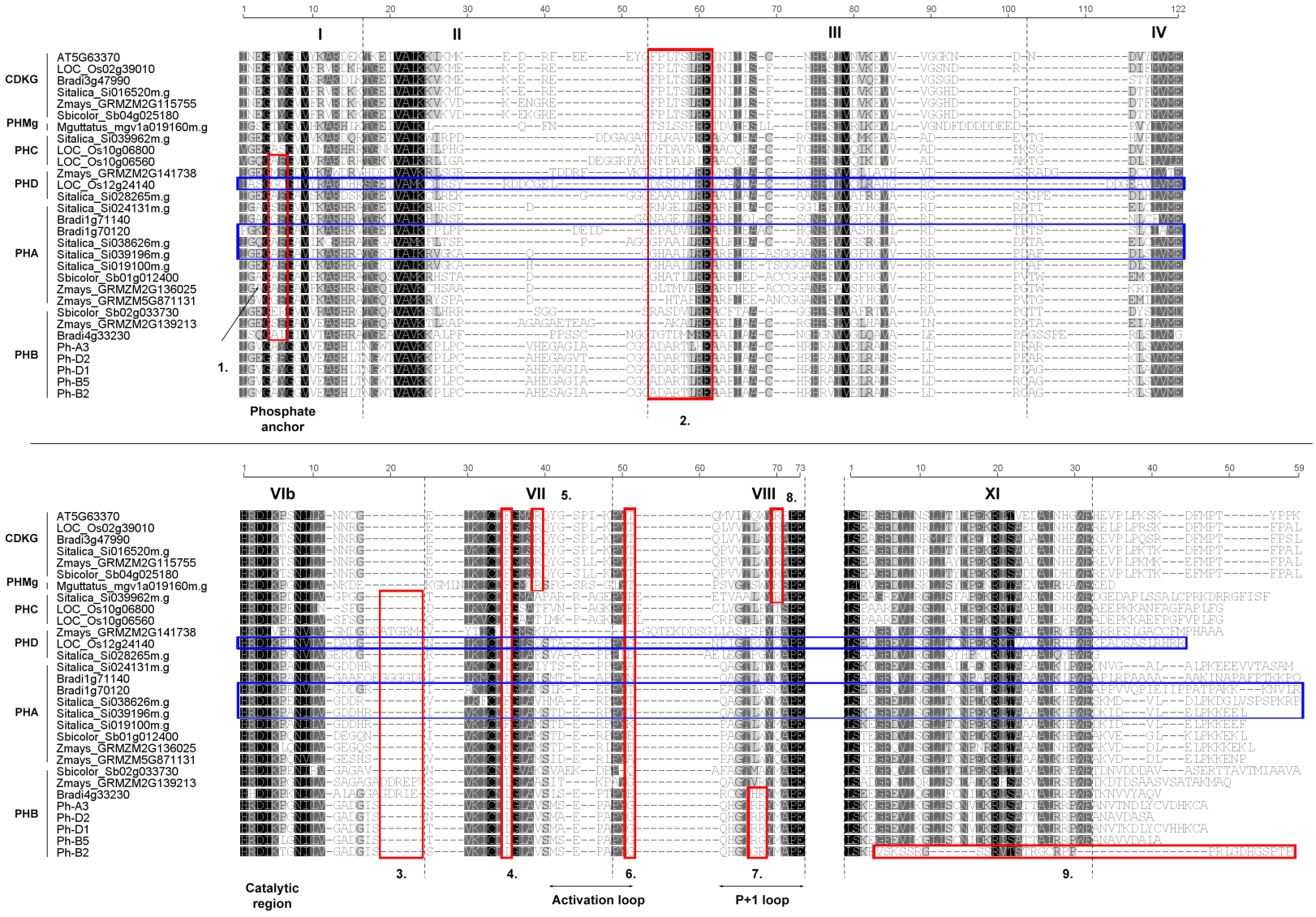


Figure 4 Comparison of PH kinase group with other kinase families.

HMM logos of each kinase family/group were aligned based on kinase sub-domain. See context at section 2.3 for details.

Figure 5 Sequence comparison of PH proteins with CDKG proteins.

The representative protein sequences of PH gene clusters and CDKG from PH gene containing genomes were extracted from the multi-alignment and labelled with kinase sub-domain numbers. Numbered red frames were used to indicate unique PH protein features at important kinase functional regions/sites. Blue frames indicate grass PH proteins that share more kinase features with CDKG. See context at section 2.3 for details.


CDKG
PHMg
PHC
PHD
PHA
PHB
AT5G63370
LOC_Os02g39010
Bradi3g47990
Sitalica_Si016520m.g
Zmays_GRMZM2G115755
Sbicolor_Sb04g025180
Mguttatus_mgv1a019160m.g
Sitalica_Si039962m.g
LOC_Os10g06800
LOC_Os10g06560
Zmays_GRMZM2G141738
LOC_Os12g24140
Sitalica_Si028265m.g
Sitalica_Si024131m.g
Bradi1g71140
Bradi1g70120
Sitalica_Si038626m.g
Sitalica_Si039196m.g
Sitalica_Si019100m.g
Sbicolor_Sb01g012400
Zmays_GRMZM2G136025
Zmays_GRMZM5G871131
Sbicolor_Sb02g033730
Zmays_GRMZM2G139213
Bradi4g33230
Ph-A3
Ph-D2
Ph-D1
Ph-B5
Ph-B2
I
II
III
IV
1.
Phosphate anchor
2.
VIb
VII
5.
VIII
8.
XI
3.
4.
6.
7.
9.
Catalytic region
Activation loop
P+1 loop

Although the PH group is located within the CDK family and is closely related to CDKG in the Maximum likelihood phylogenetic tree , there is no PITSLRE signature in either the PHB1 small group or the whole PH kinase group. In fact, apart from PH and CAMK2 kinase families, all the other kinase families investigated do have some kind of motifs at the PITSLRE equivalent sites. A detailed look at the multi-alignment confirmed that all the CDKG genes have a totally conserved PLTSLRE motif at this site, but no clear motif could be found even among genes of the same PH sub-groups (Fig. 5, red frame 2).

The functions of Sub-domain IV and VIa appear to be mostly structural and like Sub-domain II are not well conserved.

Sub-domain VIb is comprised of two small hydrophobic strands with a highly conserved loop between them. The loop between the two strands contains five invariant or highly conserved residues that form the kinase signature HRD-K--N. This region is termed as the ‘catalytic region’, because the invariant aspartate of the HRD-K--N motif is intimately involved in catalysis. As shown at Fig. 4, the sub-domain VIb of all the investigated kinase families are highly conserved with the only difference being that PHB1 has a few more amino acids at the end of its sub-domain VIb. It was shown by the multi-alignment that the extra sub-domain VIb end residues are not exclusive to PHB1 sub-group but also exist in a few other PH genes, but that none of the wheat Ph1-CDKs, members of PHB1 sub-group, have such extra end residues (Fig. 5, red frame 3).

Sub-domain VII contains the well-conserved triplet of ‘D[FL]G’, which lies in the loop between the strand–loop–strand or strand–loop–helix structure. The invariant aspartate has the function of orienting the gamma-phosphate of ATP for its transfer to the substrate. PH kinases tend to have a DLG motif rather than the DFG motif that all other kinases have, and it was shown by the multi-alignment that this replacement of Phenylalanine by Leucine is more commonly seen in the PHB, PHC and PHD sub-groups than in PHA sub-group (Fig. 5, red frame 4). There is a conserved charged amino acid Arginine or Lysine three amino acids down the D[FL]G triplet in all three CMGC kinase families as well as in STE11 family. However this conserved residue is missing in the PH group as well as in both CAMK families. A detailed look at the

multi-alignment showed that this charged residue is replaced by an aliphatic residue (Methionine, Alanine or Valine) in most PH kinases (Fig. 5, red frame 5).

Sub-domain VIII is another key sub-domain as well as sub-domain I and VIb. It contains the highly conserved 'APE' motif, and the residues before the APE motif plays a major role in the recognition of the substrate and are therefore often well conserved among members of the same kinase family. The activation of many protein kinases is accomplished by the phosphorylation of the Ser/Thr that is 9 or 10 amino acids left of the APE motif, which can induce a conformational change that is crucial for the activity of kinase. This Ser/Thr activation residue is present in all the investigated kinase families except in CAMK1 family and PH group. It is shown by the multi-alignment that, although most of the PH kinases do not have the activation Ser/Thr as expected from the HMM logo, there are a few proteins from PHA and PHC as well as the proteins of PHMg cluster do have this important residue (Fig. 5, red frame 6). Also there is a conserved Arginine just before the APE motif in all the three CMGC kinase families, but this Arginine is missing in both the PH group and the PHB1 sub-group. The multi-alignment showed that this conserved Arginine is missing in nearly all the PH proteins, but a similar charged Lysine is present at the PHMg cluster protein (Fig. 5, red frame 8). It is also noticable that PHB sub-group proteins have a nearly conserved Arginine at the APE -3 position and the wheat non B genome Ph1-CDKs (A3, D1 and D2) have an extra Arginine at the APE -4 position (Fig. 5, red frame 7). It was shown by the HMM logos that the presence of Arginine at the APE -3 positions is an exclusive feature of PHB sub-group, which may give PHB sub-group kinases a specific substrate preference.

Sub-domain IX has the secondary structure of a large alpha-helix. There is a Y----D-WS-G motif located in the middle of this sub-domain. The invariant aspartate in the motif acts to stabilize the catalytic loop. This sub-domain is in general well conserved in all the investigated kinase families. In all the CMGC kinase families as well as PH kinase group, there is a well conserved Cysteine right next to the Glycine in the Y----D-WS-G motif. Also in only CDKG and PH kinase families, there is a highly conserved aspartate right next to the Tyrosine in the Y----D-WS-G motif, which is not present in other kinase families

investigated. If taking these minor differences into account, the sub-domain IX of PH group is more closely related to CMGC kinase families, especially the CDK family.

Sub-domain X has the structure of a small alpha-helix that sits at the base of the large lobe. The sequence from this region is not well conserved. Sub-domain X of PH kinase group and CMGC kinase families all have a large insertion at the end of this sub-domain. Also PHB1 sub-group has the well conserved Glycine that is present in all the CMGC kinase families but not others.

Sub-domain XI is the last sub-domain present in the kinase domain of STKs. This sub-domain does not have a clearly conserved function in general and has only a few scattered conserved residues that are present in nearly all the investigated kinase families. It is noticed from the multi-alignment that, wheat Ph-CDKB2 has a nearly completely altered sub-domain XI (Fig. 5, red frame 9), which is caused by the insertion of a single nucleotide that leads to a frame shift at the beginning of sub-domain XI.

The identified important differences of PH kinases compared with other kinases, especially CDKG kinases, are summarised at Table 1. It is clearly shown by the results that PHMg genes are nearly the same as CDKG, in that in 6 out of 8 sites, PHMg genes have the same motifs/residues as CDKG, and in the remaining two sites PHMg genes have similar motifs/residues to CDKG. There are a few PHA or PHD sub-group genes (annotated by blue frames on Fig. 5) that are quite similar with CDKG with 4 out of 8 identical motifs/residues with CDKG. The rest of PHA and PHD genes together with PHB and PHC sub-group genes are less similar to CDKG with only less than two identical motifs/residues to CDKG.

When an STK kinase phosphorylates its substrate, residues surrounding the Ser/Thr on the substrate make contact with certain residues in the binding pocket of the STK kinase catalytic region. The nature of these residues, termed as Substrate Determining Residues (SDR), determines the preference of the

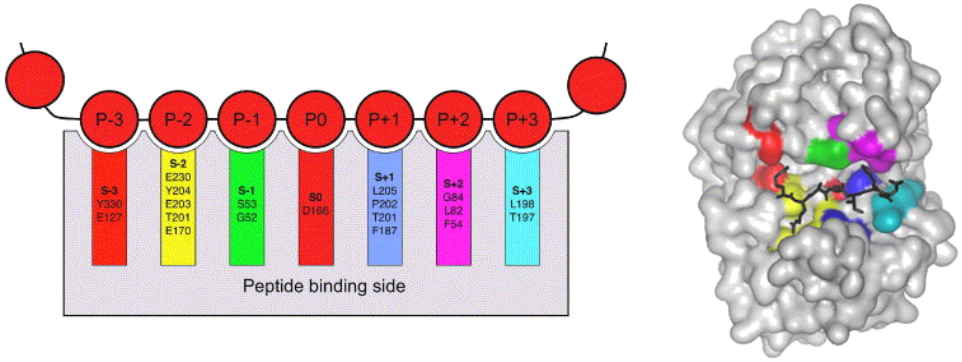


Figure 6 Diagrams showing the substrate binding in protein kinase A.

(taken from (120)).

Frame No.	2	6	7	8	1	3	5	4
Motifs	PITSLRE	Activation S/T	APE -14	APE -1	G-YG	Extra end	R	DFG
Subdomain	III	VIII	VIII	VIII	I	VIb	VII	VII
Function	Cyclin binding	Activation	Substrate	Substrate	ATP anchor	Catalytic		ATP transfer
CDKG	√	√	Aliphatic	Charged	√	X	√	DFG
PHMg	Similar	√	Aliphatic	Charged	√	X	√	DFG
PHC	X	X	Aliphatic	Aliphatic	X	X	X	DLG
PHD	X	A few	Aliphatic	Aliphatic	X	A few	X	DLG
PHA	X	A few	Aliphatic	Aliphatic	X	A few	X	DFG
PHB other	X	X	Charged	Aliphatic	X	√	X	DLG
Ph1-CDK	X	X	Charged	Aliphatic	√	X	X	DLG

Table 1 Summary of differences between PH sub-groups and CDKG kinase group.

Red color labels CDKG like sites. Blue color labels PHB like sites. Purple color labels mixture situation.

Substrate position	SDRs	
	STK	CMGC
-3	GEL+1, GEL+3, GEL+4	GEL+1, GEL+3, PEN+1
-2	APE-2, APE-3, APE-5	APE-2, APE-3, APE-5
-1	GXG+2, GXG+3	GXG+2, GXG+3
+1	APE-1, APE-4, DFG+3	APE-1, APE-4, DFG+3
+2	LRE-3, LRE-4, LRE-5	GXG+3, GXG+4
+3	APE-8, APE-9	APE-9, APE-10

Table 2 SDRs location determine system (taken from (120))

	-3	-2	-1	+1	+2	+3
STE	S--	WYS	GA	MPA	RL-	SK
CAMK	EFD	YYS	G-	VAL	E-D	D-
CAMK	EF-	YAT	G-	VAL	--L	TK
PHB	SHD	YRT	GA	M-L	ADA	-Y
PH	S--	Y-T	G-	---	---	-Y
PHB	SH-	YRT	GA	M-L	AF	-Y
PH	S--	Y-T	G-	---	-F	-Y
CDK	DK-	YWT	GT	R-L	TY	TY
MAPK	DH-	YWT	GA	RRL	AY	TM
GSK	T-Q	YYS	GS	RRS	SF	IN

Table 3 SDRs of PH kinase group and other kinase families.

	-3	-2	-1	1	2	3
CDKG	DKS	YWT	GT	RL[LM]	TY	TY
PHMg	DKG	YWT	GT	KSM	TY	TY
PHC	SR[ED]	YPT	GA	TL-	AF	-Y
PHD	SR[DE]	Y-T	GG	--L	GF	SY
PHA	SH-	YWT	G[GA]	[MQ]LL	[GA]F	[TE]Y
PHB other	SHG	YRT	GA	MML	AF	GY
Ph1-CDK	SHG	YRT	GA	MML	AY	GY

Table 4 SDRs of PH sub-group and CDK

Ser/Thr surrounding residues on the substrate and therefore makes a major contribution to the substrate specificity of the kinases (Fig. 6).

The accurate positions of six SDRs determined by examining X-ray structures of STK kinases bound to substrate heptapeptides were reported before, and are confirmed by analysis of yeast cell cycle control and DNA damage checkpoint pathways (120,121). The SDRs of PH kinase group as well as other kinase families were identified using the reported SDR determining system. Since the locations of SDRs are different in CMGC kinase families and other STK kinase families, the SDRs of PH kinase group were determined by considering them as CMGC or not (Table 2). As the result shows, SDR patterns at -1 and -2 positions are quite similar for all the investigated kinase families apart from the fact that the PHB1 kinase sub-group has a unique Arginine at the second SDR of -2 SDR pattern. The +1 SDR patterns of CMGC kinase families are characterized by one or two Arginines, which are missing in the PH kinase group. Instead, PH group has a +1 SDR pattern that is similar to STE11. The +2 SDR pattern of PH group is closer to CMGC kinase families, which is featured by Aromatic Tryptophan or Phenylalanine at the third SDR. The +3 SDR pattern of PH group is close to CDK family, which both contain a Phenylalanine at the third +2 SDR. The -3 SDR pattern of PH group is close to STE11 family, which have a Serine at the first -3 SDR. For the CDK family and MAPK family, their -3 SDR patterns share the same Aspartate at the first -3 SDR followed by a charged Lysine or Histidine.

The SDRs of different PH sub-groups were also identified and compared with CDKG kinase group (Table 3). The analysis showed that PHMg cluster has nearly the same SDRs as the CDKG group. In contrast, the SDRs of PHB and PHC sub-groups are different from CDKG's at all 6 positions. SDRs of PHA and PHD sub-groups are more closely related to those of PHB and PHC subgroup, but they do have one or two SDR positions that are close to CDKG rather than PHB or PHC sub-groups. The results suggested that PH kinase group, especially the PHC and PHB sub-groups where the wheat Ph1-CDKs are located, is unlikely to phosphorylate the same set of substrates as CDKG.

The detailed comparison of PH group proteins with CDKG proteins showed that proteins in different PH sub-groups have different levels of similarity compared

with CDKG. PHMg proteins are close to CDKG with nearly the same kinase motifs and SDRs. A few proteins in the PHA and PHD sub-groups (annotated by blue frames at Fig. 5) are more close to CDKG than others with more CDKG similar kinase motifs and SDR residues. Proteins in the PHB and PHD sub-groups are more distantly related to CDKG, with more modifications at important kinase motifs and almost distinct SDRs. The gradual divergence of PH protein sequences suggested that PH kinase group was likely to be derived from or at least shared a common ancestor with CDKG proteins, and evolved more quickly in grass species than in *M. guttatus* genome. It is however unclear from the bioinformatics analysis whether PH proteins still have the same or a similar function to CDKG or whether they developed new diverse functions during the process of evolution. The existence of PH group protein in both dicot *M. guttatus* and grasses suggests that the origin of the PH kinase group preceded the monocot-dicot divergence; however it is unknown that why the PH kinase group is not present in other dicot species investigated apart from *M. guttatus.*

3. Materials and methods

3.1 Ph1-CDK homologous sequence collection

In the first round of sequence collection, Ph1-B2, Ph1-D2, Bradi4g33230, Bradi2g were aligned by AlignX module of VectorNTI (Ver. 11) software suite and the consensus sequence that was constituted by identical amino acids was interspersed with letter X for non-identical amino acid was used as template to search for homologous sequences of Ph1-CDKs in *A. thaliana, O. sativa* and *B. distachyon* genomes respectively via protein BLAST web server at http://www.phytozome.net/.

For each genome only the top 200 putative protein sequences per genome were retrieved. The original 600 protein sequences were selected in Vector NTI (Ver. 11) software with a manner that only one representative sequence per gene and only sequences within the size range of 200-700 aa were kept. As a result 359 putative protein sequences were selected for the first round of sequence alignment.

In the second round of sequence collecting, the same template used above was used again to BLAST against each of the 25 plants genomes (apart from *A. thaliana, O. sativa* and *B. distachyon*) available on the web server of http://www.phytozome.net/. A threshold that just below the hit of *A. thaliana* CDKG was applied to select only the most closely related protein sequences in each genome. Again, only one representative sequence per gene within the range of 200-700 aa was selected. Another 125 putative protein sequences from 23 genomes were selected and combined with the previous collection of 359 sequences to be aligned in the final round of sequence alignment.

3.2 Multiple alignment and phylogenetic tree construction

For the first round of sequence alignment, the collection of 359 protein sequences were uploaded onto and aligned by the MUSCLE (122) multi-alignment method at http://www.ebi.ac.uk/Tools/msa/. The full length alignment was then imported into ClustalX software platform (Ver. 2.1) and the Neighbor joining phylogenetic tree was constructed by the “Bootstrap N-J tree” function with default setting and bootstrap analysis of 1000 re-sampling.

For the final round of sequence alignment, the collection of 484 protein sequences were imported into Geneious software platform (Ver. 5.4.6), and the alignment was performed by the E-INS-i iterative refinement method of the MAFFT (123) module (Both MAFFT and MUSCLE alignment modules were tested and comparable results were obtained. The alignment of MAFFT was selected because a slightly better structured phylogenetic tree was obtained with this alignment.).

The full length MAFFT alignment was trimmed in the Gblock (Ver. 0.91b) software (124) to keep only the conserved blocks of the alignment which is 316 aa in total. Both Neighbor joining phylogenetic tree and Maximum likelihood phylogenetic tree were constructed based on the conserved blocks only alignment via the “Geneious tree build” module or the PHYMaximum likelihood (125) module of the Geneious software platform (Ver. 5.4.6) respectively. For the Neighbor joining phylogenetic tree construction, default setting was used and the branch confidence was estimated by the bootstrap analysis of 1000 re-samplings. For the Maximum likelihood phylogenetic tree construction, “WAG

substitution model", "0 invariable sites", "4 substitution rate categories", "estimated gamma distribution parameter" and "optimization for tree/brach" options were selected, and the branch confidence was estimated by the Chi2 method (126). For both trees, only braches with more than 50% confidence support were kept in the result phylogenetic trees.

The classifications of *O. sativa* kinases were obtained from the rice kinase database at http://phylomics.ucdavis.edu/kinase/ (119). The classifications of *A. thaliana* kinases were obtained from the Arabidopsis kinase database at http://bioinformatics.cau.edu.cn/athKD/browse.htm. The classifications of *B. distachyon* kinases was based on the classification of *O. sativa* kinases and/or *A. thaliana* kinases within the same or nearby branches. The annotations and graphic outputs of the phylogenetic trees were handled by Treegraph software (Ver 2.0.47-206 beta) (127).

The genome positions of the PH kinases were retrieved and displayed on the corresponding Gbrowses at http://www.phytozome.net.

3.3 HMM profile building and display

The HMM profiles (128) of particular kinase families were generated at http://www.genome.jp/tools/motif/MOTIF3.html, which use the Hmmbuild module of the HMMER software suite to generate the HMM profiles. To display the HMM profiles, HMM logos were generated at http://www.sanger.ac.uk/cgi-bin/software/analysis/logomat-m.cgi (129). The obtained HMM logos were saved as PNG image files to be arranged and handled by the PS (Ver. CS3) software.

CHAPTER III Studies of *Ph1* candidate genes in *S. cerevisiae*

1. Background

1.1 Budding yeast is an ideal eukaryotic model organism

Budding yeast (*Saccharomyces cerevisiae*) is the first eukaryotic organism that was developed as a model system for biology studies. Budding yeast is very easy to handle and propagate. It shares many essential processes with higher eukaryotic organisms. More than this, decades of development by the yeast community has given this model system large stocks of material and technical resources. The budding yeast genome was fully sequenced years ago. It was the first organism for which a mutant library was established. Transformation of yeast cells can be done within a few days. Handling of target genes (mutation, tagging) can be done with a simple PCR based strategy (130). Compared with other higher eukaryotic model organisms like mouse or Arabidopsis, working with budding yeast can save the researcher a lot of effort. In fact, budding yeast has been the pioneer organism for many break-through experiments in eukaryotic biology. The complementation of yeast mutants by counterparts from other higher eukaryotes has helped to identify many genes, including the CDK genes in mammals (131) and plants (132). The genome sequence of this simple but informative microorganism has given powerful insights into later genome sequencing projects including the human genome sequence. Due to the convenience of propagation and the existence of many efficient protein tagging systems, budding yeast is also an excellent system for protein studies.

Budding yeast, as a simple microorganism, is quite different from the much more complicated higher eukaryotic organism in many ways. However, given all the advantages discussed above and considering the time and effort that could be saved, for researchers that focus on fundamental biological processes, Saccharomyces cerevisiae still serves as a good model system since these events are often conserved within the eukaryotic world to a large extent.

Budding yeast has always been a popular system for CDK kinase studies as well as meiosis studies, both relevant for studies of the *Ph1* candidate genes. Therefore, it was decided to test the function of the *Ph1* candidate genes in the Saccharomyces cerevisiae cells.

1.2 Functions of the Ph1 candidate gene homologue in budding yeast.

Ime2 together with HsCDK2 are the two genes whose DNA sequences share the biggest similarities with *Ph1* candidate genes outside the plant kingdom. Actually the similarities between *Ph1* candidate genes and Ime2, HsCDK2 are even higher than compared with AtCDKG1.

Ime2 is a kinase in budding yeast that is only expressed during meiosis and is crucial for different steps of the process. The most important reported role of Ime2 is its involvement in transcriptional regulation during meiosis. Genomic microarray studies (133,134) have revealed that the budding yeast meiotic genes can be categorised into four major waves based on their transcriptional profiles. Early meiosis genes, the first wave, are mostly controlled by two upstream elements UAS and URS1 (135,136) via transcription factors Abf1 and Ume6. URS1 is controlled by Ume6, a transcription factor that can serve as repressor or activator depending on the binding and phosphorylation of Ime1. Ime2 is one of these early meiosis genes, and the transcription of Ime2 is controlled by the Ume6-Ime1 complex via the URS1 element. Ime2 is required for the initiation and maintenance of transcription of the middle meiosis genes, the second wave. Middle meiosis genes normally contain a regulatory element called MSE (middle sporulation element). In the vegetative condition, MSE sites are occupied by a repressor called Sum1. During meiosis, Ime2 removes Sum1 by phosphorylation to allow the binding of the activator transcription factor Ndt80. There also exists an MSE site in the upstream sequence of Ndt80; thus the positive feedback loop will greatly increase the transcription of Ndt80 and other middle meiosis genes. Many of these middle meiosis genes are required for the meiotic divisions later on, therefore the deletion or mutation of Ime2 will lead to a pachytene stop.

Ime2 is also required for the proper initiation of meiotic DNA replication. The activation of Cdc28 by binding B type cyclins Clb5 and Clb6 is also required for the initiation of meiotic DNA replication in budding yeast. Sic is a repressor of this complex, which needs to be degraded to allow binding of Cdc 28 and Clb5 and Clb6. The activity of Ime2 is required for the degradation of Sic1, and thus it is essential for the initiation of meiotic S phase.

1.3 HsCDK2 can up-regulate the transcription of Hop1

The similarity between Ime2 and CDK family kinase is not limited to the amino acid sequence. Analysis of purified Ime2 protein has displayed that Ime2 even shares some important kinase properties like substrate preference, crucial lysine site and divalent metal ion preference with CDK family kinase (137).

Higher eukaryotic organisms like mammals and higher plants have several different CDKs within one genome. So far five out of eleven CDKs in mammal, CDK1-4, and CDK6, have been confirmed to be involved with cell cycle progress directly. In the classic cell cycle model, from G1 onwards, cells are driven in order by CDK4,CDK6-Cyclin D (138)., CDK2-CyclinE (139,140), CDK2-CyclinA and finally CDK1-Cyclin B complex to go through a full cell cycle and cell division afterwards. It had been believed for a long time that most of the CDKs in higher eukaryotic organisms are essential and could not be deleted from the genome, until it was reported that in mouse, the knockouts of all the other CDKs apart from CDK1 were surprisingly not lethal. Detailed study later on revealed that, among those CDKs that are normally involved in cell cycle but not essential for mitotic division, CDK2 is the only one that is essential for the process of meiosis. Mice with a CDK2 knock out in the genome were sterile. (141)

As discussed above, Ime2 is a CDK-like kinase, and a meiosis-specific one. Mammalian CDK2 is especially important for meiosis as well. Moreover, it has been showed recently that HsCDK2 can partially complementary the function of Ime2 in budding yeast (142). One of the results that supported this conclusion is that the over-expression of HsCDK2 could induce the expression of an early meiosis gene Hop1, the same effect as that of Ime2. Hop1 is an early meiosis gene encoding a component of the axial element (AE) and is required for both

programmed SPO11 catalyzed meiotic DSB formation (143) and for proper meiotic recombination with inter-homolog (IH) bias by promoting the dimerization of Mek1 and thereby preventing the inter-sister (IS) recombination. (144-147). The phosphorylation of a [S/T]Q cluster domain (SCD) within HOP1 by MEC1, the homolog of mammalian ATR, is required for preventing IS recombination during meiosis but not for programmed meiotic DSB formation (148). Interestingly, it was also reported recently that the transcription of TaASY1, the homolog of Hop1 in bread wheat, is also up-regulated in the absent of the *Ph1* locus, while the transcription of other meiotic genes including TaRAD51A, TaRAD51B and TaMSH4 remained at a similar level with or without the presence of the *Ph1* locus (149). As mentioned in the *Ph1* locus introduction, there are 14 copies of *Ph1* candidate genes clustered within the *Ph1* locus and its related locus on the D and A genomes. In the absence of the *Ph1* locus, where 7 copies of *Ph1* candidate genes from the B genome are deleted and the other 7 copies remain in the A and D genomes, the total transcription of *Ph1* candidate genes is actually up-regulated (6). Thus, the up-regulation of TaASY1 transcription could be explained as the downstream effect by *Ph1* candidate gene over-expression.

1.4 Our hypothesis

The amino acid sequences of Ime2 and CDK2 both share especially high similarity with *Ph1* candidate genes, and HsCDK2 is reported to be the functional homolog of Ime2 in budding yeast. With the hope of utilizing the rich budding yeast research resources to further the *Ph1* locus study, we designed this experiment to test if *Ph1* candidate genes are also homologs of Ime2 as is HsCDK2.

Functional complementation of Ime2 by HsCDK2 was indicated by the up-regulation of Hop1 transcription under vegetative growth condition. Also, the over-expression of *Ph1* candidate genes in bread wheat itself led to the up-regulation of the Hop1 homolog transcription. We therefore asked whether heterologous expression of *Ph1* candidate genes in budding yeast under vegetative growth conditions could up-regulate the transcription of Hop1. This

work was carried out in the laboratory of Prof Luis Aragon, MRC Clinical Science Centre, Hammersmith, London.

2. Results

2.1 Cdc28-4 ts mutant complementary

Single colonies from Cdc28-4 ts mutant strain transformed with either Ph1-B2 or Ph1-D2, together with Cdc28-4 ts mutant strain and WT strain were streaked out on URA- plates using either glucose (GLU) or galactose (GAL) as the sugar source. All the plates were divided by a cross into four even sectors with the same layout as below: top sector for Cdc28-4 ts mutant strain, left sector for Ph1-B2, right sector for: Ph1-D2 and bottom sector for wild type strain. All the plates were first grown at 25C for 4 hrs to induce the GAL1 promoter leading transcription and then delivered to the four test temperatures (25C, 30C, 32C, 34C) and grow for 5 days. The final results are shown in Fig. 7. Wild type strain grew well at all temperature on both GLU and GAL plates as expected. Cdc28-4 ts mutant strain grew well at 24C and 30C, showed less vigour at 32C and totally stopped growing at 34C on both GLU and GAL plates. Cdc28-4 ts mutant strains that contain Ph1-B2 or Ph1-D2 acted the same as Cdc28-4 ts mutant strain itself that they also failed to grow at 34C even on the GAL plates, on which the transcription of Ph1-B2 or Ph1-D2 were inducted. The results suggested that the transcription of Ph1-B2 or Ph1-D2 cannot complement the function of Cdc28 in the cdc28-4 ts mutant strain.

The budding yeast CDK complementation assay is a routine method used to identify and/or confirm CDKs from other organisms. There are a number of different budding yeast lines that contain a malfunctioning Cdc28 locus and can serve as the loss of function strains under certain conditions. The complementation effects of the same external construct varied among different Cdc28 mutant strains. For example, mammalian CDK2 was reported to complement Cdc28-13 ts mutant well but not the Cdc28-4 ts mutant (150,151). Therefore, the failure of the *Ph1* candidate genes to complement Cdc28 in budding yeast could be due to the simple fact that the *Ph1* candidate genes are not the homologues of CDK, especially considering that there are no obvious

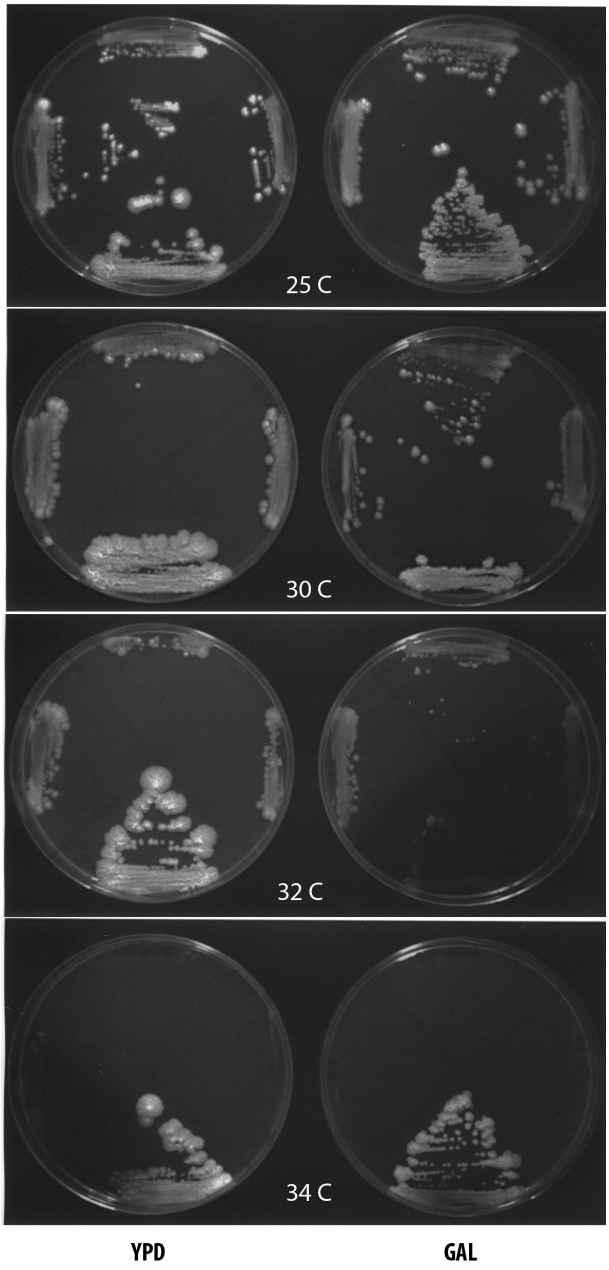


Figure 7 Cdc28-4 ts mutant complementation assay.

Plates were put at 25C for 4hr to induct the GAL1 promoter and then delivered to the four test temperatures and grown for 5 Days. From the top one, clockwise: Cdc28-4, TaCDK-D2, Wild type, TaCDK-B2.

PASTRI sites in the *Ph1* candidate genes. Or this failure could still possibly be due to the choice of Cdc28 mutant strains. So far, the reported successful Cdc28 complementation by plant CDKs were normally not achieved with the Cdc28-4 ts mutants (132,152,153). Use of other Cdc28 mutant strains like Cdc28-13 or Cdc28-1N might lead to different results.

2.2 Ime2 complementation under vegetative growth conditions

2.2.1 Genome DNA contamination level.

Column based (a, b) or Tri-agent based (c,d) RNA extraction methods together with genome DNA digestion by different DNases (Ambio Turbo for a,c; Roche DNase for b,d) were used to prepare RT- samples. Genomic DNA contamination level of these RT- samples is shown at Fig. 8A. No matter which of the two DNases were used, Tri-agent based RNA extraction method always generated RT- samples that contain only negligable (<2%) DNA contamination. On the other hands, column based method always generated RT- samples that were highly contaminated by genomic DNA that could not be efficiently digested by either DNase. Therefore, Tri-agent based RNA extraction method followed with Roche DNase (economic consideration) digestion was used as the routine method to generate total RNA samples.

2.2.2 UBC6 was chosen as reference gene

Typical household genes like Act1 or 18S rDNA have often been used as reference genes in quantitative RT-PCR (QRT-PCR) analysis. However, genome wide budding yeast transcriptomee studies under various conditions using micro-array or other high throughput methods have suggest some novel genes whose transcription was more consistent from conditions to conditions (154). We measure the transcription level of Snr17 and UBC6. The results showed that the transcription of UBC6 is much more stable than that of Snr17 among different samples (Fig. 8B, C). Therefore, UBC6 was used as the reference gene to quantify the transcriptional level of our genes of interest.

2.2.3 Transcription of Ph1 candidate genes was efficiently induced

Transcriptional levels of *Ph1* candidate genes were tested by QRT-PCR. Results showed that both Ph1-B2 and Ph1-D2 constructs containing strains showed high level of *Ph1* candidate gene transcription after the adding of Galactose, which means that the transcription of *Ph1* candidate genes were efficiently induced by Galactose (Fig. 8D).

2.2.4 The transcription of early meiosis genes were not affected with the induction of Ph1 candidate genes.

The transcription levels of Hop1 in different strains before and after the adding of Galactose were tested by QRT-PCR (Fig. 8E). When glucose was used as sugar source, the transcription level of Hop1 remained low and was consistent among Ime2, Ph1-B2, Ph1-D2 or pRS426 containing strains. When galactose was used as sugar source to induce the transcription of GAL1 promoter, the transcription level of Hop1 in Ime2 containing strain was significantly up-regulated (more than two fold), while the transcription level of Hop1 in Ph1-B2 and Ph1-D2 vectors containing strains remains comparable with that in pRS426 containing strains. These results showed that the transcription of Hop1 is only up-regulated in the Ime2 over-expression strain, but not in the Ph1-B2 or Ph1-D2 over-expression strains.

We further investigated the transcription of Spo13, another typical early meiosis gene, with the same experiment setting. Again, the transcription of Spo13 was not up-regulated by the over-expression of *Ph1* candidate genes under vegetative growth condition (Fig. 8F).

It was reported that the episomal expression of Ime2 increased chromosome recombination frequency under vegetative growth condition, in a Spo11 dependent manner (155). We initiated a collaboration with DrAlastair Goldman's group from the University of Sheffield, and investigated whether the episomal expression of *Ph1* candidate genes could increase the chromosome recombination frequency under vegetative growth condition. Again the episomal expression of *Ph1* candidate genes failed to make any noticeable difference compared with that of empty pRS426 vectors.

All these finding suggested that unfortunately *Ph1* candidate genes could not complement the function of Ime2 in budding yeast. Therefore we transferred our efforts to explore the effect of *Ph1* candidate genes in other model systems.

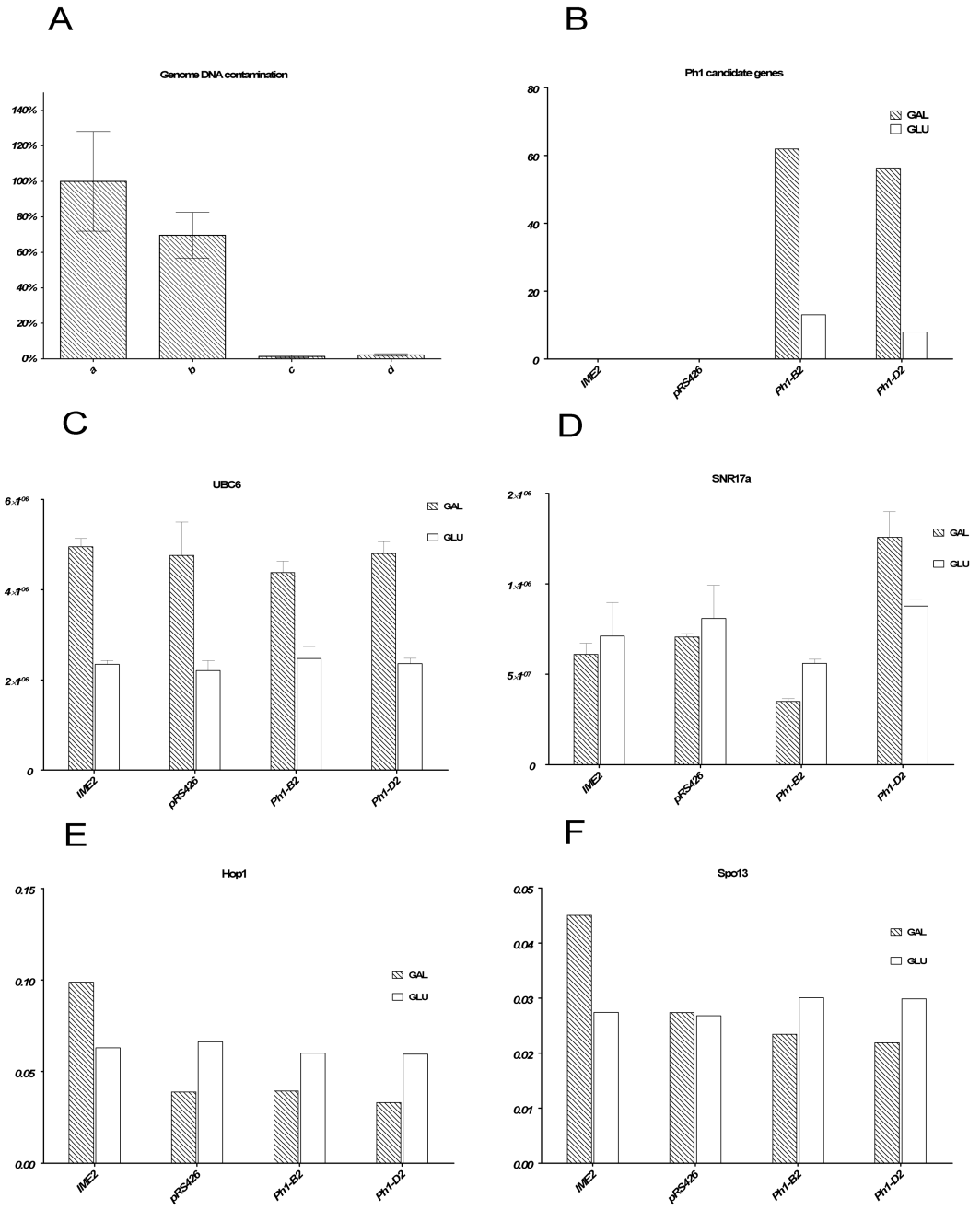


Figure 8 Ime2 complementation test.

Transcription of early meiosis genes was not up-regulated by the over-expression of *Ph1* candidate genes.

3. Materials and methods

3.1 Strains and constructs

Strain W303, from the Aragon lab, was used for Cdc28-4 mutant complementary assay. Strain SK1, was used for Ime2 transcriptional level assay.

Shuttle vectors pRS416 and pRS426 from the pRS4xx vector series were used as backbones to generate the desired constructs if not specifically defined. Both vectors pRS416 and pRS426 employ URA3 as the selection maker. pRS416 is a centromeric plasmid (YCp) that carries CEN and ARS elements. 1 ~ 3 copies per haploid yeast cell.

pRS426 is an episomal plasmid (YEp), that carries REP3 and FRT elements for high copy propagation. ~ 20 copies per haploid yeast cell.

pIME2:TaCDKB2 and pIME2:TaCDKD2 were ordered from Eurofins MWG Operon. The synthesised coding region DNA sequence is inserted into pRS416 vector by XhoI and EcoRI sites.

For the pGAL1: TaCDKB2 and pGAL1:TaCDKD2 constructs, GAL1 promoter was generated by PCR with KpnI tag at the 5' end and HindIII tag at the 3' end. The PCR products were inserted into the pCR-Blunt II-TOPO vector by the pCR-Blunt II-TOPO cloning kit. GAL1 promoter DNA sequence with KpnI and HindIII sticky ends were generated by duo enzyme digestion and ligased with pRS416 vector arms with the same sticky ends, which were generated by KpnI and HindIII duo enzyme digestion from pIME2:TaCDKB2 and pIME2:TaCDKD2 constructs.

For the pGAL1: Ime2 and pIME2:Ime2 constructs, Ime2 coding region DNA sequence was generated by PCR with NotI tag at the 3' end. The PCR products were inserted into the pCR-Blunt II-TOPO vector by the pCR-Blunt II-TOPO cloning kit. Ime2 coding region DNA sequence with HindIII and NotI sticky ends were generated by enzyme digestion and ligased with pRS416 vector arms with the same sticky ends, which were generated by HindIII and NotI enzyme digestion from pGAL1:TaCDKD2 and pIME2:TaCDKD2 constructs.

All the constructs generated in this study were confirmed by dual enzyme digestion and sequencing.

3.2 Galactose induction

- Pick single colonies from SC Glu –Ura plates, and make starter media with 10ml SC Raf –Ura media, shake for overnight.
- Dilute the starter media into 100ml SC Raf –Ura media to a OD600=0.2 (~ 30 folders), grow until OD_{600}=0.6 (exponential phase, ~3hr)
- When OD_{600} reach 0.6, take 10ml media as control, then add 5ml 40% Galactose and start time counting.
- Take 10ml media as 1hr sample.

3.3 Transformation of S. cerevisiae

- Make the following solutions:

10× LiOAc stock solution

LiOAc	5.101 g
1M Tris–HCl (pH7.5 or 8)	5 ml
0.5M EDTA	1 ml
Water	to 50 ml

Note: Filter sterile

LiOAc working solution

10× LiOAc stock solution	10 ml
Water	to 100

Note: Can be stored at room temperature for several months

PEG

Resolve 40g PEG3350 (Sigma)	in 90 ml
10× LiOAc stock solution	10 ml

Note: Can be stored at room temperature for several months

- 50ml Saccharomyces cerevisiae cells were harvested at 0.7 OD_{600} (500g, 5', RT),
- wash with 25ml sterile water,
- then with 5ml LiOAc working solution,

- re-suspended in 400µl LiOAc working solution, aliquot 50µl for 8 tubes,
- add 5 µl plasmid DNA and 6µl DMSO, mix gently and thoroughly
- add 300µl PEG, mix gently and thoroughly, incubated at 30C for about 30',
- then at 42C for exactly 15'.
- spin down the cells (2000 rpm, 3'), and re-suspended the cells in 150µl -Ura medium, plate evenly onto a –Ura plate, put in 30C incubator.

3.4 Q-PCR assay

3.4.1 RNA extraction

- For Saccharomyces cerevisiae culture, resuspend cell pellet with 350 µl of TRI Reagent and add 300 mg of acid washed 600nm glass beads (Sigma, G8772) in a screw capped 2 ml tube. Vortex at 1600 rpm for 2 x 5 min to break the cells.
- For plant tissue, grind 100mg of plant tissue thoroughly with a mortar and pestle in a 2 ml tube with liquid nitrogen. Allow the liquid nitrogen to evaporate, but do not allow the tissue to thaw.
- Then follow the steps afterwards.
- Add TRI Reagent to 1 ml, incubate for 10 min at RT, then transfer the lysate to a new tube, spin at 12,000g x 10 min, 4 °C to remove the debris and/or high molecular mass DNA.
- Add 0.1 ml of 1-bromo-3-chloropropane (BCP). Cover the sample tightly, shake vigorously for 15s, incubate for 10 min at RT, and spin at 12,000g x 10 min, 4 °C to separate the mixture into 3 phases: a red organic phase (containing protein), an interphase (containing DNA), and a colorless upper aqueous phase (containing RNA).

Note: Bromo-3-chloropropane is less toxic than chloroform and its use for phase separation decreases the possibility of contaminating RNA with DNA. The chloroform used for phase separation should not contain isoamyl alcohol or other additives

- Transfer the aqueous phase (about 430 µl) carefully to a fresh tube and add 0.5 ml 2-propanol (Sigma, I9516, or ACS grade) and mix, incubate for 10 min at RT, and spin at 12,000g x 15 min, 4 °C to form the RNA pellet.

Note: Store the interphase and organic phase at 2–8 °C for subsequent isolation of the DNA and proteins.

- Remove the supernatant and add 0.9 ml of 75% ethanol. Vortex gently to wash the RNA, then spin at 12,000g x 5 min, 4 °C to re-settle the pellet.

Note: Store the interphase and organic phase at 2–8 °C for subsequent isolation of the DNA and proteins.

- Remove most of the supernatant with 1 ml tips, spin at 12,000g x 5 min, 4 °C, and then remove the remaining trace liquid carefully with 10 µl tips. Briefly air-dry the RNA pellet for 5 min. Resolving the pellet with 60 µl super-water for exactly 2 min at 55°C, then put on ice for Nano drop.

Note: Remaining ethanol may interfere with the downstream applications, so make sure all the ethanol is gone before resolving the RNA. However, over dry the RNA pellet will greatly decrease its solubility.

- Measure the RNA concentration by Nanodrop.

3.4.2 Genome DNA contamination digestion

- Make the digestion mixture as below:

10x buffer	6 µl
DNAse I (Roche, 04716728001, stores item or	6 µl
Total RNA	12 µg
H2O	to 60 µl

- RT for 25 min, then add 60µl Acid-Phenol:Chloroform (Ambion, AM9720), vortex for 15 min and incubate for 10 min at RT, and spin at 12,000g x 10 min, 4 °C to separate the mixture into 3 phases,
- Transfer the aqueous phase carefully to a fresh tube, mix with 17µl 8M LiCl (Sigma, L7026) and let RNA precipitate overnight at 4 °C.
- (Day 2) Spin the mixture above at 12,000g x 20 min, 4 °C to form the RNA pellet.
- Remove the supernatant and add 1 ml 75% ethanol (ACS grade ethanol mix with super water). Vortex gently to wash the RNA pellet, then spin at 12,000g x 5 min, 4 °C to re-settle the pellet.
- Remove most of the supernatant with 200 µl tips, spin at 12,000g x 5 min , 4 °C, and then remove the remaining trace liquid carefully with 10 µl tips. Briefly air dry the RNA pellet for 5 min. Resolving the pellet with 25 µl super-water for 2 min at 55°C, then put on ice for Nano drop.

Note: Remaining ethanol may interfere with the downstream applications, so make sure all the ethanol is gone before resolving the RNA. However, over dry the RNA pellet will greatly decrease its solubility.

- Measure the RNA concentration by Nanodrop, also run samples on agar gel.

3.4.3 cDNA synthesis

- Make the primer annealing mixture as below:

DNase treated RNA, from above	2µg
Primer Mix (0.25µg/µl Oligo-dT and 0.15µg/µl Random Primers)	1µl
10mM dNTPs	1µl
Distilled water	to 12µl

- 65 °C 5 min then on ice whilst you prepare the reverse transcription master mix

3.4.4 Reverse Transcription

Primer-annealed RNA from above	12µl
5x First Strand Buffer	4µl
0.1M DTT	2µl
RNA-guard*	1µl

- Divide the mixture into two tubes, each 9.4µl, then add 0.45µl Superscript III Reverse Transcriptase (Invitrogen, Cat. No. 18080-044) for the RT+ (sample) tubes, and 0.45µl water for the RT- (control) tubes. Incubate at 42°C for 50 min, then 70°C for 15 min (use PCR machine). Use directly for Real-Time PCR or store at -80°C immediately.

3.4.5 Real-time PCR

- Make the primer mixture as below:

Master Mix	2.5µl
Primers (15µM pair)	0.2µl

- Mix the same primer sets in one 1.5ml tube then add 2.7µl mixture to the plates by 2-20µl electronic pipette.
- Then add 2.3µl cDNA (1:3 diluted), and seal the plate with film applicator.
- Spin at 1500g for 2 min before putting the plate in the Real-Time PCR machine (LightCycler® 480 System, Roche Applied Science Ltd.)
- Run cycles at below:

Program Name	Cycles	Analysis Mode
Preincubation	1	None
Amplification	45	Quantification
Melting curve	1	Melting Curves

	Target (°C)	Acq Mode	Time	Ramp	Acq (per °C)	Sec Target	Step Size	Step Delay
Preincubation	95	None	2 min	4.8	5	0	0	0
Amplification	95	None	25 sec	4.8	5	0	0	0
	58	None	25 sec	2.5	5	0	0	0
	72	Single	40 sec	4.8	5	0	0	0
Melting curve	95	None	5 sec	4.8	5	0	0	0
	65	None	1 min	2.5	5	0	0	0
	97	Cons	0 sec	0.11	5	0	0	0

3.4.6 Q-PCR data analysis

The raw (not baseline-corrected) PCR data was transferred into LinRegPCR (156) software. The baseline fluorescence and PCR efficiency of individual PCR reaction were calculated based on the kinetic model of each PCR reaction in this software. The mean PCR efficiency of each amplicon, C_t threshold and C_t value of each PCR reaction were then calculated based on the baseline fluorescence and individual PCR efficiency. The start concentration value (N0) of each well was calculated based on the mean PCR efficiency and C_t value finally. N0 values were then transferred into Excel 2003 (Microsoft, USA) software to calculate the relative cDNA amounts of each sample (normally containing three biological replicates) by dividing the mean N0 values of target gene cDNA with the mean N0 values of reference gene cDNA. The fold changes of expression increase/reduction were calculated by dividing the relative cDNA amounts of two target genes. The standard error of mean (SEM) values were given by Excel 2003 software. The combining of SEMs in the equation a=b/c or a=b*c was calculated based on the equation: $(\Delta a/a)^2 = (\Delta b/b)^2 + (\Delta c/c)^2$, Δ=SEM.

CHAPTER IV Meiosis studies in Brachypodium

1. Background

1.1 Brachypodium as a model system

Brachypodium distachyon is a *Pooideae* sub family wild grass native to Mediterranean and Eurosiberian areas. The main reason that we chose *B. distachyon* as a model system to study the mechanism of the *Ph1* locus is that close homologues of *Ph1* candidate genes, Ph1-CDK*s*, were found to be present in the genomes of *Brachypodium* species but not other two model plants - *Arabidopsis thaliana* and *Oryza sativa*. This is why a *Pooideae* model plant was needed when a *Poaceae* model, *O. sativa,* was already established with rich research resources already available. In fact *B. distachyon* is a popular newly emerged model plant not only because it is genetically close to temperate cereal and forage grasses, but also because there are many biological characteristics which make *B. distachyon* a good model organism for research on its own merits.

A small genome size is the first reason that *B. distachyon* was chosen to be the *Pooideae* model. Most *Brachypodium* species have genome sizes around 300-400Mb/1C, which are similar to that of rice and are the smallest among all the Pooideae species (157). *B. distachyon* was selected from the various *Brachypodium* species because it is the only true annual species and has the smallest chromosome number within the genus. The genome size of a diploid plant mainly depends on the abundance of repetitive sequences within its genome. In common with the other two model plants, *Arabidopsis thaliana* and *O. sativa* which have also small genome sizes, *B. distachyon* has a significantly lower repetitive sequence content (less than 15%) than most other plants. The compact genome and low level of repetitive sequence of *B. distachyon* is a valuable trait in a model organism, since it greatly saves time and effort when doing things like genome sequencing, genetic map development, BAC FISH etc.

B. distachyon also has a growth habit and biological characteristics that make it easy to handle for large-scale applications like mutagenesis; *B. distachyon* plants are inbreeding and require only simple conditions and a small space to grow. In a given area *B. distachyon* plants can produce 3-6 fold more seeds than *O. sativa* plants (158). The seed to seed life cycle of *B. distachyon* is only 8-10 weeks, which is similar to the life cycle of *A. thaliana* and about half the life cycle of *O. sativa* (158).

The development of *B. distachyon* as a model system has had much input from research communities all over the world, which has led to the rapid development of genetic and genomic research resources such as efficient transformation (159,160), BAC libraries (161-163), genetic linkage map (164), physical maps (165), mutant collections (166) ,and, importantly, the completely sequenced and annotated genome together with various bioinformatics databases (http://www.brachybase. org, http://www.phytozome.net, http://www.modelcrop.org).

However since *B. distachyon* was a newly emerged model plant that received very little previous attention, a number of techniques needed first to be developed to begin the study of meiosis in *B. distachyon*.

1.2 Centromere structure and regulation

1.2.1 Centromere general speaking

As mentioned in CHAPTER I, in order to understand the *Ph1* locus, a great deal of effort has been focused on the study of chromosome dynamics and the effect of the *Ph1* locus on it during early meiosis in wheat. Both telomeres and centromeres are undergoing dynamic changes at the time of meiotic chromosome pairing in wheat. In order to investigate any potential effect of the *Ph1* locus in Brachypodium, it was important to first describe chromosome dynamics during early meiosis in Brachypodium. Then the possible effect of *Ph1* locus on these dynamics could be determined. Most plants including Brachypodium share an identitical 7 bp telomere tandem repeat sequence (TTTAGGG). The protocol for labelling such a telomere sequence was well

established. Brachypodium telomeres were successfully labelled in preliminary experiments using this sequence. However, the centromere probe CCS1 that was previously used in our group to label cereal centromeres failed to label Brachypodium centromeres in the preliminary experiments. To find a suitable probe for centromere labelling and also to better understand the results, it is important to understand the structure of cereal centromeres.

Centromeres are fundamental eukaryotic chromosome components that are responsible for essential cell division processes like chromosome segregation and sister chromatid cohesion. The word centromere refers to the complex constituted by the protein structure, the kinetochore, and the centromeric DNA that binds the kinetochore. The kinetochore binds tightly to centromeric DNA based on sequence and/or epigenetic markers. During mitotic or meiotic cell division, spindle fibres attach to the kinetochores and pull sister chromatids towards opposite poles while sister chromatid cohesion binds to both sister chromatid and holds them together. This counteraction of spindle and sister chromatid cohesion helps to align chromosomes along the metaphase plate and ensures the faithful segregation of sister chromatids.

Centromeres of most eukaryotic species are restricted to a specific and narrow locus named the primary constriction. This type of centromere is defined as a monocentric centromere, to differentiate it from the uncommon cases of dicentric, polycentric or holocentric centromeres with which two loci, multiple loci or the entire chromosome acts as the centromere.

The main structures of centromeres are in general conserved among distinct lineages. Briefly, a typical centromere is composed of a centromeric core sequence, the pericentromeric heterochromatin surrounding the core sequence and the centromeric specific proteins that binds to the centromeric DNA. However the sequence, size and composition of centromeric DNA vary significantly among species of different lineages, or for the case of centromeric satellites, vary dramatically even among closely related species. Here only the details of the cereal centromere structure will be described.

1.2.2 Cereal centromeric DNA composition

The investigation of cereal centromeres was started more than 10 years ago with the finding of the first cereal centromeric sequence CCS1. CCS1 is a 249 bp centromeric sequence isolated from *Brachypodium sylvaticum* and homologous sequences were found among various cereal species including brachypodium, wheat, barley, rye, maize and rice (168). Soon after the finding of CCS1, another cereal centromeric sequence named pSau3A9 was reported. pSau3A9 was isolated from *Sorghum bicolor*, and homologous sequences were found in rice, maize, wheat, barley, rye, and oats (169). It was shown by FISH labelling that both CCS1 (and its corresponding homologues) and pSau3A9 are located specifically at the centromeric regions of different cereal chromosomes, which confirmed them to be centromeric sequence. It was also found that both sequences are cereal specific and are not present in the dicot species. Further analysis of rice and barley centromeric sequence revealed that both CCS1 and pSau3A9 are parts of a specific Ty3/gypsy retrotransposon family (170,171). Unlike usual retrotransposon families that have diverged rapidly during evolution, members of this particular retrotransposon family were found to be present in various cereal genomes with highly conserved motifs in the long terminal repeats (LTRs) and primer binding site (PBS) (172). FISH labelling of this retrotransposon family produced centromere specific or centromere enriched signals in all the cereal species investigated, which indicates that this is a common cereal centromeric retrotransposon. This particular retrotransposon family is now commonly referred to as the cereal centromeric retrotransposons (CCRs).

Centromeric satellite DNA is another important centromeric DNA element. Unlike CCRs, which are conserved among cereal species, the sequences of centromeric satellites have diverged significantly between species from different cereal genera, although the size of the satellite unit is often found to be in the range of 150-180 bp (). The centromeric satellites from close related species within the genus of *Oryza* have been identified (173-175). It was shown by these studies that the 155/165 bp centromeric satellite CentO was presented in all the *Oryza* species studied but not in the wild rice *Zizania palustris* (174). Further

analysis showed that although CentO is the universal centromere satellite enriched in each of the 12 chromosomes of *O. sativa* (AA, 2n = 24) and *O. punctata* (BB, 2n = 24) (173,176), in *O. officinal* and *O. rhizomatis* (both CC, 2n = 24) it is much less abundant and is enriched only at centromere of chromosome 7 (175). In contrast, another 126 bp centromeric satellite CentO-C was found to be the dominant centromere satellite repeat that is exclusively enriched in most of the centromeres (175). Such an intra-genus diversity of centromeric satellite was also reported in other genera including *Arabidopsis* (177), *Brassica* (178) *Trifolium* (179)and *Coix* (180), which suggests it is a common scenario for all plants. The diversity of centromeric satellite is not limited to centromeric satellites of different types, but also exists among monomers of the same centromeric satellite type. Detailed analysis of *O. sativa* chromosome 4 (181) and chromosome 8 (182) centromeres showed that the overall identity of CentO repeats is more than 90%. Sequence variation tends to happen randomly at a low frequency in most sites, however with a preference to several specific sites, and is absent from certain conserved sites or domains. Interestingly, 6 out of 8 highly variable sites of chromosome 4 were clustered at the 5' region of the monomers, which is similar to the DNA element interacting with centromeric proteins Cse4p and Mif2 in budding yeast (182).

It has long been known that cereal centromeres are mainly constituted by centromeric satellite and various TEs (Transposable Elements) including the CCR. However the detailed arrangement of these elements within the centromeres was only discovered recently by deep sequencing and alignments of the whole centromere region with several chromosomes of rice and Maize (172,181,182). It was shown by these studies that, in a typical cereal centromere, a few to hundreds of satellite monomers align head-to-tail to form satellite tracts that are separated by clusters containing mainly TEs. Although the orientations of the monomers within the same tract are the same, orientations of different tracts can be different. TEs within the cereal centromeric region are mainly members of the CCR family mentioned before, although species specific retrotransposon and other types of repetitive sequences are also commonly found in the cereal centromeric regions with a lower abundance. It was reported previously that in *A. thaliana* the distribution of dominant centromeric satellites

pAL1 and major centromeric retrotransposon Athila is distinct, with pAL1 mainly locate at core centromeric regions and Athila at pericentromeric regions respectively (183,184). Together with the fact that pAL1 is the only major centromeric DNA that binds CENH3 in the CHIP assay (185) this suggested that centromeric satellite is the only dominant core centromeric sequence in *A. thaliana*. However no obvious difference was found in the distribution of centromeric satellite and CCR in the cereal centromeric regions. The mixture of centromeric satellite and TEs are features for both core centromeric regions and peri-centromeric regions in cereals. It was also shown that both centromeric satellite and CCR strongly binds to CENH3 in the CHIP assay, with no obvious difference in terms of abundance. Taking together all the evidence described, it is clear that unlike *A. thaliana*, both CCR and centromeric satellites are crucial components of core centromeres in cereal. It has been suggested that the unique composition of cereal core centromere sequences represents their status as newly formed centromeres, which is partially supported by the finding that active genes exist in the centromeric regions of rice chromosome (186).

Centromeric DNA is mainly constituted by unique centromeric specific sequences that are located at defined loci in cereal and also other eukaryotic species. Such a non-random composition of centromeric DNA suggests that the primary sequence of the centromeric DNA has a role in centromere function and specification. Indeed it has been reported in *S.pombe* (187) and Humans (188) that functional centromeres can be re-established by simply reintroducing certain centromeric DNA fragments into the cells, which showed that the primary sequence of centromeric DNA itself is able to trigger the centromere formation process in certain circumstances.

Apart from the unique but diverse centromeric DNA sequences, eukaryotic centromeres all share a similar centromeric chromatin architecture characterised by the centromeric H3 variant CENH3 (CENP-A in mammals) and various specific epigenetic marks. CENH3 is a specific variant of H3 that shares conserved histone fold domains with canonical H3 but has a unique and highly variable N-terminal tail. Because of its universal crucial role in all eukaryotes, CENH3 is considered as the hallmark of centromere. It is specifically loaded on

functional centromeres of both higher plants and other eukaryotes and is absolutely required for centromere function during chromosome segregation. Furthermore the centromere loading of CENH3 is not affected by the deletion of many other centromeric proteins (189); in contrast the loading of other centromeric proteins is dependent on the presence of CENH3 (190). In all eukaryotes, CENH3 is loaded at the core centromeric region and interspersed by canonical H3 (except for S. cerevisiae where a single CENH3 nucleosome functions as the core centromere). Euchromatin markers H3K4me2 or H3K4me were found to be enriched at core centromeric canonical nucleosomes with the absence of heterochromatin marker H3K9me2 in *S. pombe*, *D. melanogaster*, mammals and Arabidopsis. The mosaic mixture of euchromatin-marked canonical nucleosomes and CENH3 nucleosomes define the core centromere that is surrounded by peri-centromeric heterochromatin enriched with marker H3K9me2. Other heterochromatin marks like DNA methylation and heterochromatin protein binding have also been found to be enriched at the peri-centromeric region but absent from the core centromeric region, which further support the idea that core centromeric and peri-centromeric regions contain distinct chromatin features. It was proposed that it is the peri-centromeric heterochromatin that limits the spread of CENH3 and thus core centromeres. However in rice, H3K4me2 was found to be not present in the centromeric region of chromosome 8 in two independent studies; instead it was the H3K9me2 that was enriched in the entire centromeric region. It is unclear what the reason is for this and the consequence for rice to have such a unique core centromeric chromatin.

There is now accumulating evidence showing that between centromeric DNA and epigenetic regulation, the latter is more essential for centromere function and specification. For instance it was reported that neocentromeres (newly formed centromeres) can be formed at previously non-centromeric loci without any conventional unique centromeric DNA (191). It was also reported that centromeres with same centromeric DNA sequences can be in either active or inactive modes (187,192,193), and the conversion from one mode to the other is controlled in an epigenetic pathway without altering the primary sequence of the centromeric DNA (187). Another obvious piece of evidence is that although

species specific centromeric satellites are major components of core centromeres in a wide range of eukaryotic species including *Drosophila* species, mammals and higher plants, in all the centromeric satellite carrying species investigated, only a subset of the whole centromeric satellite population were found to be located in the core centromeric region and loaded with CENH3. The distinction of core or peri- satellite is epigenetically dependent rather than dependent on primary sequence.

1.3 Site specific BAC clone screening

Apart from the centromere and telomere dynamics, the dynamics of chromosome arms, especially the chromatin conformation change during meiosis has been of special interest in the study of the *Ph1* locus in wheat recently. As mentioned in CHAPTER I, it was shown by labelling specific rye sub-telomeric heterochromatin regions that synchronised chromatin unfolding of heterochromatin at the onset of meiosis is strongly correlated with chromosome pairing and is tightly controlled by the *Ph1* locus in the wheat background (57,92). It would be interesting to answer questions such as whether such a chromatin conformation change is universal for all chromatin regions or is limited to certain chromatin regions and if such a chromatin conformation is globally controlled by the *Ph1* locus by investigating the global chromatin conformation changes at the onset of meiosis.

BAC FISH is by far the most convenient and practical method to visualise specific chromosomal loci *in situ*. It has been shown previously that it is possible to study global chromatin dynamics (194) as well as local chromatin conformation changes (195,196) by analyzing distance between two or more BAC FISH labelled chromatin loci with known genome positions. However, since wheat has a genome highly entiches with repetitive sequences , wheat BAC clones are very likely to contain repetitive sequences that are also widely dispersed in the wheat genomes, therefore it is difficult if not impossible to find site-specific BAC clones in wheat. This makes the use of BAC FISH in wheat highly problematic. In contrast, the genome of *B. distachyon* is highly compact and contains much less repetitive sequence, which make it a good candidate for site specific BAC FISH labelling. We were therefore interested in studying

meiotic chromatin conformation changes in Brachypodium by BAC FISH. The labelling of several randomly picked, site-specific BAC clones was already previously reported for *B. distachyon* ecotype ABR1 and ABR5 (197). However, since the correlation of spatial distance and DNA sequence of two adjacent chromosomal loci is only valid within a certain range (< 2.5 Mb in mammal), and also chromatin dynamics varies at different parts of the chromosomes (heterochromatin versus euchromatin for instance), a set of site specific BACs with known chromosomal positions is necessary for a dynamic study of chromatin conformation on *B. distachyon.* In this study, a novel site specific BAC clone screening strategy combining the use of genome located BAC libraries, a bio-informatics pre-screening and an efficient BAC DNA amplification method was developed to produce such a site set of specific BAC clones. Several site specific BAC clones with known genome positions were identified using this strategy with a high efficiency. The established screening strategy and identified site specific BAC clones provide useful groundwork for the future study of chromatin conformation change in *B. distachyon.* The site-specific BAC clone screening strategy described in this study can also be easily adapted to other model system to establish site specific BAC clone sets for various chromosome behaviour and/or chromatin conformation studies.

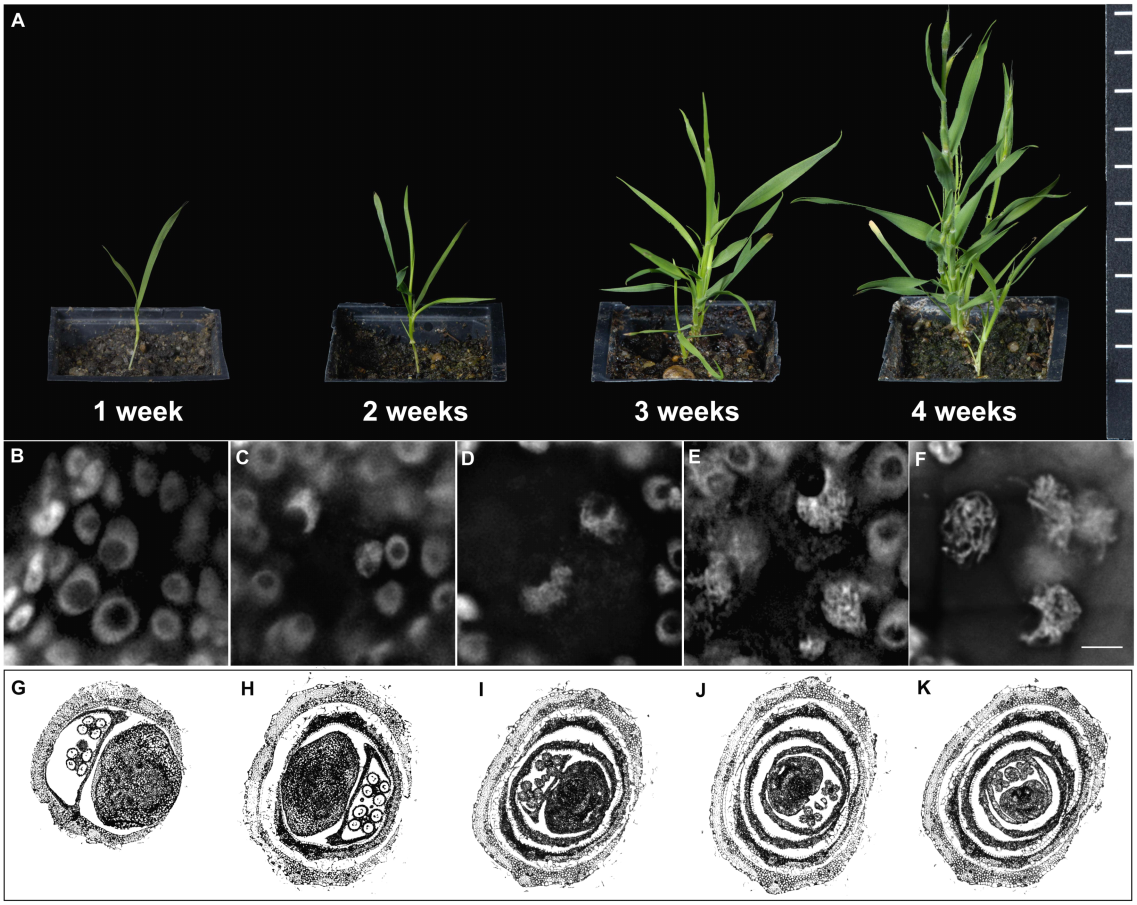
2. Results and discussion

2.1 Sample preparation from young B. distachyon spikelets

Anther squash is a convenient and frequently used technique to prepare PMC (Pollen Mother Cell) samples. However the strong external physical force of squashing modifies the 3D spatial structure of the nuclei and may produce artefacts in the chromatin structures (198). As an alternative, we prefer thin tissue sectioning in our studies of chromosome behaviour because of its good maintenance of nuclear fine structures. Among various tissue sectioning techniques that are available for plant sample preparation, Vibratome sectioning is our preferred choice because of its minimum chemical treatment compared with other tissue sectioning techniques, which helps to produce better FISH

Figure 9 Staging of *B. distachyon* meiosis in PMCs.

A: *B. distachyon* plants at different stages from young seedling to flowering. **B-F**: *B. distachyon* PMC meiocytes at different stages; Bar = 5 μm. **B**: Meiocytes at pre-meiosis stage; Nuclei of PMCs at this stage are large, smooth and round; Nucleoli are intact and located in the centre of the nuclei. **C**: Meiocytes at leptotene stage; Thin chromatin threads start to appear in the nuclei of PMCs at this stage, but in general nuclei still look smooth; Nucleoli moves to the edges of the nuclei, and as the results the nuclei appears as crescent shapes; **D**: Meiocytes at zygotene stage; Chromatin threads become thicker at this stage, which give the nuclei a fluffy appearance; Nucleoli are still at the edges of the nuclei, and some nuclei start to lose their shapes. **E**: Meiocytes at late zygotene stage; Thick chromatin threads become obvious and start to separate at this stage, which leaves holes in the nuclei; Nucleoli are still at the edges of the nuclei. **F**: Meiocytes at pachytene stage; Thick chromatin threads separate completely from each other at this stage; Nuclei mostly appear as irregular shapes and nucleoli are dispersed in most nuclei of this stage. Meiocytes only stay for a very short time (likely within 2 hr or less) at leptotene stage (**C**), and then change from zygotene (**D**) to late zygotene (**E**) stage gradually during a longer period (approximately 10 hr). Meiocytes quickly move from late zygotene (**E**) to pachytene (**F**) stage and stay at pachytene (**F**) for a few hours. After pachytene, it takes meiocytes only about 8hr to go through all the remaining steps until the tetrad stage. **G-K**: Wax embedding sections of one *B. distachyon* spikelet at meiosis. **G** shows the image of the base floret. Images from **G** to **K** show florets in sequence from base to top of the spikelet. Florets on the same spikelet go through meiosis successively. Meiosis starts first in the base floret and last in the top floret with an interval of 8-10 hr. For instance, if meiocytes of the base floret (**G**) were at pachytene stage (**F**), the second from base floret (**H**) would be at about zygotene stage (**D**), and the third from base floret (**I**) would be at pre-meiosis stage (**B**).


A
1 week
2 weeks
3 weeks
4 weeks
B
C
D
E
F
G
H
I
J
K

results afterwards with lower background fluorescence and less unspecific labelling. Vibratome sectioning of wheat young spikelets was established in our lab previously (198,199) and successfully applied in various chromosome behaviour studies in wheat (44,57,87). However, due to the distinct structure of Brachypodium spikelets (Fig. 9C), the contact free florets are very likely to drop out during the process of sectioning, which makes it very difficult, if not impossible to section young Brachypodium spikelets with the original Vibratome sectioning protocol. Various modifications have been tested to adapt the original wheat Vibratome section protocol to Brachypodium study; one of the modifications which was successful is described.

In detail, the outer glume was removed carefully to expose the base floret and the whole spikelet was cut in the middle with the bottom 5 mm left. The trimmed young spikelet with the base floret exposed was glued with super glue vertically on the section block with the base of the spikelet facing up. After the super glue had dried, the whole spikelet was carefully coated with another thin layer of super glue. This super glue coating helped to keep the base floret structure from dropping during Vibratome sectioning, but it would also cause the sections to curve and cause sample loss if dried directly. To solve this problem, 50% (v/v) DMSO/water was dropped directly onto sections on the sample holding glass slides to dissolve the super glue coating. After the 50% DMSO was evaporated, the florets containing sections were attached nicely to the glass slides and were ready to be used in the FISH procedure.

The efficiency of sectioning was improved to an acceptable level with this modification. However, even with the modified protocol, only the base floret of each spikelet is usable since the other, upper florets still dropped without the assistance of super glue during sectioning. Given the sequential arranged meiosis of florets within one spikelet (described at section 2.1 of this chapter), such a limitation loses valuable information for meiosis staging. Moreover, the size of the spikelet at and before the onset of meiosis is very small, which hinders the manipulation of spikelets that is required in the modified protocol and thus reduces the efficiency of sample preparation of pre-meiosis samples. Considering all the drawbacks of the modified Vibratome sectioning protocol,

wax embedding sectioning was tested later to assess its impact on final FISH outputs. No special manipulation of spikelets is needed in the wax embedding section protocol. Spikelets of various sizes can be sectioned continuously from base to top without any loss of samples (Fig. 9G-K). And importantly, the final FISH results of the wax embedding sectioned samples are comparable with those of the Vibratome sectioned samples (Fig. 11). Therefore, although Vibratome sectioning may still have an advantage for other applications like immunostaining, wax embedding sectioning was preferred in this study and is recommended for the sample preparation of young *B. distachyon* spikelets for FISH in future work.

2.2 Staging of B. distachyon meiosis in PMCs

The schedule of plant growth varies significantly under different growth conditions, therefore the staging of Brachypodium meiosis was measured under fixed growth condition specified in the “materials and methods” section, and will only apply to *B. distachyon* plants that are grown under the same conditions.

It is common to stage plant meiosis by staining meiotic chromosomes with Acetocarmine or similar dyes on anther squash samples. However, the anther squash samples of *B. distachyon* failed to be stained properly by any of the dyes tested including Acetocarmine, Carbol fuchsin and DAPI, although they did produce strong and clear staining with wheat anther squash samples (data not shown). Alternatively, *B. distachyon* meiosis was staged with cell wall digested and DAPI stained young spikelet thin sections. The meiotic stage of a given meiocyte was determined based on the chromatin morphology, sizes of the spikelets, the timing of sample collection and also the pattern of labelled centromeres and/or telomeres if available. DAPI stained meiocytes at different meiotic stages are shown in Fig 1 A. In detail, meiosis starts at day 25 (with a variation within two days) after germination and the whole process from early prophase to tetrads lasts for approximately one day. One typical *B. distachyon* plant has several inflorescence holding tillers, 2-3 spikelets per rachis node, 5-10 florets per spikelet and 2 anthers per floret. Meiosis starts first in the tallest tiller, and one day later in the next two tillers. It also starts earlier in the bigger spikelet of the same rachis node. In this study, only the biggest spikelets from

the top three tall tillers of the plants were used. The wax section samples of base to top florets within one spikelet are shown at Fig. 9G-K in order. The meiosis processes were in general synchronised among PMCs within the same floret (although may not be strictly synchronised, see section 2.4 this chapter). For the same spikelet, meiosis starts first in the base floret and a few hours later in the next upper floret and so on. For instance, if the base floret is at pachytene stage (Fig. 9G), the second from base floret (Fig. 9H) would be at a stage around leptotene to zygotene, and the third from base floret (Fig. 9I) at pre-meiosis stage. For a tray of *B. distachyon* plants germinated and grown at the same time, meiosis is in general synchronized among the majority of the plants although it is common to see a few plants starting the process of meiosis one day prior to or after the majority. Since young spikelets are still fully covered by stems when meiosis already starts in the base floret, it is hard to judge when to collect the early meiosis samples based on a single plant alone. As an alternative, newly emerged young spikelets of a few plants beginning meiosis were used as a sign for sample collection on the following day if early meiosis base floret samples were needed for vibratome sectioning. Otherwise, emerged spikelets were collected for wax embedding sectioning, and early meiosis florets could be found in the upper parts of such spikelets.

2.3 Identification of centromeric satellite sequence in B. distachyon

The fully assembled genome sequence of *B. distachyon* was not available until early 2010 , and the centromere sequence was not identified at the time of this study. To identify the centromeric satellite repeat sequence of *B. distachyon*, the reported cereal centromeric sequences CCS1 (Genebank ID: U52217.1 (168)) and pSau3A9 (Genebank ID: U68165.1 (169)) were used to BLAST against 4× coverage preliminary genome sequence assembly at www.modelcrop.org with default setting to identify the centromeric sequence segments containing both homologues of CCS1 and pSau3A9. Two such segments (super_2:2420000-2430000 10 kb, super_3:21200002-21250000 50 kb) were obtained at www.modelcrop.org. A bio-informatics tool (Phobos, Ver 3.2.6) was used to search for tandem repeat sequences within these two segments.

Three long tandem repeat sequences with good scores were found in the super_3 segment. Interestingly, these three sequences were highly similar both in size (156 bp, 157bp, 158 bp respectively) and in sequence (Fig. 10D) with each other. Consensus sequence of these three was used to BLAST against 4× coverage preliminary genome sequence assembly again. Highly conserved sequences of these three were found in many super-contigs with a uniform pattern as expected for centromeric satellite repeats. These 156 bp sequences were therefore suspected to be the centromeric satellite repeat sequences of *B. distachyon.*

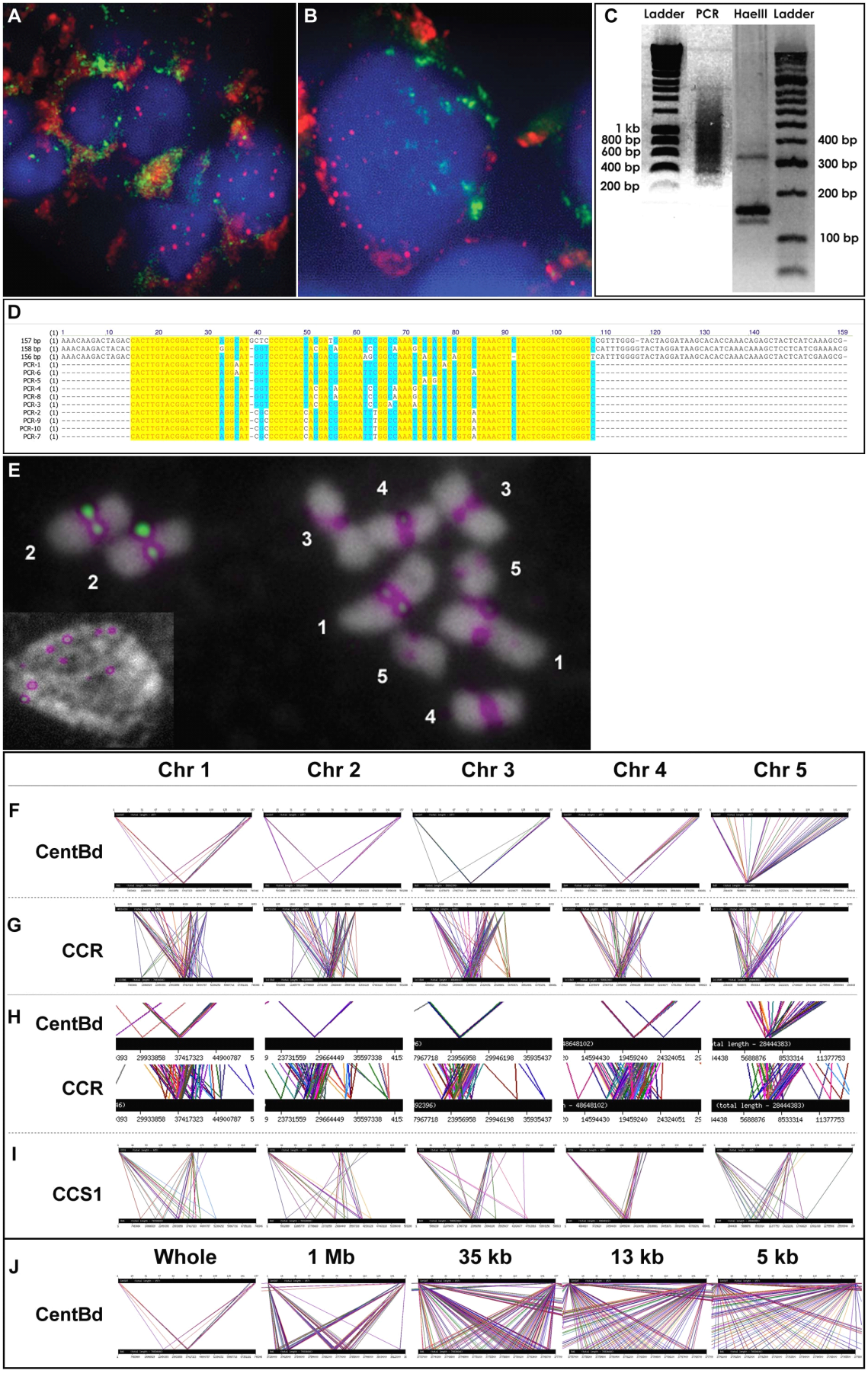
PCR primers selected from the conserved regions were used to amplify the suspected centromeric satellite repeat from *B. distachyon* genome DNA. The PCR products appeared as DNA ladder on gel with an interval of about 156 bp as expected (Fig. 10C). After fully digesting the PCR products with HaeIII restriction enzyme, one strong band and two faint bands were found on the agar gel. The strong band was located at approximately 156 bp, with one faint band being slightly lower next to it. These two bands represent satellite monomers with two different sizes and the majority of the population is the 156 bp one. There was another faint band located at about 310 bp, which likely represented the mutated HaeIII recognition sites in a small portion of the satellite monomers (Fig. 10C). The PCR products were then cloned and sequenced. Sequencing data showed that the obtained PCR products were highly conserved with the suspected satellite sequence with the same pattern of conserved and variable regions.

To further check if these sequences are specifically located in centromeric regions, the above PCR products were used to label metaphase chromosomes by FISH (Fluorescent In Situ Hybridization). The signals specifically covered centromeric regions of all the five chromosomes (Fig. 10E), and thus the identified tandem repeats were confirmed as the centromeric satellite sequence of *B. distachyon* and named as CentBd. It was also found by the FISH labelling that the abundance of the CentBd repeats varies significantly among different chromosomes, with chromosome 5 containing only a tiny amount of CentBd (Fig. 10E). The verified CentBd sequence was analyzed together with other

Figure 10 Identification and analysis of B. distachyon centromeric satellite sequence.

A-B: Dual labelling of telomere (red) and CCS1 sequence (green) on B. distachyon (A) and T. aestivum (B) somatic interphase nuclei using the same probe set; Telomere probe labelled well on both B. distachyon and T. aestivum nuclei, which verified the quality of the FISH procedure itself; in contrast, CCS1 probe produced clear specific labelling only in T. aestivum but not in B. distachyon nuclei. C: PCR amplification of the presumptive B. distachyon centromeric satellite sequences from genome DNA, and the HaeIII digestion of the PCR product; PCR product appeared as a ladder on the agar gel with an interval of around 156 bp, which was as expected from a tandem repeat with a unit size of 156 bp. A unique HaeIII recognition site was found with the monomers of most presumptive B. distachyon centromeric satellite repeats; after the full digestion of HaeIII, the PCR product mentioned above appeared as three bands on the agar gel. A strong band was located at approximately 156 bp, with another faint band being slightly lower next to it. These two bands represent satellite monomers with two different sizes and the majority of the population is the 156 bp one. There is another faint band located at about 310 bp, which likely represent the mutated HaeIII recognition sites in a small portion of satellite monomers. D: Alignment of presumptive B. distachyon centromeric satellite sequences with 10 actual sequences of PCR products. As shown by the picture, the presumptive sequences and actual PCR product sequences were highly similar in both sequence and structure with each other, which suggested that the PCR product were indeed amplified from the presumptive satellite sequences. E: FISH labelling of the presumptive B. distachyon centromeric satellite sequences onto metaphase chromosome spread by FISH. As shown by the picture, presumptive sequences were specifically located at centromeres of all chromosomes, which confirmed these sequences as the true centromeric satellite repeat sequences. It was also shown by the labelling that the abundance of the B. distachyon centromeric satellite (CentBd) repeats varies significantly among different chromosomes, with chromosome 5

(arrow head) containing the smallest amount of CentBd. F-J: Bioinformatics analysis of B. distachyon centromeric regions. Centromeric sequences were BLASTed against the assembled B. distachyon genome sequence to show their distribution within the genome. Each small picture shows the BLAST result of one particular centromeric sequence against one particular chromosome. In each small picture, the upper bar represents the full length of the query sequence and the lower bar represents the full length of one chromosome. The left and right borders of one homologous domain were linked by a pair of lines with the same color. F: Distribution of CentBd homologous sequences on each chromosome. CentBd homologous regions were enriched at a few loci on each chromosome, with one major locus per chromosome containing thousands of CentBd homologous regions. G: Distribution of S. bicolor CCR homologous sequences on each chromosome. On each chromosome, CCR homologous sequences were enriched in large numbers at one major locus. H: The magnified images of F and G at the same narrow centromeric region of each chromosome. The CentBd enriched loci were surrounded by an enrichment of CCR homologous sequences. I: Distribution of CCS1 homologous sequences on each chromosome. The CCS1 homologous sequences are comparatively less abundant. CCS1 homologous regions were enriched at the centromeric region of chromosome 3 and 4, but were more dispersed on chromosome 1, 2 and 5. J: Serial magnified pictures of the major CentBd enriched locus of chromosome 1. The major CentBd enriched locus contains two CentBd blocks that are 300 kb from each other. The CentBd blocks are about 100 kb in size and are formed by tandemly arranged CentBd monomers.


A
B
C
Ladder
PCR
HaeIII
Ladder
1 kb
800 bp
600 bp
400 bp
200 bp
400 bp
300 bp
200 bp
100 bp
D
157 bp
158 bp
156 bp
PCR-1
PCR-6
PCR-5
PCR-4
PCR-8
PCR-3
PCR-2
PCR-9
PCR-10
PCR-7
E
Chr 1
Chr 2
Chr 3
Chr 4
Chr 5
F
CentBd
G
CCR
H
CentBd
CCR
I
CCS1
J
Whole
1 Mb
35 kb
13 kb
5 kb
CentBd

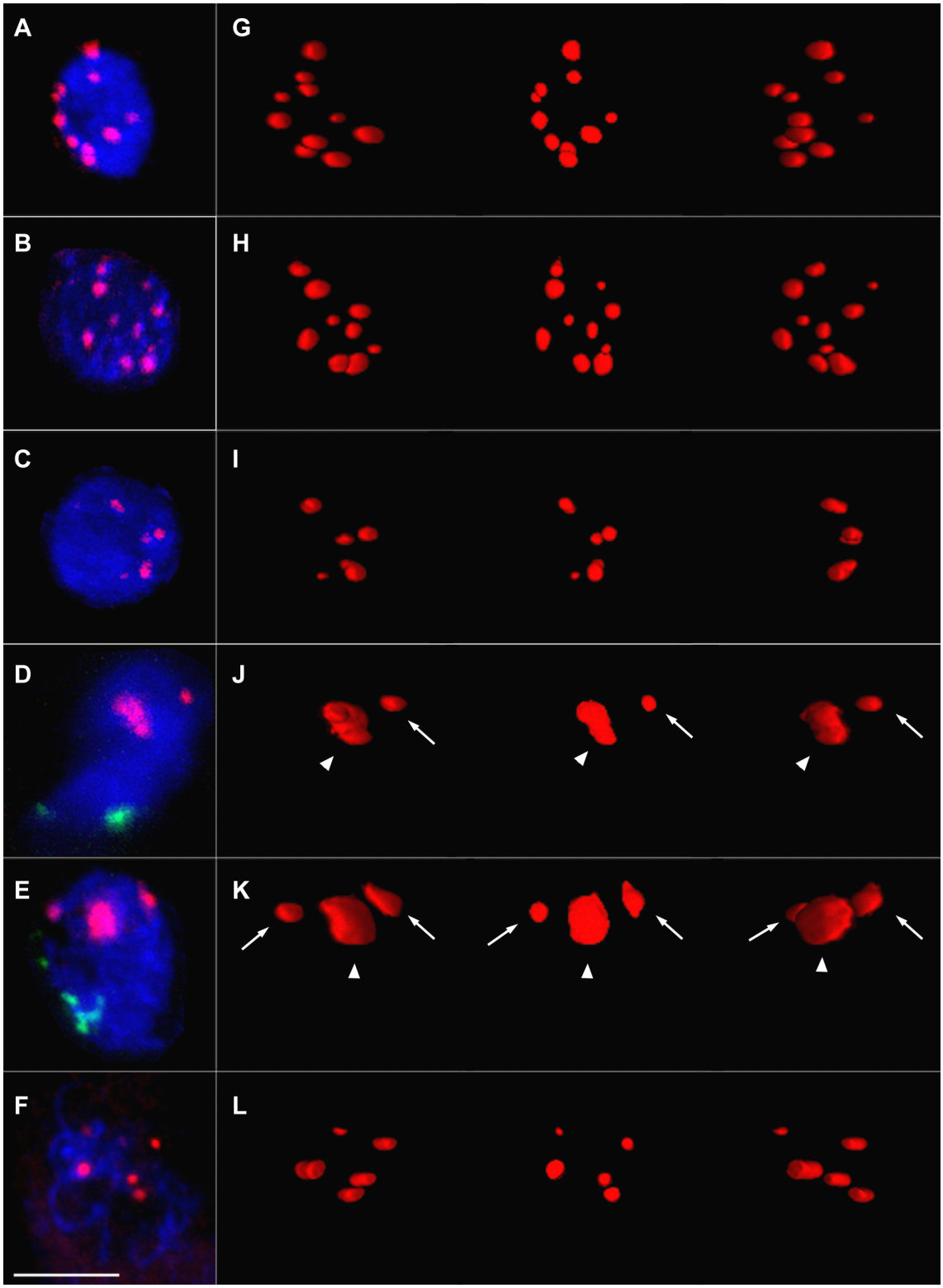
CCR sequences when the whole genome sequence of *B. distachyon* was assembled later. As shown at Fig. 10F, G, both *S. bicolor* CCR homologous sequences (Fig. 10G) and CentBd sequences (Fig. 10F) are specifically enriched at the centromeric regions of each chromosome with a high abundance. It was also shown by the magnified pictures (Fig. 10H) that, on each *B. distachyon* centromeres, most CentBd repeats were concentrated at one narrow locus (about 100 kb) and surrounded by CCR homologous sequences that cover a much wider range (about 5Mb) on each centromeres. Serial magnified pictures of one CentBd concentrated locus are shown at Fig. 10J. It was shown by these pictures that CentBd repeats were further separated into a few blocks within the locus, and the entire blocks are composed of tandemly arranged CentBd repeats. The *B. distachyon* centromere structures are consistent with previously reported cereal centromere structures as described in section 1.2 this chapter. As suggested from the studies of other cereal centromere structures, the CentBd concentrated locus should approximately define the core centromere positions and the CCR covered regions define the peri-centromeric regions, although knowledge of the distributions of CenH3 and centromere related histone modifications is required in order to confirm the organization of peri- and core centromeres in *B. distachyon*. Finally, it was also found by the analysis that the CCS1 homologous sequences are much less abundant than those of CentBd and CCR homologous sequences in the genome of *B. distachyon*, and their centromere specific enrichment was not obvious on chromosomes 1, 2 and 5 (Fig. 10I). This result brings into question the status of CCS1 as a centromeric sequence in *B. distachyon*, which is in contrast with the fact that the CCS1 sequence was first identified in a species closely related to *B. distachyon* (*B. sylvaticum*). Nevertheless, it may explain the lack of success of labelling of *B. distachyon* centromeres by CCS1 sequences in the preliminary experiments.

2.4 Centromere and telomere behaviour during early meiosis

Telomeres (in green) and centromeres (in red) were labelled by FISH on thin tissue sections (10-30μm) to examine chromosome behaviour during the process of chromosome pairing in *B. distachyon*. *B. distachyon* is a diploid with

Figure 11 B. distachyon centromere behaviour during early meiosis.

A-F: Confocal maximum projection images of B. distachyon PMC nuclei (blue) at different early meiosis stages, labelled with centromeres (A-F; red) and telomeres (D, E; green); Bar = 5 μm. G-L: 3D reconstruction images of centromeres in the corresponding nuclei from A-F; In each image of G-L, -30°, 0° and +30° y axial rotation images of the same reconstruction are shown from left to right in sequence. A, G: A typical anther somatic nucleus with 10 normal size and dispersed centromere dots. Centromeres were not associated in anther somatic nuclei at any stages. B, H: A typical pre-meiotic PMC nucleus with 10 normal size and dispersed centromere dots. Centromeres were not associated in most pre-meiotic PMC nuclei. C, I: A late pre-meiotic PMC nuclear with 5 normal size centromere dots, which shows the pairing of centromeres in some late pre-meiotic PMC nuclei that are about to go through meiosis. D, J: A typical leptotene PMC nuclear with one centromere block (J, arrow head) and one normal size centromere dot (J, arrow), which shows that centromeres are clustered at this stage. E, K: A typical zygotene PMC nucleus with one centromere blot (K, arrow head) and two normal size centromere dot (K, arrow), which shows that the clustering of centromeres was maintained during this stage. F, L: A typical pachytene PMC nucleus with 5 normal size centromere dots. Centromere clusters were resolved as paired centromeres at this stage.


A
G
B
H
C
I
D
J
E
K
F
L

a chromosome number of 10. Considering the tiny centromeres that chromosome 5 has (Fig. 10E), unassociated centromeres would be labelled as 8 bright dots plus 2 faint dots, paired centromeres as 4 bright dots plus 1 faint dot, and clustered centromeres (with more than two centromeres associated as one cluster) as less than four bigger dots. The faint labelling of chromosome 5 centromeres may or may not be detectable depending on the actual FISH sensitivity. As for the telomeres, unpaired telomeres should have a number close to 20 per cell and paired telomeres close to 10 per cell. The results are displayed in Fig. 11 and summarised in Fig. 12.

Centromeres were not associated and were randomly dispersed across the whole nuclear in most of the pre-meiotic PMCs (Fig. 11B, H) as well as in the surrounding somatic cells (appear as 8.61±0.19 dots) of all stages (Fig. 11A, G). Pre-meiosis is a vaguely defined stage that contains PMCs of different status with some of them more close to the onset of meiosis than others. Due to the lack of defined pre-meiosis markers, it is hard to sub-stage pre-meiosis precisely. However, it is possible to compare the ages of pre-meiotic PMCs of different florets by comparing the stages of their adjacent older florets within the same spikelets (as described at section 2.2, this chapter). As shown in Fig. 12A, the centromere numbers are slightly decreased in the pre-meiotic PMCs (7.91±0.55) compared with that of the surrounding somatic cells (8.61±0.19). The decrease is caused by the reduced centromere number (close to 5, Fig. 11C, H) in some of the late pre-meiotic PMCs. This kind of centromere pairing in PMCs was rarely seen in PMCs at early pre-meiosis. However, even in the late pre-meiotic florets, only a few PMCs were found to have centromeres paired. These results suggested that centromeres remain unassociated during the long period of pre-meiosis, and only started to be paired near the transition point of pre-meiosis to meiosis.

The PMC centromeres showed dramatic changes around the onset of meiosis, when PMCs were changing from pre-meiosis to leptotene. During this narrow time period (less than 2 hr), the average number of centromere dots reduced (may quickly pair first as mentioned above) from 7.91±0.55 to 1.84±0.24, which is coupled with the emergence of an unusual centromere organisation that

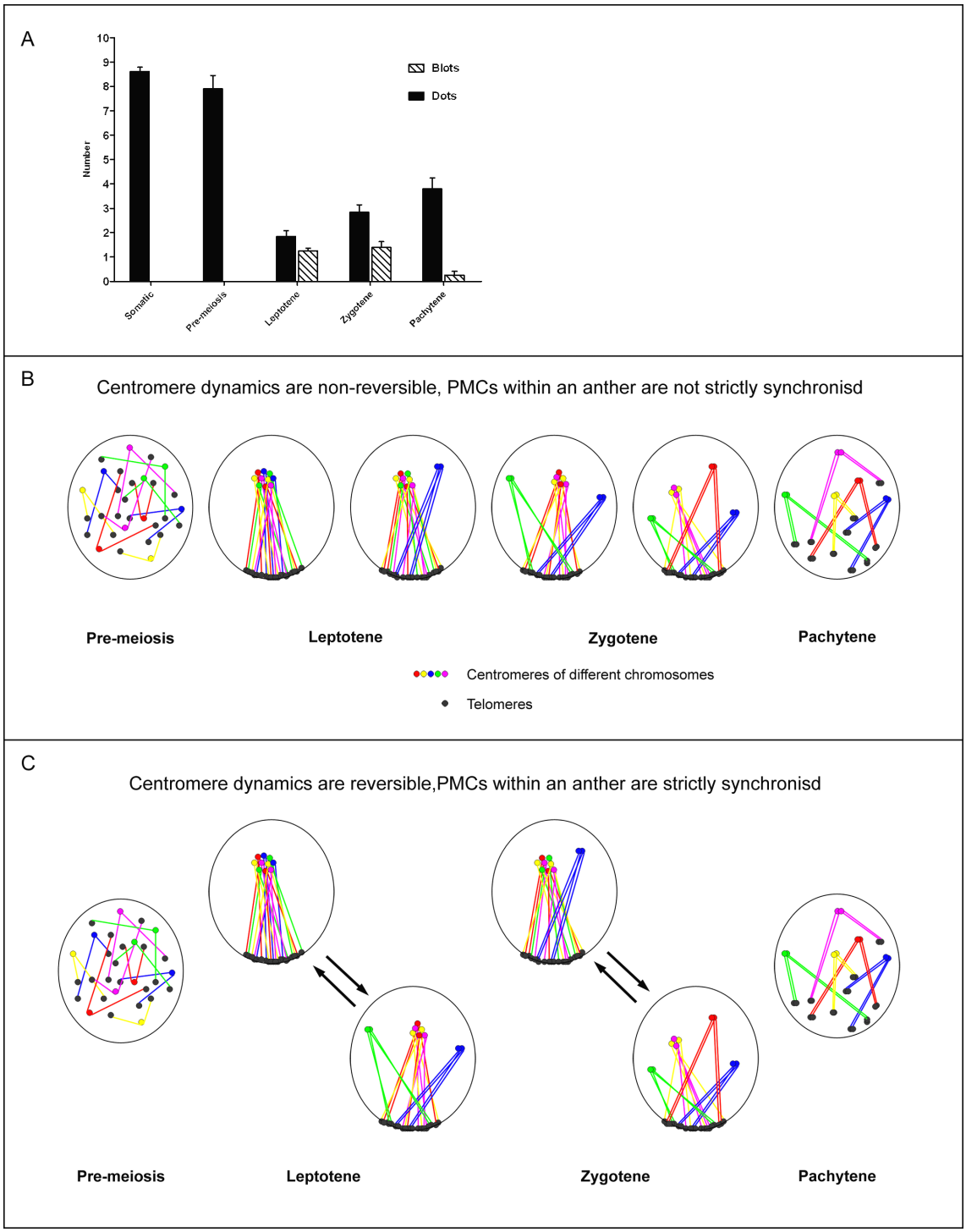
appeared in the form of large blocks (Fig. 11J, K; arrowheads). The small total number (3.08±0.20) of centromere dots and blocks indicated that a centromere clustering event occurred at leptotene. The average number of centromere dots per PMC increased from 1.84±0.24 to 2.83±0.31 and then to 3.80±0.44 when PMCs went through the stages of leptotene to zygotene and then to pachytene. Meanwhile, the average number of centromere blocks was nearly maintained (1.24±0.19 and 1.39±0.24) at zygotene but almost disappeared (0.25±0.16) later at pachytene. The average centromere dot number of 3.80±0.44 in a pachytene PMC was about half of the average centromere dot number in anther somatic cells (8.61±0.19), which indicated that centromeres were fully paired at pachytene. By dual labelling telomeres and centromeres together, it was shown that the centromere clustering event and emergence and maintenance of centromere blocks were accompanied by the formation and maintenance of telomere bouquet structures (Fig. 11D, E).

The numbers of centromere blocks and dots varied among PMCs of the same anther during early meiosis, with a few extreme examples that contained only one large centromere block and no centromere dot. However such extreme examples were only seen in leptotene anthers but not in zygotene anthers. Presumably the behaviour of centromeres in *B. distachyon* PMCs are controlled by the same mechanism and follow the same process; the variation of PMC centromere patterns (different number of blocks and dots) at early meiosis could then be explained by two possible models (Fig. 12B, C).

In the case that centromeres are undergoing a one-way non-reversible process, with which centromeres cluster into one block and homologous centromeres resolve as pairs one after another afterwards, variable centromere patterns would then represent PMCs at different time points of this process. PMCs within the same anther would not be strictly synchronised in this case (Fig. 12B). Or in the case that centromeres are undergoing a dynamic and reversible process, in which centromeres commute rapidly and frequently between clustered blocks and paired dots and homologous centromere pairs are established gradually during the process, then variable centromere patterns would represent PMCs that are fixed at different phases of the same dynamic process. In this case,

Figure 12 Centromere number at different early meiosis stages.

A: Bar represent the SEM. Centromere number decreased slightly in the pre-meiosis PMC cells (7.91±0.55) compared with that of anther somatic cells(8.61±0.19); A centromere clustering event which brought down centromere dot number to 1.84±0.24 occurred at leptotene stage with the emergence of centromere blocks (indicated by arrow heads in **J, K**); Centromere block number was maintained during zygotene, but centromere dot number was increased to 2.83±0.31; Centromere blocks almost disappeared (0.25±0.16) at pachytene, and centromere dot number was further increased to 3.80±0.44, which is just about half of the centromere number in the anther somatic cells. **B**: Model of centromere behaviour during early meiosis if it is a one-way non-reversible process, in which centromeres cluster into one block and homologous centromeres resolve as pairs one after another afterwards. In this case, PMCs within the same anther would not be strictly synchronised. Variable centromere patterns represent PMCs at different time points of the process. **C**: Model of centromere behaviour during early meiosis if it is a dynamic reversible process, in which centromeres commute frequently between clustered blocks and paired dots; homologous centromere pairs are established gradually during the process. In this case, PMCs within the same anther would be synchronised and undergoing the same dynamic process. Variable centromere patterns would represent PMCs fixed at different phases of the same dynamic process.


A
Number
Blots
Dots
Somatic
Pre-meiosis
Leptotene
Zygotene
Pachytene
B
Centromere dynamics are non-reversible, PMCs within an anther are not strictly synchronisd
Pre-meiosis
Leptotene
Zygotene
Pachytene
Centromeres of different chromosomes
Telomeres
C
Centromere dynamics are reversible,PMCs within an anther are strictly synchronisd
Pre-meiosis
Leptotene
Zygotene
Pachytene

PMCs within the same anther could be treated as synchronised in a dynamic process (Fig. 12C).

To clarify which of the two models represent the truth, it would be necessary to record the complete centromere dynamic of each PMC within one anther during early meiosis process especially near the onset of meiosis and also to label centromeres of specific chromosomes, which may be possible by live imaging of fluorescent protein fused CenH3 and BAC FISH of site specific BAC clones near centromeres.

To my knowledge, the study reported here is the first time that a meiotic centromere clustering event was reported in diploid plants. However centromeres with similar clustering pattern at meiosis has also been found in Arabidopsis PMCs (Zheng, personal communication), which suggests that it may represent a universal meiotic event for all diploid plants.

It was reported before that in hexaploid and tetraploid wheat, centromeres clustered into seven groups during early meiosis when the telomere bouquet was formed, with each group containing six homoeologous chromosomes of a particular chromosome group (60). It was suggested that paired homoeologous centromeres were resolved as homologous centromere pairs with this centromere clustering event. Both hexaploid and tetraploid wheat are constituted by ancestral genomes of same chromosome base numbers, where centromeres can be easily grouped together based on their chromosome number. However for many other allopolyploid plants like *B. rapa* (2n=20+18=38), *A. suecica* (2n=10+16=26) and *Brachypodium pinnatum* (2n=10+18=28) which are constituted by ancestral genomes of different base number, grouping of centromeres based on chromosome number may cause problems. Here the reported centromere clustering event of *B. distachyon* provides another possible model to help explain chromosome pairing in polyploids. In detail, the rapidly evolving sequences of centromere satellite repeats may serve as genome identity tags to sort centromeres based on their genome origin at the beginning of meiosis. Telomeres do still cluster to form the bouquets and the searching and pairing of homologous chromosomes are still initiated from telomeres. However the sorting and grouping of centromeres of the

same genome set may help promote the pairing of true homologous chromosomes or prevent homoeologous pairing.

In meiotic centromere clustering events of both wheat and *B. distachyon* in this study, the formation of telomere bouquets was found to be concurrent with the clustering of centromeres. It is therefore of interest to know if the clustering events of these two important chromosome components are related in some way. The formation of telomere bouquets at early meiosis is a universal meiotic event that is found in almost all the species studied. It was reported recently in rice that the formation of telomere bouquet was dependent on several synapsis-related axis proteins including OsAM1, OsREC8 and PAIR3. One possible way to answer the above question would be to investigate if the clustering of centromeres is affected by these synapsis-related genes as well. The formation of the telomere bouquet is widely considered to promote homologous pairing and recombination by bringing all chromosome ends together to facilitate the search of homologous chromosomes. Here the reported meiotic centromere clustering event also has the similar function of gathering chromosomes and thus the potential of promoting homologous pairing.

The unusually large size of the centromere block seen at leptotene and zygotene possibly represents the decondensation of the centromere satellite chromatin that forms the block (Fig. 10J, K). It was known that a sub-telomere heterochromatin decondensation event occurred concurrently with the formation of telomere bouquet in wheat and rye, and was important for the pairing of homologous or homoeologous chromosomes. It would be intriguing to determine if such a sub-telomere chromatin decondensation event does also occur in *B. distachyon* and if (in the case that it does exist) it is related with the possible centromere decondensation event. Chromatin conformation change is known to be related with *in situ* epigenetic changes. It would be interesting to determine if any of the typical centromere satellite epigenetic markers like H3K9me2 or DNA methylation are specifically modified in the possibly decondensed centromere block.

2.5 Site specific BAC clone screening

A BAC-based physical map of *B. distachyon* ecotype BD21 was constructed and integrated with whole genome shotgun sequence (WGS) assemblies using BAC end sequences (165), which allows the targeted selection of BAC clones at particular chromosome regions of interest. To utilize the genome sequence data of *B. distachyon*, the BAC libraries used to establish the physical map were chosen in this study. BAC clones from the libraries were kindly provided by Dr. Febrer from Prof. Bevan's group at John Innes Centre. Several randomly picked BAC clones were tested for BAC FISH on chromosome spreads of *B. distachyon* in the preliminary experiments, and more than a half of them failed to produce clear site specific signals (data not shown). Therefore, a bio-informatics pre-screening (as shown at Fig. 13A) was applied at the BAC selection step in the hope of eliminating BAC clones enriched in repetitive sequence and improving the chance of getting site specific BAC clones. The repetitive sequence content of specific chromosome regions was assessed in the genome browser of www.modelcrop.org. A total of 31 BAC clones distributed across the genome were selected to be pre-screened by this tool. As shown at Table1, 19 out of 31 BAC clones were found to be enriched in repetitive sequences, 4 BAC clones contained a moderate amount of repetitive sequences and the remaining 8 BAC clones were found to be nearly repetitive sequence-free and were thus selected to be the candidates for the next step. All the 8 repetitive sequence-free BAC clones were located at near telomeric regions, and almost all the centromere-located BAC clones were found to be repetitive sequence-enriched, even when several adjacent BAC clones were selected for analysis (Table1, see chromosome 5).

Another improvement to speed up the screening of site specific BAC clones concerned the BAC DNA preparation method. The use of high concentration DNA (normally 1 ug/ul) as template is required in nick translation probe labelling to ensure a proper concentration of the probes produced. The traditional method to obtain high concentration BAC DNA is by maxi prep, which is a time and effort-consuming process and only a limited number of samples (depending on the centrifuge used, normally no more than six) can be processed for the

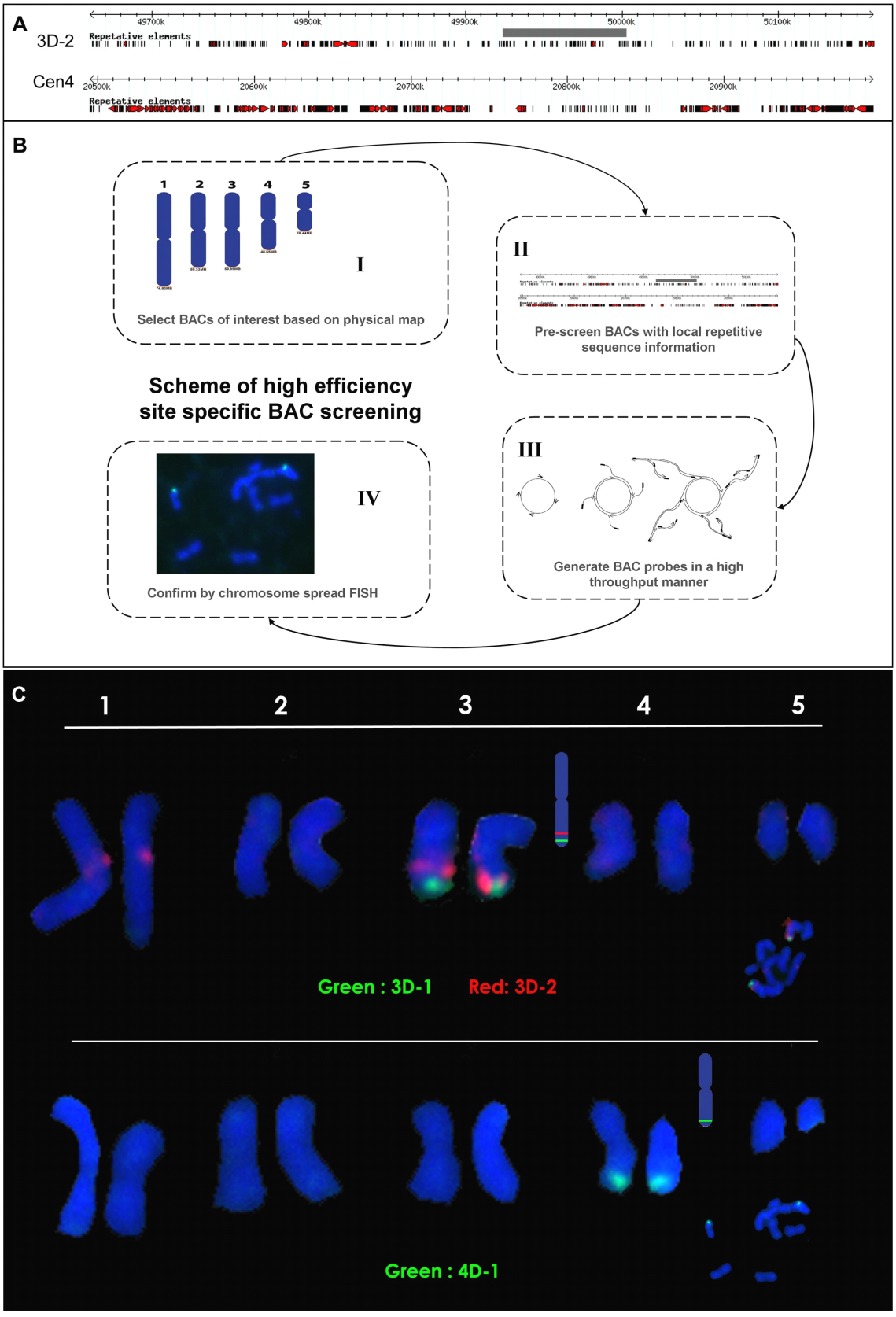
Table 5 Screening of site specific BAC clones in *B. distachyon*.

A total of 31 BAC clones distributed across the genome were selected to be pre-screened by bioinformatics tool. 8 out of 31 BAC clones were found to be repetitive sequence rare in the pre-screening, and 5 pre-screened BAC clones were amplified and labelled on root tip squash slides by FISH. All of the 5 BAC clones generated strong and clear site specific labelling in the interphase nuclei. Nice chromosome spreads were found on 2 root tip squash slides, with which the labelling of BAC clone 3D-1, 3D-2 and 4D-1 were confirmed to be located at expected chromosome positions.

BAC Code	BAC Name	Chr No.	Genome location	Pre-screen results	Confirmed by FISH
	Centromere 1	1	37.38-38.18MB		
b0023C23	1D-6	1	42.34-42.49MB	BAD	
b0047G12	1D-5	1	47.42-MB	BAD	
a0043J18	1D-4	1	52.88-53.06MB	BAD	
b0047B23	1D-3	1	59.86-60.03MB	BAD	
b0047E15	1D-2	1	67.19-67.30MB	GOOD	x
b0042M23	1D-1	1	74.45-74.62MB	GOOD	
	Centromere 2	2	28.99-29.15MB		
b0023N03	2D-5	2	34.06-34.18MB	BAD	
b0008P16	2D-3	2	43.91-44.04MB	BAD	
b0047A10	2D-2	2	48.96-49.11MB	MODERATE	
b0008E06	2D-1	2	54.02-54.15MB	GOOD	x
a0043K08	3U-1	3	5.95-6.15MB	GOOD	
b0047G01	3U-2	3	10.94-11.08MB	BAD	
b0008D20	3U-3	3	15.83-15.97MB	BAD	
a0024J19	3U-4	3	21.05-21.22MB	BAD	
	Centromere 3	3	25.21-25.66MB		
b0008D14	3D-6	3	30.35-30.51MB	BAD	
b0023F15	3D-5	3	35.20-35.36MB	MODERATE	
b0047N08	3D-4	3	40.01-40.15MB	MODERATE	
a0043H04	3D-3	3	45.28-45.11MB	BAD	
a0005I18	3D-2	3	49.92-50.00MB	GOOD	x
b0008N19	3D-1	3	54.74-54.89MB	GOOD	x
	Centromere 4	4	20.64-21.02MB		
a0005L11	4D-4	4	28.33-28.47MB	MODERATE	
a0024E06	4D-3	4	33.24-33.38MB	BAD	
a0043G19	4D-2	4	-38.73MB	BAD	
b0023G20	4D-1	4	43.35-43.50MB	GOOD	x
	Centromere 5	5	7.29-7.70MB		
b0008K23	5D-3a	5	12.50-12.65MB	BAD	
b0008B02	5D-3b	5	12.92-13.05MB	BAD	
a0024H10	5D-2a	5	17.47-17.65MB	BAD	
a0024D15	5D-2b	5	18.05-18.22MB	BAD	
b0008M05	5D-1a	5	22.75-22.90MB	BAD	
a0005D23	5D-1b	5	22.78-22.92MB	BAD	
b0008M07	5D-1c	5	23.75-MB	GOOD	

Figure 13 Screening of *B. distachyon* site-specific BAC clones.

A: Sample *B. distachyon* chromosomal regions in genome browser with local repetitive sequence information. The upper figure shows a common chromosomal region with average abundance of repetitive sequences and the lower figure shows a typical repetitive sequence enriched region. Grey bar in the upper figure shows the range of the BAC xxx (see its labelling at **C**). Repetitive sequence-enriched sub region (around 49.8 Mb) was avoided during the selection. **B**: Diagram showing the scheme of the developed site specific BAC screening strategy. Detailed description of this scheme can be found in the corresponding text of section 2.5. **C**: The labelling of three screened BAC clones by FISH on *B. distachyon* metaphase chromosome spreads. Images of each chromosome pair within one chromosome spread were cropped out and arranged in order. Physical positions of the labelled BAC clones is shown next to the FISH images of the corresponding chromosomes, and the untouched images of the corresponding chromosome spread is shown in the bottom right corners. For each BAC clone, strong and clear site specific labelling was seen at the expected loci. BAC xxx and xxx only labelled specifically at the expected loci. Site specific faint labelling of BAC xxx was also found on chromosome 1, which may represent the existence of BAC xxx homologous sequences at chromosome 1.


A
3D-2
Repetative elements
49700k
49800k
49900k
50000k
50100k
Cen4
20500k
20600k
20700k
20800k
20900k
Repetative elements
B
1 2 3 4 5
I
Select BACs of interest based on physical map
II
Pre-screen BACs with local repetitive sequence information
Scheme of high efficiency site specific BAC screening
III
Generate BAC probes in a high throughput manner
IV
Confirm by chromosome spread FISH
C
1
2
3
4
5
Green : 3D-1
Red: 3D-2
Green : 4D-1

same run. Also the quality of the BAC DNA is quite variable depending on the protocol and reagents used and the status of bacterial culture. Such a BAC DNA extraction method is one of the major bottleneck steps in the screening of site specific BAC clones. As an alternative, a rolling-circle-amplification (RCA) method was applied to improve the efficiency of BAC DNA extraction in this study. In this method, BAC DNA was amplified directly by DNA Polymerase Phi29 with the RCA reaction, which took only a few minutes to set up. High concentration BAC DNA (more than 1 μg/ul) was obtained reliably after a typical 10 hour amplification. The amplified BAC DNA was ready to be used directly as the DNA template for the following nick translation step, which allows the whole probe labelling process to be done on multi-well plates in a high throughput manner.

5 pre-screened BAC clones were amplified and labelled on root tip squash slides by FISH. All of the 5 BAC clones generated strong and clear site-specific labelling in the interphase nuclei. Nice chromosome spreads were found on 2 root tip squash slides, with which the labelling of BAC clones 3D-1, 3D-2 and 4D-1 were confirmed to be located at the expected chromosome positions (Fig. 13C).

Taking together all the steps described above, a high efficiency strategy to screen site specific BAC clones in *B. distachyon* was established (Fig. 12B). BAC clones are first selected from places of interest based on the physical map, and then pre-screened by bio-informatics tools, amplified by RCA and labelled by nick translation. BAC clones will be finally confirmed to be site specific and correct by FISH on chromosome spreads. Since BAC clone probes prepared in this way will in most cases (5 out of 5 in this study) produce site specific labelling, the effort needed in the last time consuming FISH step will be minimized, which will in return greatly improve the efficiency of BAC clone screening. One possible improvement of this strategy would be with the bio-informatics tools used to pre-screen BAC clones. The bio-informatics tools used in this study only distinguish repetitive sequences from non-repetitive sequences but give no further information about the distribution of a given repetitive sequence within the genome. For repetitive sequence-enriched regions where a BAC clone can

not avoid containing some repetitive sequences (like peri-centromeric regions shown at Table1), the differentiation of different repetitive sequences will be needed to select candidates. A bio-informatics tool that can provide such information can certainly improve the applicability of this strategy, especially in repetitive sequence-enriched regions.

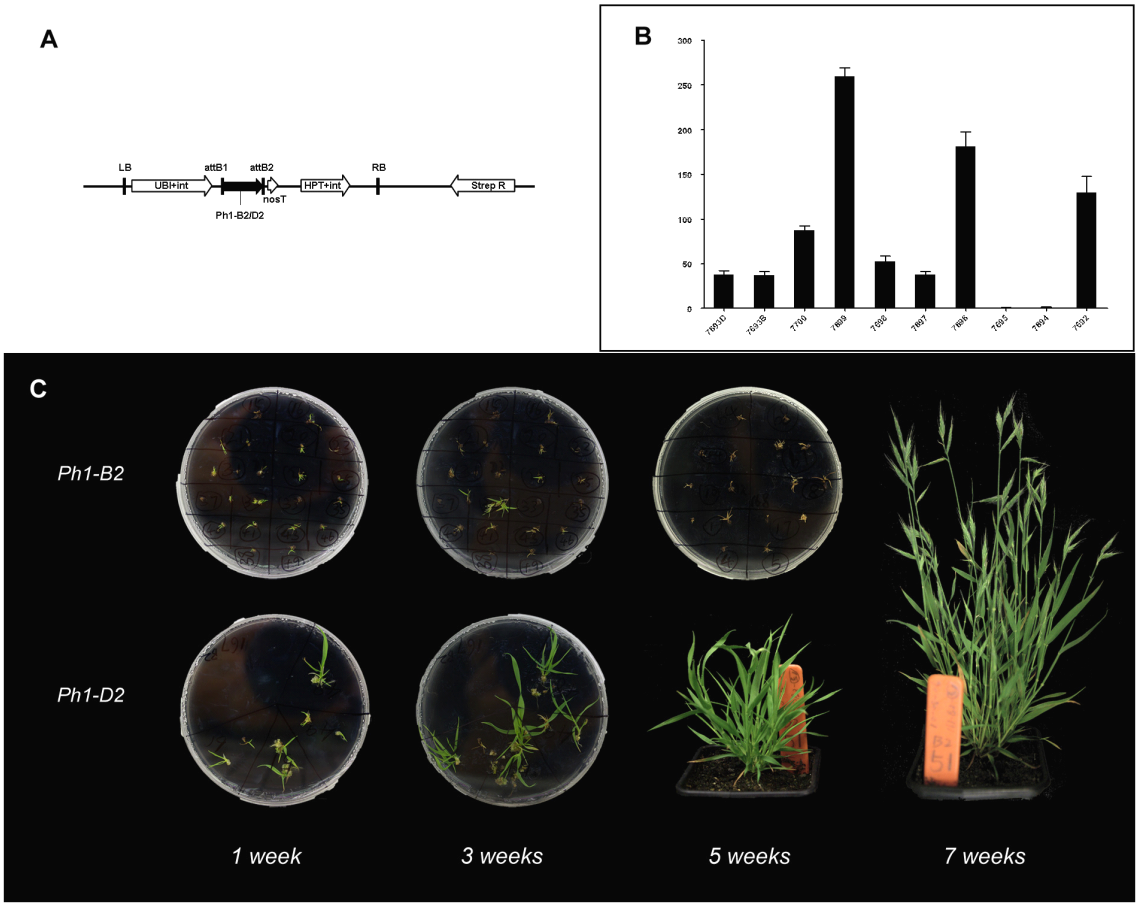
2.6 Over expression of wheat CDK genes is lethal for B. distachyon

Over expression constructs were made by inserting the coding sequences of Ph1-B2 or Ph1-D2 into the binary pVec8-Gateway vector by Gateway cloning (Fig. 14A). The expression of Ph1-B2 or Ph1-D2 was controlled by the maize ubiquitin promoter contained in the pVec8-Gateway vector. pVec8-Gateway vector was kindly provided by Dr. Chulmin Kim and was made by replacing the GFP coding sequence of pVec8-GFP vector (200) with the Gateway cassette. The Ph1-B2 and Ph1-D2 over-expression constructs were transformed into *B. distachyon* to study the effect of Ph1-B2 and Ph1-D2 genes in *B. distachyon*. Two independent transformations were done with the same constructs and protocol. The first transformation was done in Dr. Vain's lab and the other was done by myself several months later. The two transformations produce comparable results, and the one done by myself was described into detail below.

100 calli produced from young embryos were transformed separately with each of the two constructs with the same protocol. Similar amount of callus lines (50 for Ph1-B2 and 41 for Ph1-D2) were obtained after several rounds of antibiotic selection and no noticeable difference were found at this stage between calli transformed with different constructs. Calli transformed with both constructs generated shoots successfully on the regeneration germination media, although the Ph1-D2 transformed calli grew slower. However, after being transferred onto the germination media, the final step before transferring into soil, Ph1-D2 transformed calli stopped growing gradually. Lots of Ph1-D2 transformed calli died on the media after 3 weeks, and by the time of 5 weeks, all Ph1-D2 transformed calli died on the media (Fig. 14C, top row). In contrast, the Ph1-B2 transformed calli grew as normal on the germination media and later in the soil

Figure 14 Over-expression of wheat CDK genes in *B. distachyon* plants.

A: Diagram of over-expression constructs used for the transformation; The constructs were made by inserting the coding sequences of Ph1-B2 or Ph1-D2 into the binary pVec8-Gateway vector by Gateway cloning and the expression of Ph1-B2 or Ph1-D2 were controlled by the maize ubiquitin promoter with intron; **B**: The expression of Ph1-B2 in the young leaves of T1 pVec8-Gateway::Ph1-B2 *B. distachyon* plants. The figure shows the increase of Ph1-B2 expression in each transgenic lines compared to the wild type plants. Bar represents the SEM, and expression of each sample was corrected by the expression of GAPDH of the same sample. Significant expression of Ph1-B2 was found in the young leaves of 8 out of 10 transgenic lines. High expression (>100 fold increase) of Ph1-B2 was found in young leaves of lines 7692, 7695 and 7699. **C**: Growth of Ph1-B2/Ph1-D2 transgenic calli. Shoots of Ph1-D2 transformed calli were much smaller than those of the Ph1-B2 transformed calli after 1 week of growth on germination media, however most Ph1-D2 transformed calli were still alive by that time. Many Ph1-D2 transformed calli started to die after 3 weeks of growth on germination media, but in contrast, Ph1-B2 transformed calli grew fast and as normal. 5 weeks after being transferred onto the germination media, all the Ph1-D2 transformed calli died, whereas Ph1-B2 transformed calli successfully regenerated into mature plants. Two weeks later, seeds were generated by the Ph1-B2 transformed calli with a normal seed set.


A
LB
attB1
attB2
RB
UBI+int
HPT+int
Strep R
nosT
Ph1-B2/D2
B
300
250
200
150
100
50
0
C
Ph1-B2
Ph1-D2
1 week
3 weeks
5 weeks
7 weeks

(Fig. 14C, bottom row). A total of 11 Ph1-B2 transformed individual lines did regenerate to mature plants finally and produced healthy seeds with moderate seed set. The sequences of Ph1-B2 and Ph1-D2 are highly similar with each other. To check if the different phenotypes observed between Ph1-B2 and Ph1-D2 transformed plants are results of silenced Ph1-B2 expression, qRT-PCR were performed with the T1 generation Ph1-B2 transgenic plant sample to confirm the expression of Ph1-B2 coding sequence. As shown at Fig. 14B, the expression of Ph1-B2 in 8 out of 10 individual transgenic lines was more than 30 fold higher than in wild type plants. Expression increase of more than 100 fold was found in 3 of these transgenic lines. Despite the highly variable Ph1-B2 expression levels found in different Ph1-B2 transgenic lines, they all showed a wild type-like phenotype. Parameters including plant height, tiller number and 2C/4C nuclei ratio were measured in each transgenic line and compared with their Ph1-B2 expression level, and Ph1-B2 expression correlated phenotypes were not found. It is therefore concluded that the expression of Ph1-B2 is in general neutral for the *B. distachyon* plants and the expression of Ph1-D2 is in contrast lethal for the *B. distachyon* plants.

3. Materials and methods

3.1 Seedling growing

- Seed should be stored at 5°C.
- Choose bottom 4-6 large seeds from each spikelet; seeds above this are generally inferior. 40-50 seeds will provide 40 good seedlings.
- Soak seeds in sterile water for 30 minutes (optional).
- Remove top glume under microscope.

Note: there is an advantage of 7 days growth when removing one glume over not removing any

- Rinse in 70% ethanol for 30 sec and then three times with sterile water.
- Soak in 10% sodium hypochlorite solution (Fluka 71696) 100ml with 1 drop of Tween for 3 mins on shaker
- Rinse three times in sterile water.
- 10-20 seeds to sterile petri dish containing two layers damp sterile filter paper.
- Wrap in foil and put at 5°C for two days.
- Incubate in dark then light at 25°C until shoots and roots well formed (approx 5 days).
- Pot into 40 cell trays containing Medicago mix, then into 7cm pots or rootrainers
-

3.2 Adhesive slide preparation

- Wash either normal glass slides (for wax embedding sectioning) or 8 well multitest slides (for Vibratome sectioning, MP Biomedicals) in 5% Decon 90 (Decon) for 1-3 hrs with shaking.

Note: Put slides on slide rack and use a blue box for slide wash. Can wash overnight

- Rinse slides with sterile water 3 times and dry the slides.
- Put the slide in a freshly prepared 2% 3-aminopropyltriethoxysilane (APTES, Sigma) in acetone for 10 sec.

Note: Keep the APTES treatment time at precisely 10 sec, otherwise high background may be introduced to the slides.

- Wash slides with pure acetone briefly. Let foil paper-covered slides dry in a fume hood.

Note: Slides can be stored at this step. Only move to the next step if the slides are to be used.

- Put the slide in 2.5% glutaraldehyde in 1×PBS for 30 mins to activate the slides. Rinse the slide with sterile water afterwards and let slides dry in the fume hood.

Note: The glutaraldehyde solution can be store at 4°C for up to 1 month for reuse. Dispose the solution when the color turns to yellow.

3.3 Vibratome sectioning

- Soak samples in freshly prepared fixative (4% paraformaldehyde in 1×PBS with 0.01% Tween, pH7.0). Vacuum infiltrate samples until they start to sink, and then replace with fresh fixative and leave at 4°C overnight.
- Remove fixative and wash fixed samples two times with 1×PBS for 15 min.
- Remove the outer glume carefully under stereo microscope to expose the base floret
- Cut spikelet in the middle with the bottom 5 mm left, and then use super glue to attach the trimmed spikelet vertically on the section block with the base of the spikelet facing up.
- After the attachment super glue has dried, carefully coat spikelet with another thin layer of super glue and let it dry.
- Fit the section block onto the Vibratome machine, fill the Vibratome with sterile water and section sample into 30 μm thick thin slices and put slices onto APTES treated adhesive slides.
- Absorb water brought from the sections with sterile filter paper, and then quickly cover the sections with 10 μl 50% DMSO.
- Let the slices dry at 42 °C in the chamber overnight.
-

3.4 Microtome sectioning

- Soak samples in freshly prepared fixatives (4% paraformaldehyde in 1×PBS with 0.01% Tween, pH7.0). Vacuum infiltrate samples until they start to sink, and then replace with fresh fixative and leave at 4°C overnight
- Remove fixative and wash fixed samples with 1×PBS for 15 min. After wash, transfer samples into embedding cassettes and store cassettes in 70% ethanol.
- Put embedding cassettes into the embedding chamber and run embedding protocol as below:

	Regent	Duration	Temperature
Step 1	70% EtOH	4 hrs	35°C
Step 2	80% EtOH	4 hrs	35°C
Step 3	90% EtOH	4 hrs	35°C
Step 4	100% EtOH	3 × 4 hrs	35°C
Step 5	100% Xylene	3 × 4 hrs	35°C
Step 6	Wax	4 × 4 hrs	60°C

- Dissect and position samples in the cassettes for proper size and orientation, then cool down cassettes on cold plate to harden the wax

Note: Sample containing wax cassettes can be stored at 4 °C for up to one month or -20°C for up to one year

- Fit the cooled cassettes on Microtome and section samples into continuous 10µm think wax strips.

Note: Warm the cassettes at RT for at least 30 min first if they are taken from fridge.

- Float wax strips with sterile water on APTES treated adhesive slides and let them dry at 42°C in the chamber overnight.
- Wash slides with 100% xylene for 15 min× 2 in a glass tank to remove the wax
- Dry slides thoroughly in a fume hood. Before use, draw the borders of samples with a liquid blocker pen (Cosmo Bio, Japan).
-

3.5 Metaphase chromosome spread preparation

- Immerse the seedlings for 5 min in 10 ml of citrate buffer* in a small watch glass to remove the fixative. During this first wash, excise with a scalpel and retain the distal 3 mm length of the roots and discard the rest of the seedlings. Remove the citrate buffer with a pipette and repeat the wash twice.
- Drain the roots and put about 100 µl of enzyme mixture** for digestion of root meristems. Incubate at 37°C for one hour.
- Carefully replace the enzyme with citrate buffer, place the material on ice and wash for 15 min. Remove the root-tips from the rest of the root. The root tip is white and opaque, in contrast to the rest of the root, which is more translucent.
- Carefully transfer one root-tip to a clean slide with a cut end 10µl pipette. Remove the solution carefully first with the same pipette and then with filter paper, then put a drop of 8 µl 45% acetic acid and put a 18-18 mm cover slip on.
- Put the slides in the wet chamber and wait for about 15 mins, making sure that acetic acid does not dry out.
- Squash the root-tip through filter paper firmly but not too hard.
- Freeze the preparation thoroughly on dry ice or in -80°C, flick off the cover slip with a blade and flood the slide with ethanol: acetic acid fixative (pre-cooled to -20°C) with a Pasteur pipette.
- Drain the slide briefly by touching the long edge onto absorbent tissue, immerse it immediately in absolute ethanol for 15–30 min and allow it to air-dry overnight.
- Post fix the slides with 4% paraformaldehyde for 20 mins at RT, then wash with TBS buffer twice and dehydrate by methanol series (30%, 50%, 70%, 100%, 2 mins each) and allow it to air dry.

- Store the slides in a sealed box with desiccant at -20°C. Before use, draw the borders of samples with a liquid blocker pen (Cosmo Bio, Japan).

*Note: * Citrate buffer: 0.1 M Citric acid monohydrate and trisodium citrate dihydrate in sterile distilled water. Mix the two solutions at a ratio of 4:6 respectively and adjust pH to 4.8 with one or other solution. Dilute 1:10 with distilled water for use. Store both stock solutions and diluted buffer at 4°C.*

*Note: **Enzyme mixture: 2% (wt/vol) cellulase (Onozuka, R-10), 1% (wt/vol) pectolyase (Sigma, cat. no. P-3026), 1% (wt/vol) pectolyase (Duchefa Biochemie B.V., Y-23, cat. no. P8004), in 10 mM citrate buffer. Dissolve at 4 C and divide enzyme mixture into 0.1 ml aliquots and store at -20°C*

3.6 PCR amplification of telomere

The protocol was adapted from Ljdo *et al* (201).

- Make the PCR reaction mixture as below:

10 μM Primer A	0.5 μl
10 μM Primer B	0.5 μl
2× Qiagen master mix (Qiagen)	25 μl
Sterile distilled water	24 μl
Total	50 μl

- Run PCR with the following cycle setting:

94°C	3 min
Start 10 cycles	
94°C	1 min
55°C	30 sec
72°C	1 min
End of 10 cycles, and start of 30 cycles	
94°C	1 min
60°C	30 sec
72°C	1 min
End of 30 cycles	
72°C	5 min
10°C	Infinite

- PCR product can be used directly for nick translation labelling

3.7 RCA amplification of BAC DNA

The RCA amplification protocol was adapted from the protocol of Berr *et al* (202).

- Make mixture A as below:

BAC bacterial glycerol stock (1:5 diluted)	0.5 µl
2 × annealing buffer (80 mM Tris-HCl, 20 mM $MgCl_2$; pH 8.0)	2.5 µl
500 µM thiophosphate-modified random hexamer primers (5'-NNNN*N*N-3', *= PTO bond; Metabion, Martinsried, Germany)	1 µl
Sterile distilled water	1 µl
Total	5 µl

- Denature the mixture at 95°C for 3 min, cool rapidly on ice,
- Make mixture B with denatured mixture A as below:

mixture A	5 µl
10 × phi29 buffer	2 µl
4mM dNTP mixture	2 µl
phi29 DNA polymerase (NEB)	1 µl
Sterile distilled water	10 µl
Total	20 µl

- Incubate mixture B in a thermal cycle at 30°C for 10 hr.
- The amplified BAC DNA can be directly used for probe labeling

3.8 Nick translation labelling of probes

- Make the reaction mixture as below:

dNTP mixture (dA,G,CTP, 0.5 mM each)	2 µl
dUTP mixture (0.34 mM dTTP with 0.17 mM of Biotin-16-dUTP or Digoxigenin-11-dUTP)	2 µl
Template DNA	1-1.5 µg
Nick Translation Mix for in situ probes (Roche)	4 µl
Sterile distilled water	to 20 µl

- Incubate mixture at 15°C water bath for 90 min (telomere PCR products), 2 hr (centromere PCR products) or 3-6 hr (BAC RCA products)
- Stop the reaction by adding 1 µl 0.5 M EDTA (pH 8.0) and heating to 65°C for 10 min.
- Clean labelled probes with standard ethanol precipitation procedure
- Resuspend air-dried probe pellet with hybridization pre-mixture (hybridization mixture of 3.9 without probes), use immediately for FISH or store at -20°C

3.9 In situ hybridization of genome DNA

- Process section or squash slides sequentially through dehydration series of 30%, 50%, 70% and 100% methanol for 2 min each at RT, and then air dry the slides.
- Drop 20 µl of enzyme mixture* to wells of slide and put small cover slip on each well. Incubate at 37°C for 1 hr.
- Wash slides in IxTBS for 10 min at RT.
- Repeat dehydration series as step 1 and air dry the slides.
- Make hybridization mixture as below:

Deionised formamide	50 µl
100 mM PIPES, 10 mM EDTA, pH 8.0	10 µl
50% dextran sulfate	20 µl
5M NaCl	6 µl
Salmon sperm blocking DNA (1 mg/ml)	5 µl
Probes	9 µl each

- Denature hybridization mixture in boiling water or I00 C dry heat blocks for 8 min. Immediately cool mixture on ice for at least 5 min.
- In a cold room, add proper amount of cooled hybridization mixture (9 µl each well for multitest slides or 20-40 µl each slide for common slides) onto slides and put plastic coverslips for each slide (or each well for multitest slides).
- Place slides onto OMNISLIDE thermal cycler, run program as below to denature the genome DNA and let it to anneal with probe DNA gradually

Step 1	82°C	12 min
Step 2	55°C	3 min
Step 3	50°C	5 min
Step 4	45°C	5 min
Step 5	42°C	5 min
Step 6	40°C	5 min
Step 7	38°C	5 min
Step 8	37°C	infinity

Note: Allow OMNISLIDE machine to start and pause at 82 °C before putting slides in,

- After the temperature dropped to 37°C, move slides into humidity chamber and incubate at 37 °C O/N.

Note: It is very important to keep the humidity of the samples.

- Carefully remove coverslips in 2×SSC at 42 °C
- Wash slides in 20% formamide, 0.1×SSC for 2×5 min, then in 2×SSC for 2×5 min at 42 °C

- Wash slides in 2×SSC for 5 min, then in 4×SSC, 0.2% Tween 20 for 2 × 5 min at RT
- Put 100 µl blocking solution (5% BSA in 4×SSC, 0.2% Tween 20) onto each slide, and incubate slides for 5 min at RT

Note: No coverslips required

- Remove blocking solution by briefly touching the slide edges with filter paper, then put 100 µl detection solution (1:300 sheep anti-digoxigenin-fluorescein, Roche and 1:300 ExtrAvidin−Cy3, Sigma in blocking solution) onto each slide and cover each slide with one large plastic coverslip.
- Incubate at 37 °C for 45 min in wet chamber .

Note: From this step on, slides must be kept in dark. Use foil paper if necessary.

- Carefully remove coverslips in 4×SSC, 0.2% Tween 20 at RT
- Wash slides in 4×SSC, 0.2% Tween 20 for 3×5 min at RT
- Remove wash solution by briefly touching the slide edges with filter paper, then mount slides carefully with 60 µl Vectashield containing 1µg/ml DAPI. Try to avoid the production of bubbles during the procedure.
- Absorb surplus mounting solution with paper towel and fasten coverslips with nail polish at corners.

3.10 Transformation of B. distachyon

The protocol was established by Alves *et al* (203). The detail regent/media recipes can be found in this paper, and was not included here.

3.10.1 Production of CEC (compact embryogenic callus)

- Growing the plants to get seedlings as in step 1.
- Collect tillers from 7-9 weeks-old *B. distachyon* plants when the immature seeds are swollen but still green.
- Select immature seeds with a soft endosperm.
- Remove the top glumes of the seeds and put them in water. Drain well before sterilisation.
- Sterilise approximately 20 seeds for 30 sec with 20 ml of 70% ethanol in a sterile Petri dish.
- Drain ethanol and rinse three times with sterile water.
- Gently shake the seeds in 20 ml of 1.3% sodium hypochlorite solution for 4 min.
- Rinse three times with sterile water.
- Isolate immature embryos that are less than 0.3 mm in length.

Note: Only very small immature translucent embryos will produce homogeneous CEC at high frequency.

- Culture immature embryos, scutellum facing up, onto MSB3+Cu0.6 solid medium for 3 weeks at 25°C in the dark.

Note: CuSO4 significantly promotes the growth and embryogenicity of B. distachyon CEC.

- Excise the shoots under sterile conditions, as they elongate during the first 2-3 days of culture.
- At week 3, fragment CEC with a creamy colour and pearly surface in 1-3 pieces. Transfer pieces of CEC onto fresh MSB3+Cu0.6 solid medium for another 2 weeks at 25°C in the dark. Discard all non-CEC tissue.
- At week 5, split CEC with a creamy colour and pearly surface in 4-6 pieces. Transfer pieces of CEC onto fresh MSB3+Cu0.6 solid medium for another week at 25°C in the dark. Discard all non-CEC tissue.
- At week 6 (on the day of transformation), split CEC one last time in 4-6 pieces and place 50-100 CEC pieces on fresh MSB3+Cu0.6 solid medium before inoculation with Agrobacterium.

3.10.2 Preparation of Agrobacterium inoculum

- Suspend 5 µl of vector carried Agrobacterium tumefaciens (AGL1 strain) glycerol stock into 1 ml of LB liquid medium with suitable antibiotics.

Note: 50 mg/l streptomycin for pVec8-Gataway binary vector.

- Grow overnight at 28°C and 200 rpm.
- Plate 200 µl overnight culture onto solid MGL medium supplemented with with suitable antibiotics. Culture plates (upside down) for 2 days at 28°C in the dark.

Note: 50 mg/l streptomycin and 30 mg/l acetosyringone for pVec8-Gataway binary vector.

3.10.3 Agrobacterium-mediated transformation of CEC

- Scrape Agrobacterium layer with a L-shaped sterile glass Pasteur pipette and add to a 50-ml disposable sterile tube containing 10 ml of MSB+AS45 liquid medium. Strongly shake by hand to suspend Agrobacterium well.
- Shake suspension at 28°C and 220 rpm for 45 min to continue dispersing *Agrobacterium*.
- Dilute suspension with MSB+AS45 liquid medium to OD_{600}=1.
- Flood CEC plates with 13 ml Agrobacterium suspension and leave for 5 min at room temperature.
- Pipette out the bacterial suspension completely and transfer callus pieces straight onto a sterile filter paper covered empty Petri dish.
- Leave CECs uncovered under the laminar flow hood for 7 min as a desiccation treatment.
- Co-culture CECs from on MSB3+AS60 medium plates for 2 days at 25°C in the dark.

3.10.4 Selection of transformed calli

- Transfer co-cultured CECs onto MSB3+Cu0.6+H40+T225 solid medium. Culture for 3 weeks at 25°C in the dark.

- 3 weeks after transformation, dissect the healthy parts of each hygromycin resistant CEC and culture them as independent T-DNA lines onto MSB3+Cu0.6+H30+T225 solid medium for another 3 weeks at 25°C in the dark.

3.10.5 Regeneration of T-DNA plants

- 6 weeks after transformation, transfer hygromycin resistant calli onto the MSR26+H20+T225 regeneration medium for 2-3 weeks at 25°C under 16-h photoperiod. Keep all hygromycin resistant callus pieces from the same independent T-DNA line together for regeneration.
- 8-9 weeks after transformation, transfer shoots (rooted or not) onto the MSR63+Ch7+T112 germination medium. Culture for 2-3 weeks at 25°C under 16-h photoperiod.
- 10-11 weeks after transformation, transfer seedlings to CER with common growth condition.

CHAPTER V Studies of *Ph1* candidate genes in *A. thaliana*

1. Background

S. cerevisiae is the most convenient and versatile model organism in which to carry out biological studies. However in studies of Ph1-CDKs in *S. cerevisiae* as described earlier , no effect was observed with the expression of Ph1-B2 and Ph1-D2. *B.distachyon* is the closest model organism to *T. aestivum* and the expression of Ph1-D2 in *B.distachyon* did induce an interesting lethal effect. However, to study the function of a lethal gene, precise molecular biological tools like inducible promoters are needed. Such precise tools and other important research resources like mutant library are yet to be developed in *B.distachyon*.

On the other hand, *A. thaliana* is a well established model plant that can support well the most complicated studies. The direct use of buds rather than the time consuming callus in the transformation of *A. thaliana* greatly improves the transformation efficiency compared to other model plants. The transgenic study in *A. thaliana* was further helped by various vectors that were equipped with advanced cloning techniques (204). *A. thaliana* also has a well established T-insertion mutant library that covers most of the *A. thaliana* genes. Many important meiotic genes were uncovered in *A. thaliana* thanks to the well developed mutant library, which in turn provided *A. thaliana* the most complete meiotic molecular genetic knowledge in plants. In the study reported in this chapter, a set of expression constructs was generated and transformed to study the effects of *T. aestivum* Ph1-CDK genes in *A. thaliana*. It was found by the study that the expression of Ph1-D2 or Ph1-A3 genes greatly reduced the fertility or viability of *A. thaliana* plants, which is different from the neutral effect of Ph1-B2 gene in *A. thaliana* plants. The different effect of Ph1-D2/Ph1-A3 and Ph1-B2 was different from what we found in *B.distachyon*. It was confirmed by the study that *A. thaliana* is the suitable model organism to study the function of

Ph1-CDK genes. The vectors and transgenic lines made in this study will also provide a useful platform for future study.

Apart from the study of Ph1-CDKs in wild type *A. thaliana* plants, constructs were also generated and transformed into *cdkg1/cdkg1* to investigate if Ph1-CDKs can complement the function of CDKG1 in *A. thaliana*. CDKG is a newly discovered CDK kinase group that has large unique N-terminal domains (98). There are two copies of CDKG (CDKG1: At5g63370; CDKG2: At1g67580) in the genome of *A. thaliana*, which have similar expression patterns. It was found recently that CDKG1 is required for the formation of bivalents at meiotic metaphase I at high temperature (personal communication with Doonan J, unpublished data). Since PH kinases which include Ph1-CDKs are closely related with CDKG (see chapter II for more details), and the function of CDKG is also meiosis related like *Ph1* locus, we were interested to investigate if CDKG and Ph1-CDK are functional homologues with the collaboration of John Doonan's group. As part of the collaboration, Ph1-CDKs under the control of the CDKG1 promoter were transformed into *cdkg1/cdkg1* plants with or without the fusion of CDKG1 specific N-terminal domain, to test if Ph1-CDKs can complement the function of CDKG1 or at least the kinase domain of CDKG1. So far, the T2 seeds of transformation have been collected. The phenotyping of T2 plants will be performed under controlled temperature in the future.

2. Results

2.1 Preparation of constructs

A total of 14 Ph1-CDKs were found in the *Ph1* regions of A, B and D genomes of hexaploid wheat *T. aestivum*, and 5 of them contain intact open reading frames. Representative intact Ph1-CDKs CDS from each of A, B and D genome (Ph1-A3, Ph1-B2 and Ph1-D2) were transformed into *A. thaliana Col0* plants driven by the over expression promoter (35S), a meiotic promoter (DMC1) or an inducible promoter (XVE). For each 35S promoter-driven or DMC1 promoter-driven constructs, both N-terminal GFP tagged versions and tag free versions were generated and transformed. The CDS of Ph1-A3, Ph1-B2 and Ph1-D2 were also transformed into *CDKG1* insertion lines under the control of the CDKG1

promoter to test if Ph1-CDKs can complement the loss of CDKG1. Since Ph1-CDKs do not have the CDKG1 specific N-terminal domain, the CDS of Ph1-CDKs fused CDKG1 specific N-terminal domains were also transformed to test if Ph1-CDKs can at least complement the loss of CDKG kinase domain.

Full length or kinase domain of CDKG1 CDS was also transformed as positive controls. As a result, a total of 23 expression constructs were generated and transformed (Fig. 15, Table 6).

NAME	Type	Methods	fragments	Backbone Vectors
pWREV1	Entry Vector	BP	Ph1-A3 CDS	pDONR207
pWREV2	Entry Vector	BP	Ph1-B2 CDS	pDONR207
pWREV3	Entry Vector	BP	Ph1-D2 CDS	pDONR207
pWREV4	Entry Vector	BP	kinase domain	pDONR207
pWRDV1	Destination Vector	Ligation	DMC1 promoter	pMDC32
pWRDV2	Destination Vector	Ligation	DMC1 promoter	pMDC43
pWRDV3	Destination Vector	Ligation	CDKG1 promoter	pMDC32
pWRDV4	Destination Vector	Ligation	CDKG N-ter	pMDC32
pWRAT1	Constructs	LR	Ph1-A3	pMDC32
pWRAT2	Constructs	LR	Ph1-B2	pMDC32
pWRAT3	Constructs	LR	Ph1-D2	pMDC32
pWRAT4	Constructs	LR	Ph1-A3	pMDC43
pWRAT5	Constructs	LR	Ph1-B2	pMDC43
pWRAT6	Constructs	LR	Ph1-D2	pMDC43
pWRAT7	Constructs	LR	Ph1-A3	pWRDV1
pWRAT8	Constructs	LR	Ph1-B2	pWRDV1
pWRAT9	Constructs	LR	Ph1-D2	pWRDV1
pWRAT10	Expression Constructs	LR	Ph1-A3	pWRDV2
pWRAT11	Expression Constructs	LR	Ph1-B2	pWRDV2
pWRAT12	Expression Constructs	LR	Ph1-D2	pWRDV2
pWRAT13	Expression Constructs	LR	Ph1-A3	pMDC7
pWRAT14	Expression Constructs	LR	Ph1-B2	pMDC7
pWRAT15	Expression Constructs	LR	Ph1-D2	pMDC7
pWRAT16	Expression Constructs	LR	Ph1-A3	pWRDV3
pWRAT17	Expression Constructs	LR	Ph1-B2	pWRDV3
pWRAT18	Expression Constructs	LR	Ph1-D2	pWRDV3
pWRAT19	Expression Constructs	LR	CDKG1 kinase domain	pWRDV3
pWRAT20	Expression Constructs	LR	Ph1-A3	pWRDV4
pWRAT21	Expression Constructs	LR	Ph1-B2	pWRDV4
pWRAT22	Expression Constructs	LR	Ph1-D2	pWRDV4
pWRAT23	Expression Constructs	LR	CDKG1 kinase domain	pWRDV4

Table 6 List of vectors/constructs made in this study.

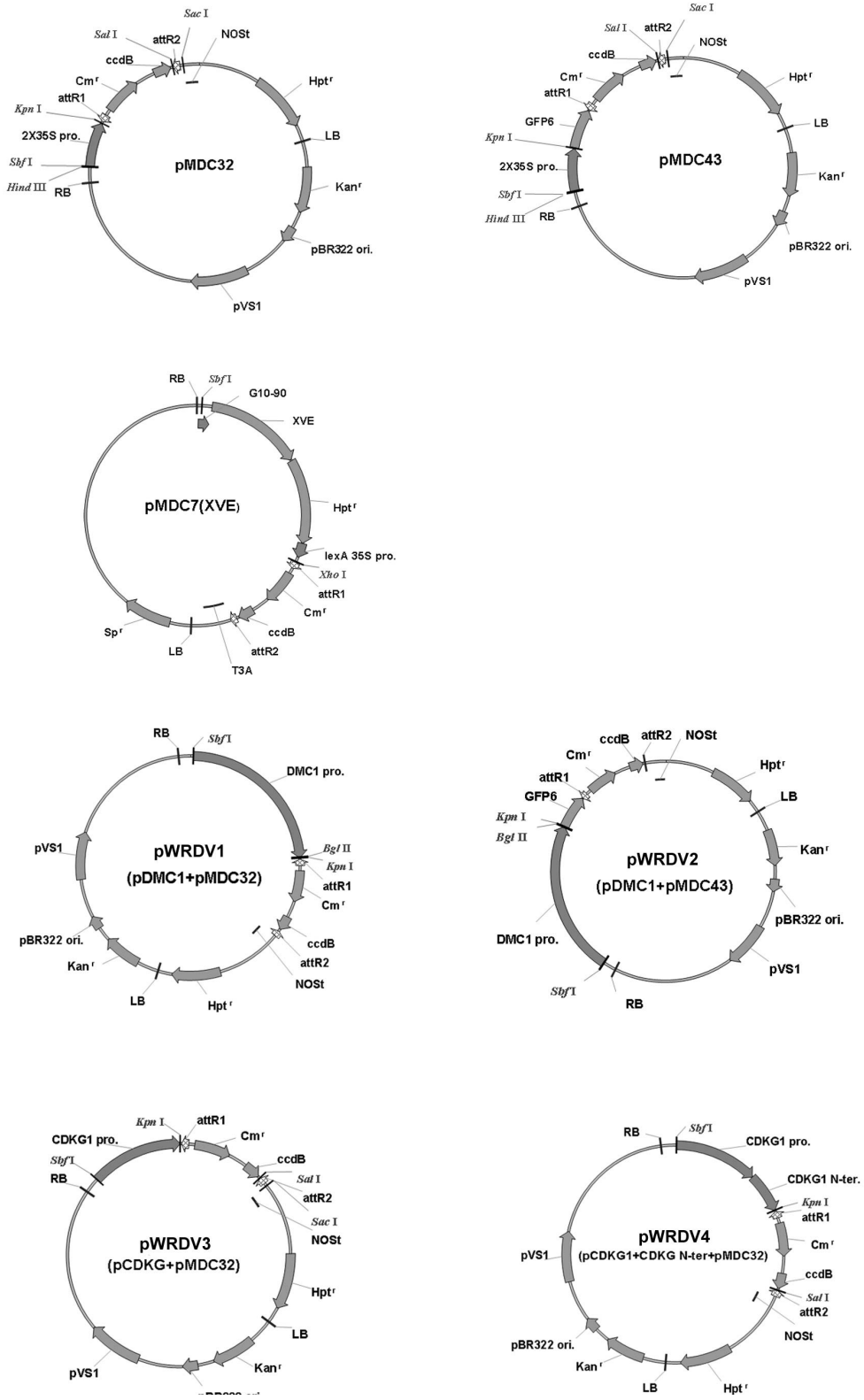


Figure 15 The maps of Gateway destination vectors used in this study.

2.2 Expression of Ph1-D2/A3 genes reduced the fertility or vitality of A. thaliana plants

The selection of resistant, positive T1 plants was accomplished in 19 out of 23 transgenic lines. The phenotypes of 35S promoter-driven and DMC1 promoter-driven transgenic lines are summarised in Table 7. For other transgenic lines whose phenotypes need to be confirmed under specific conditions, the phenotyping work will be done in the future with T2 transgenic plants.

Under normal growth condition, wild type phenotypes were seen in the T1 transgenic plants transformed with constructs pWRAT1, pWRAT2, pWRAT3, pWRAT5, pWRAT8, pWRAT10 or pWRAT11.

No resistant positive plant was collected from the T1 transgenic seedlings of construct pWRAT6 (35Spro:GFP:Ph1-D2), although a large number of T1 transgenic seeds were germinated at normal rate on the selection media. These results suggest that over expression of GFP fused Ph1-D2 protein is lethal for the *A. thaliana* plants. However the conclusion needed to be confirmed by the controlled expression of Ph1-D2 (pWRAT15, driven by XVE inducible promoter) to exclude the possibility of simply unsuccessful transformation.

A large number of resistant positive plants were produced from the T1 transgenic seedlings of construct pWRAT4 (35Spro:GFP:Ph1-A3). 18 of these resistant positive pWRAT4 transgenic plants were transferred into pots to generate T2 seeds. All the 18 plants grew as normal at the vegetative growth stage. However, 3 out of 18 pWRAT4 transgenic plants stopped at this stage without emergence of inflorescence shoot (Fig. 16A, *pWRAT4-7B*). The other 15 plants did flower; however, 7 out of 15 pWRAT4 transgenic plants were both sterile and dwarf (Fig. 16A, *pWRAT4-8C*), with another 3 being sterile only and another 4 being dwarf only (Fig. 16F).

Similar sterile phenotypes were also observed in the T1 transgenic plants of construct pWRAT12 (DMC1pro:GFP:Ph1-D2). Since the T1 seeds of pWRAT12 were unfortunately over sterilised, only a small amount of seeds germinated on

Figure 16 Phenotypes of Ph1-CDKs transformed *A. thaliana* plants.

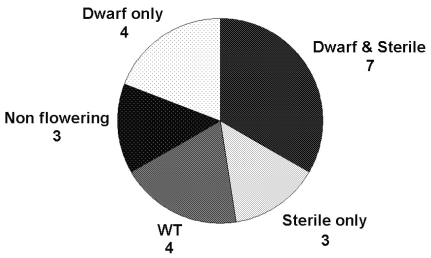
A. The representative T1 plants of construct pWRAT5 (35Spro:GFP:Ph1-B2), pWRAT4 (35Spro:GFP:Ph1-A3) and pWRAT12 (DMC1pro:GFP:Ph1-D2). *pWRAT5-5B* plant is wild type looking with normal size plant and healthy siliques. pWRAT4-8C plant is dwarf looking with short and less siliques. *pWRAT4-7B* plant stopped at vegetative growth stage without the emergence of inflorescence shoot. *pWRAT12-2D* plants are dwarf looking with short and less siliques. **B.** Comparisons of siliques from pWRAT5, pWRAT4 and pWRAT12 transgenic plants. Bar = 1cm. The siliques of pWRAT5 plants were about 1.5 cm long, and most of the siliques were stacked and plump. By contrast, siliques of pWRAT4 and pWRAT12 sterile plants were much shorter, and most of them were shrunken. **C-E.** Alexander staining of mature anthers of pWRAT5 **(C)**, pWRAT4 **(D)** and pWRAT12 **(E)** transgenic plants. pWRAT5 transgenic plants had normal size anthers and pollens were all clearly stained in pink colour. Arrow indicates one escaping pollen grain **(C)**. By contrast, pWRAT4 and pWRAT12 had shrunken anthers and no pollen looking particle could be found within the anthers **(D, E)** or in other places of the slides. **F.** Charts summarising the phenotypes of pWRAT4 and pWRAT12 T1 transgenic plants.

NAME	External fragments	Positive T1 ratio	Pheno
pWRAT1	35Spro:Ph1-A3	High	WT
pWRAT2	35Spro:Ph1-B2	Low	WT
pWRAT3	35Spro:Ph1-D2	High	WT
pWRAT4	35Spro:GFP:Ph1-A3	High	**Sterile**
pWRAT5	35Spro:GFP:Ph1-B2	High	WT
pWRAT6	35Spro:GFP:Ph1-D2	None	**Lethal**
pWRAT7	DMC1pro:Ph1-A3		
pWRAT8	DMC1pro:Ph1-B2	Moderate	WT
pWRAT9	DMC1pro:Ph1-D2		
pWRAT10	DMC1pro:GFP:Ph1-A3	High	WT
pWRAT11	DMC1pro:GFP:Ph1-B2	High	WT
pWRAT12	DMC1pro:GFP:Ph1-D2	?	**Sterile**

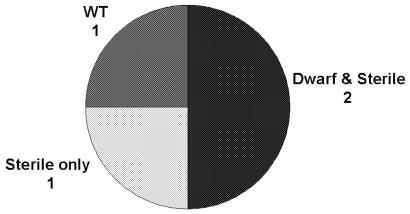
Table 7 Summary of transformation results.

WT means wild type phenotype.

F



pWRAT4



pWRAT12

the selection media and luckily 4 of them were resistant positive plants. All the 4 plants did flower, however 3 of them were sterile with 2 being both sterile and dwarf (Fig. 16A, *pWRAT12-2D*; Fig.1F).

The siliques of pWRAT4 (35Spro:GFP:Ph1-A3) and pWRAT12 (DMC1pro:GFP:Ph1-D2) sterile plants were compared with those of pWRAT5 (35Spro:GFP:Ph1-A3) plants in Fig. 16B (white bar representing 1 cm). Alexander staining of mature anthers of these plants was shown at Fig. 16C-E. The siliques of pWRAT5 plants were about 1.5 cm long, and most of the siliques were stacked and plump. By contrast, siliques of pWRAT4 and pWRAT12 sterile plants were only half the size, and most of them were shrunken. pWRAT5 transgenic plants had normal size anthers and pollen was all clearly stained as normal in a pink color. Many mature pollen grains were freely released from the anther (arrow at Fig. 16C showing one escaping pollen). By contrast, pWRAT4 and pWRAT12 had shrunken anthers and no pollen-like structures could be found within the anthers (Fig. 16D, E) or in other parts of the slides. These results confirmed that the over expression of Ph1-A3 or Ph1-D2 genes greatly reduced the fertility (pWRAT4) or even the vitality (pWRAT6) of *A. thaliana* plants. The meiotic expression of Ph1-D2 gene greatly reduced the fertility (pWRAT12) of *A. thaliana* plants.

The results reported in this chapter will be further discussed with other results in the final chapter.

3. Materials and methods

3.1 Preparation of constructs

The vectors and constructs prepared for the study reported in this chapter were summarised in Table 1. The maps of destination vectors used in this study were shown in Fig. 15.

The CDS of Ph1-A3, Ph1-D2, Ph1-B2 and the kinase domain of CDKG1 (At1g63670) were amplified with attB1 and attB2 tagged primers to introduce the attB1 and attB2 sites into fragments. Entry vectors pWREV1 - 4 were

constructed by BP reactions of pDONR207 vector and each of those attB1 and attB2 tagged fragments respectively. Gateway® BP Clonase® II enzyme mix (Invitrogen, 11789-020) was used with standard protocols on product manual.

The meiotic promoter of DMC1 was generated by dual digestion (SbfI and KpnI) of vector pPF408 (70). The promoter of CDKG1 and the promoter together with N-terminal domain of CDKG1 were amplified with SbfI and KpnI tagged primers from genome DNA respectively. Vectors pMDC32 and pMDC43 were dual digested by SbfI and KpnI and then ligased with the SbfI and KpnI tagged fragments to generate GATEWAY destination vectors pWRDV1 - 4.

The expression constructs were generated by LR reactions of various entry vectors and destination vectors as shown at Table 1. Gateway® LR Clonase® II enzyme mix kit (Invitrogen, 11791-100) was used with standard protocols on product manual.

The expression constructs were transformed with standard protocol as described at CHAPTER xx.

The sequences of DNA insertions were confirmed by sequencing the expression bacterial plasmids. The present of insertions were further confirmed by PCR before transformation in the transformation Agrobacterium strains.

3.2 Transformation of Agrobacterium

This protocol was adapted from ref. (205).

3.2.1 Preparation of Competent Cells

- Inoculate 200 ml of LB with 1 ml of an overnight culture of the chosen strain of Agrobacterium (strain AGL1 for the transformation of *B. distachyon*, and strain GV3101 for the tranformation of *A. thaliana*). Incubate overnight at 28ºC with vigorous agitation until log phase (OD_{600} 0.5-0.8).
- Pellet the cells at 5000 rpm for 10 min at room temperature.
- Wash the pellet with sterile 1X TE.
- Resuspend the cells in 20 ml LB, and aliquot to 100 µl per tube.
- Freeze in liquid nitrogen quickly and store at -70ºC.

3.2.2 Transformation and Recovery

- Thaw competent *Agrobacterium* on ice (use 250 µl per transformation reaction), and add DNA (10 µl of standard *E. coli* miniprep DNA).
- Chill the mixture on ice for 5 minutes, and then transfer to liquid nitrogen for 5 minutes.
- Incubate the mixture for an additional 5 min at 37ºC.
- Add 0.6 ml of LB to each tube, incubate at 28°C for 2-4 hr with gentle shaking.
- Transfer and spread all cells on LB agar plate containing the appropriate antibiotic (25 ug/ml rifampicin plus 100 ug/ml carbenicillin for AGL1 strain, and 25 ug/ml rifampicin plus 50 ug/ml gentamycin for GV3101 strain) and Incubate the cells for 2 days at 28ºC with shaking.
- Make glycerol stocks by adding same volumn of sterile 100% glycerol, and store them at -80ºC.

3.2.3 PCR Analysis of Agrobacterium

- Dilute 1 µl of *Agrobacterium*-saturated culture in 9 µl of distilled water in a small PCR tube,.
- Incubate diluted culture for 3 min at 100ºC in a thermal cycler then pause the thermal cycler at 85ºC for 10-20 min,
- Add 9 µl of PCR mater mix (Taq PCR Master Mix Kit, Qiagen, Cat. No. 201445) and 0.5 µl of each primer (10 mM) to the denatured culture, mix thoroughly and perform the PCR reactions as usual.

3.3 Transformation of A. thaliana

3.3.1 Plant Growth

- Place the plants in 4-inch pots at a density of 10-15 plants per pot. Grow the plants in short days (9-12 hours of light per day) using ample fertilizer to get large rosettes.

Note: If short-day conditions are not available, or if time is limited, grow the plants under long days from the beginning, but increase the density to 15-20 plants per pot.

- Remove the first inflorescence shoots as soon as they emerge (to encourage the growth of more inflorescences) and move the plants to long days (16-24 hours of light per day).

Note: Plants will be ready for transformation after about 1 week, when the secondary inflorescence shoots are ~3 inches tall.

3.3.2 Floral Dip of Arabidopsis

- 3 days prior to plant transformation, inoculate a 5 ml liquid culture of vector carried *Agrobacterium (GV3101)* and incubate at 28ºC with vigorous agitation in LB medium containing suitable antibiotics.

Note: 50 mg/ml rifampicin plus 25 mg/ml gentamycin for GV3101. Antibiotics for vector should be added as well.

- After 1-2 days, inoculate 500 ml of LB medium with 1 ml of the pre-culture and incubate again with vigorous agitation for an additional 17-20 hours at 28ºC, until OD_{600}=1.
- Pellet *Agrobacterium* by centrifuging at 6000 rpm in a GSA rotor (or equivalent) for 10 minutes at room temperature.

Note: The cell pellet is pink.

- Resuspend the cell pellet in 400 ml of infiltration medium (5% sucrose, 0.02% Silwet L-77).
- Transfer the *Agrobacterium* suspension to a convenient vessel for dipping plants,

Note: 400ml beaker for short stem flower, lid of pipette tips box for high stem flower.

- Removed the already formed fruit to increase the transformation efficiency. Seal the edge of the pots with tapes to prevent the falling out of the pots during dipping.
- Invert a pot of plants and dip the inflorescence shoots into the suspension. Rest the pot on the edge of the beaker and allow the plants to soak for 3 mins.

Note: The same suspension can be used for 10 or more pots. Try to avoid contamination of the soil with Agrobacterium, which produces a rather unpleasant smell.

- After dipping, put the pots in separate autoclave bags. Pour some sterile water into the bag, and seal the bags with tape to keep the moisture.
- After 24 hours, transfer the pots into glasshouse with normal growth condition. Remove the bags and the tapes on the pots.
- When the bottom seeds are fully mature, bag the plants and stop watering plants. Dry plants for about another 2 weeks.
- Collect and store the seeds.

Note: Store at 1.5°C, 7-10% humidity in the dark. Seeds can be stored for more than 10 years in these conditions.

3.3.3 Selection of transformed plants

Note: The whole procedure should be carried out in a fume hood.

- Place appropriate amount of seeds into opened seed bags. Place these seed bags inside a desiccators jar.
- Place a 250 ml beaker containing 100 ml bleach into the dessicator jar next to the seed bags.
- Add 3 ml concentrated HCl to the bleach carefully, then immediately sealing the jar.
- Allow sterilization by chlorine fumes to proceed for a period of between three and sixteen hours before use.

Note: The time needed varies based on the configuration of seed and the extent to which seed is contaminated. Three to four hours is often sufficient for reasonably clean seed. Overnight is usually acceptable although some seed killing may occur, especially if seed is not fully mature and dry. Try a small amount first is always recommended.

- Pour MS-S agar with appreciate antibiotics into 140mm petri-dishes for use in selection.
- Carefully scatter the sterilized seeds onto the plates, and seal the plates with Microspore tape.
- Store the seeds containing plates in a cold room for two days.
- Transfer the plates into a growth chamber and wait for the seedlings to emerge.

Note: Normally take 2 weeks.

- Seedlings that grow bigger and with roots went into the agar are more likely to be the positive ones.
- Transfer the positive seedlings into a new plate, let them grow for another 1 week and transfer them into pots.

3.3.4 Genotyping of transformed plants

Note: This protocol works well with fragments up to 1.0 kb.

- Take one tip from a side shoot and place it into PCR strips.
- Add 50μl of 0.25M NaOH for each well, seal with lid and incubate at 96 ºC for 10 min on PCR block. Spin the strips afterwards.
- Add 50μl of 0.25M HCl and 25μl 0.5M Tris-HCl (pH 8.0), and 0.25% NP40 (V/V). Spin the strips afterwards.
- Mix by pipetting up and down 5 times. Reseal with new lids and incubate at 96 ºC for 3 min on a PCR block. Spin the strips afterwards.

Note: Take care not to splash solutions from well to well.

- Use 1μl of the mixture above as the template for a 25μl PCR reaction.

Note: use a small cauline leaf and do the genotyping the same day if possible, otherwise the DNA might be still good for a month if stored at -20ºC

3.4 Alexander staining

The Alexander stain solution was made as below.

95% ethanol	10ml
malachite green	10mg
distilled water	50ml
glycerol	25ml
phenol	5g
chloral hydrate	5g
acid fuchsin	50mg
orange G	5mg
glacial acetic acid	4ml

Fresh anthers were taken from plants and directly put into Alexander stain solution for 30 min. The stained anthers were directly put onto slides, covered with coverslips and checked under light microscope.

CHAPTER VI Discussion

1. Suitable model systems to study Ph1

The effects of Ph1-CDK genes were tested in three different model organisms in this study. The ectopic expression of Ph1-CDK genes in *S. cerevisiae* did not induce any detectable changes in either the experiment of CDK complementation or Ime2 complementation. The results suggested that wheat Ph1-CDKs were not functional when expressed in *S. cerevisiae* under general conditions. *S. cerevisiae* therefore was unfortunately not suitable as a model organism for study of Ph1-CDK genes.

On the other hand, significant phenotypes were observed with the ectopic expression of Ph1-CDK genes in both *A. thaliana* and *B. distachyon*. Both *A. thaliana* and *B. distachyon* have their own pros and cons as model organisms for the study of Ph1-CDK. *A. thaliana,* as a well developed model system, is a time-efficient and economical choice. In addition, a set of transgenic *A. thaliana* lines has already been produced in this study. More transgenic lines can be generated easily with the Gateway compatible vectors made or collected in this study. The cons of choosing *A. thaliana* happen to be the pros of choosing *B. distachyon*. As mentioned at Chapter III, PH kinases are most enriched in grass genomes and are absent from most other genomes including *A. thaliana.* It is possible that PH kinases/Ph1-CDKs belong to a specific pathway whose full function is also limited to specific genomes. By contrast, *B. distachyon* has a typical PH kinases distribution, and the *B. distachyon* PHB sub-group kinases are even syntenic with wheat Ph1-CDKs. Since all the grass genomes investigated have at least two PH kinase clusters that belong to different sub-groups, it is very likely that the wheat genome also contains at least one other PH kinase cluster in addition to Ph1-CDKs. The *B. distachyon* PH kinases are likely to be the functional orthologues of corresponding wheat PH kinases. It may therefore be possible to understand the mechanism of *Ph1* locus by directly studying the functions of *B. distachyon* PH kinases.

2. Ph1-CDK genes are unlikely to be functional homologues of CDK

Ph1-CDKs were reported to be similar to the human Cdk2 gene at the time of mapping (5,6). A further bioinformatic 3D structure comparison of the human Cdk2 protein and wheat Ph1-B2 putative protein suggested that Ph1-CDK genes were likely functional homologue of human Cdk2 gene (206). The phosphorylation of Thr160 of human Cdk2 protein is needed for the kinase activity. This residue was also reported to be important for the activation of plant CDKA (207). Thr160 is not present at the corresponding residue of Ph1-B2 based on protein alignment. In the mentioned bioinformatic study, the author argued that Thr179 of Ph1-B2 was the corresponding human Cdk2 Thr160 residue, but was wrongly aligned with Thr165 of human Cdk2 by the software (206). However, it was shown by the multi-alignment of various PH kinases that this explanation was unlikely to be true, since the position of Thr160 was defined by several conserved residues surrounding it. Moving Thr179 to Thr160 would induce the mis-alignment of several highly conserved residues (Fig. 1-3, red frame 6; Table 3-1). Another obvious difference between Ph1-CDKs and human Cdk2 is at the PSTAIRE motif, which is important for the binding of cyclins. Ph1-CDKs do not have conserved PSTAIRE motifs at the corresponding region (Fig. 1-3, red frame 2; Table 3-1). In the mentioned bioinformatic study, it was suggested that the novel Ph1-B2 PSTAIRE motif, although not compatible with human cyclin-A, is biochemically comparable with most of the Arabidopsis cyclins (206). From the multi-alignment of PH kinases, it was shown that the PSTAIRE region of PH kinases were highly dynamic even among kinases of the same sub-group. On the other hand, the identified plant CDK kinases up to date all have highly conserved PSTAIRE motifs that are similar or identical to PSTAIRE (98). Therefore plant cyclins, although different from mammal cyclins, are more likely to have a preference for binding PSTAIRE-like motifs rather than the un-conserved PSTAIRE region of Ph1-CDKs. It is possible that PH kinase containing genomes have some additional unusual cyclins that are responsible for the binding of PH kinases. However, it was reported that *O. sativa* does not have additional cyclins to those of *A. thaliana* (208).

CDK kinases from different organisms are in general well conserved. The Cdc28 mutant complement assay in yeast is a routine method used to identify and/or confirm CDKs from other organisms. For instance, mammal Cdk1, Cdk2 (131) and plant CDKA (132) were confirmed to be functional CDK genes by successfully complementing the function of Cdc28 in the Cdc28 mutants of *S. cerevisiae*. However, in the study reported in Chapter V, the ectopic expression of either Ph1-B2 or Ph1-D2 failed to complement the function of Cdc28 in the Cdc28-4 ts mutant even at a less stressful temperature of 34°C (Fig. 8-1).

Putting together the results of the bioinformatic analysis of PH kinases and the yeast Cdc28 complement assay, it was concluded that Ph1-CDKs are not the functional homologues of a common CDK.

The CDKG protein, by structure, is well conserved with other typical CDK proteins apart from its specific N-terminal domain. Important CDK kinase motifs like PSTAIRE motif and activation Ser/Thr are all present in CDKG. Although CDKG kinases are closely related to PH kinases on the phylogenetic tree (Fig. 8-1C), the detailed comparison of CDKG kinases and PH kinases showed that, PH kinases, especially the PHB sub-group kinases are different from CDKG kinases at several important kinase motifs/residues. The results suggested that Ph1-CDK kinases, which belong to the PHB sub-group, are unlikely to be functional homologues of CDKG neither. However bioinformatics is only a tool for prediction. The conclusion should be drawn by the results of the CDKG complementation experiments in the future.

Even if Ph1-CDK genes are not functional homologues of CDKG genes, considering their, in general, closely related protein sequences and same field of function (meiotic chromosome pairing), they are still likely to have related functions. PH kinases are highly dynamic proteins that have mutations at several important kinase motifs/residues, which suggests that PH kinases are possibly only pseudo-kinases that function as suppressors by competing with normal kinases. It was suggested by the phylogenetic analysis of PH kinases and CDKG kinases that PH kinases may be derived from CDKG. It is therefore possible that PH kinases act as the suppressors of CDKG.

3. *Ph1-B2 is the likely candidate for the function of the Ph1 locus*

The restriction of homoeologous pairing by *Ph1* locus must be induced by specific genes on the B genome, since the deletion of *Ph1* region in either A or D genome will not produce the *Ph1* effect. “Specific genes” here could include genes that are exclusively transcribed from B genome, or genes that are only present in the B genome and transcribed. There are 7 copies of Ph1-CDKs presented in the *Ph1* region of the B genome, and only Ph1-B1 and Ph1-B2 have full length open reading frames (6). It was known from the same study that, only Ph1-B2, Ph1-B3, Ph1-B4, Ph1-B6 and Ph1-B7 were transcribed from the B genome in the presence of *Ph1* locus. Among them, Ph1-B2 has an intact open reading frame and was thus considered to be functionally similar with other intact Ph1-CDK proteins. However, as reported in Chapter IV and Chapter V, the function of Ph1-B2 is quite different from at least Ph1-D2/Ph1-A3 proteins. While the ectopic expression of Ph1-D2/A3 in both *B. distachyon* and *A. thaliana* induced clear effects like lethality or sterility, the ectopic expression of Ph1-B2 did not induce any noticeable effect in either *B. distachyon* or *A. thaliana*. As shown at Fig. 1-3, sequences of 5 intact Ph1-CDK proteins are mostly the same at conserved regions, with the exception that Ph1-B2 has a mutated C-terminal end of about 30 aa. This mutated C-terminal end of Ph1-B2 was therefore suspected to be the reason that Ph1-B2 had no effect. The other 4 transcribed B genome Ph1-CDKs would all encode truncated proteins if translated and are therefore unlikely to be functional by themselves. This kind of truncated Ph1-CDK are also present and transcribed from the *Ph1* region of the A genome in the absence of *Ph1* locus. Although it is still possible that one of the B genome truncated Ph1-CDKs has gained a specific function and is responsible for *Ph1* effect, the chance of getting unique effective truncation from many similar truncations is comparatively low. As a conclusion, Ph1-B2 was proposed to be the gene that is likely responsible for the restriction of homoeologous pairing in *Ph1* locus. The most direct and solid way to verify this hypothesis would be the complementary test. That is to test if the transformation of external Ph1-B2 gene can suppress the *Ph1* effect in the *Ph1* deletion line. However, this test will

request the use of wheat transformation, which is currently still a difficult and inefficient technique. As an alternative, it would be interesting to test the hypothesis in *B. distachyon*. As mentioned as Chapter II, *B. distachyon* PHB1 genes are close related with Ph1-CDKs and are likely the functional orthologues of Ph1-CDKs. Therefore if the hypothesis discussed here is true, then Ph1 effect can be introduced in *B. distachyon* by simply transforming a Ph1-B2 like *B. distachyon* PHB1 mutated gene into diploid or polyploidy *B. distachyon* plants.

Since all the transcribed B genome Ph1-CDKs were either truncated or mutated proteins, they are unlikely to induce the *Ph1* effect by direct positive function. On the other hand, the ectopic expression of Ph1-D2/Ph1-A3 has induced lethal or sterile effects in both *A. thaliana* and *B. distachyon*, which showed that they are functional genes. One logical presumption is that Ph1-B2 or any other possible Ph1 candidate gene from B genome (simplified as Ph1-B2 in this paragraph), acts by suppressing the function of normal Ph1-CDKs (Fig. 17, frame 2), and as a consequence to restrict homoeologous pairing, in the presence of *Ph1* locus (Fig. 17, frame 1). Ph1-B2 can suppress the function of normal Ph1-CDKs in two different ways. One possible way is to suppress simply the transcription of normal Ph1-CDKs. The idea is supported by the fact that the transcription of Ph1-D2 (a normal Ph1-CDK) was suppressed by the presence of *Ph1* locus (6). However, it was known from the analysis of PH kinases that, although PH kinases of the same sub-group tend to form gene clusters (like Ph1-CDKs), there are always one or more genes that are located away from the gene cluster (Bradi2g34210 and the cluster of Bradi4g33220/Bradi4g33230 for instance). It is therefore likely that there are other copies of normal Ph1-CDK genes located in non-*Ph1* regions and will unlikely be suppressed by the presence of *Ph1* locus. The existence of non-*Ph1* region Ph1-CDKs can be easily assessed in the near future, when the sequencing of wheat genome is completed. Another possible way is for Ph1-B2 to act as a dominant negative mutant that directly competes with the products of normal Ph1-CDKs. This can be assessed by co-expression of Ph1-B2 and normal Ph1-CDKs in *A. thaliana.* If the lethal or sterile effect of Ph1-D2/Ph1-A3 was suppressed by the expression of Ph1-B2, then Ph1-B2 is likely to act as dominant negative mutant.

4. The Ph1 regulation hypothesis

A hypothesis describing the mechanism of *Ph1* locus is given at Fig. 17 based on the discussion above. It is proposed in the hypothesis that intact normal Ph1-CDKs are functional genes that can promote homoeologous pairing or are needed for progression of homoeologous pairing. In wild type wheat, the activities of functional Ph1-CDKs are suppressed by Ph1-B2. As a result, homoeologous pairing is greatly reduced and the genome is stable. After the *Ph1* locus is deleted, the activities of functional Ph1-CDKs are released and homoeologous pairing levels are recovered. The genome of the *Ph1* deletion wheat line is unstable because of the recovered homoeologous pairing.

The hypothesis is consistent with the observed *Ph1* locus effects. For instance, *Ph1* locus is known to have a dosage effect. That is, homoeologous pairing can be further suppressed by adding of more *Ph1* copies (209). Following the hypothesis to consider Ph1-B2 as suppressor of functional Ph1-CDKs, it is then reasonable that more copies of Ph1-B2 will have a stronger suppression effect. It was also reported that genome rearrangement in *Ph1*b deletion line (AABBDD) is much stronger than in *Ph1*c deletion line (AABB) (210). In the hypothesis, functional Ph1-CDKs are the effective homoeologous pairing genes. The severe genome rearrangement in the *Ph1*b deletion line can be therefore interpreted as the effect of stronger alleles on the D genome or the combined effect of D and A genome functional Ph1-CDK copies. Last but not least, it was reported before that *Ph1* mutant could not be produced by EMS treatment, which indicated that the effective *Ph1* gene was not a functional gene (5). It was also proposed in the hypothesis that Ph1-B2 acts as mutated suppressor of functional genes rather than functional gene itself.

The hypothesis can be further extended to explain the initiation of Ph loci as shown in Fig. 17, 18. Wheat and Ph1-CDKs were used as an example in the discussion for convenience. Ancestral diploid wheat species were first divided by the rising of BDM hybrid incompatibility loci

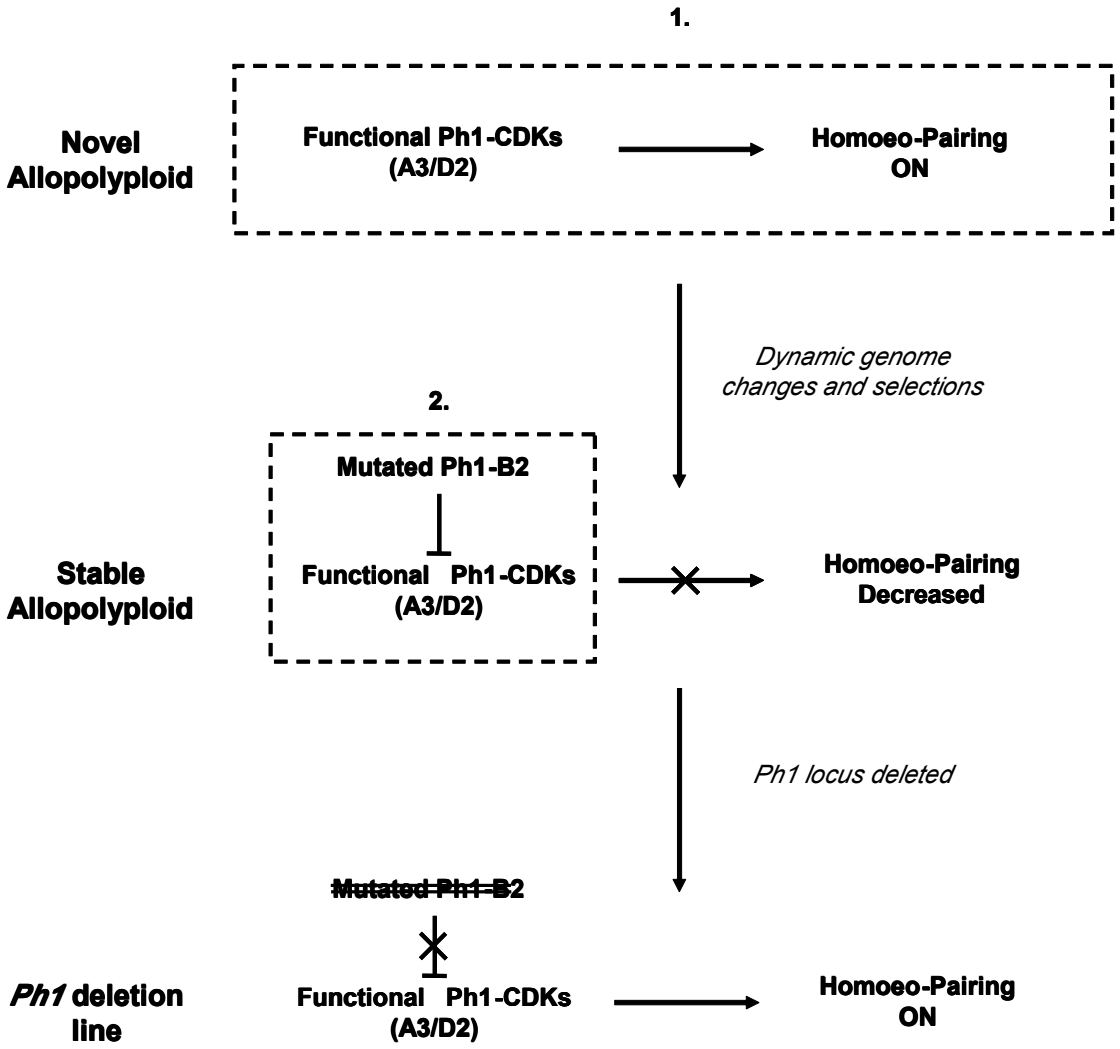


Figure 17 Hypothesis describing the mechanism of *Ph1* locus.

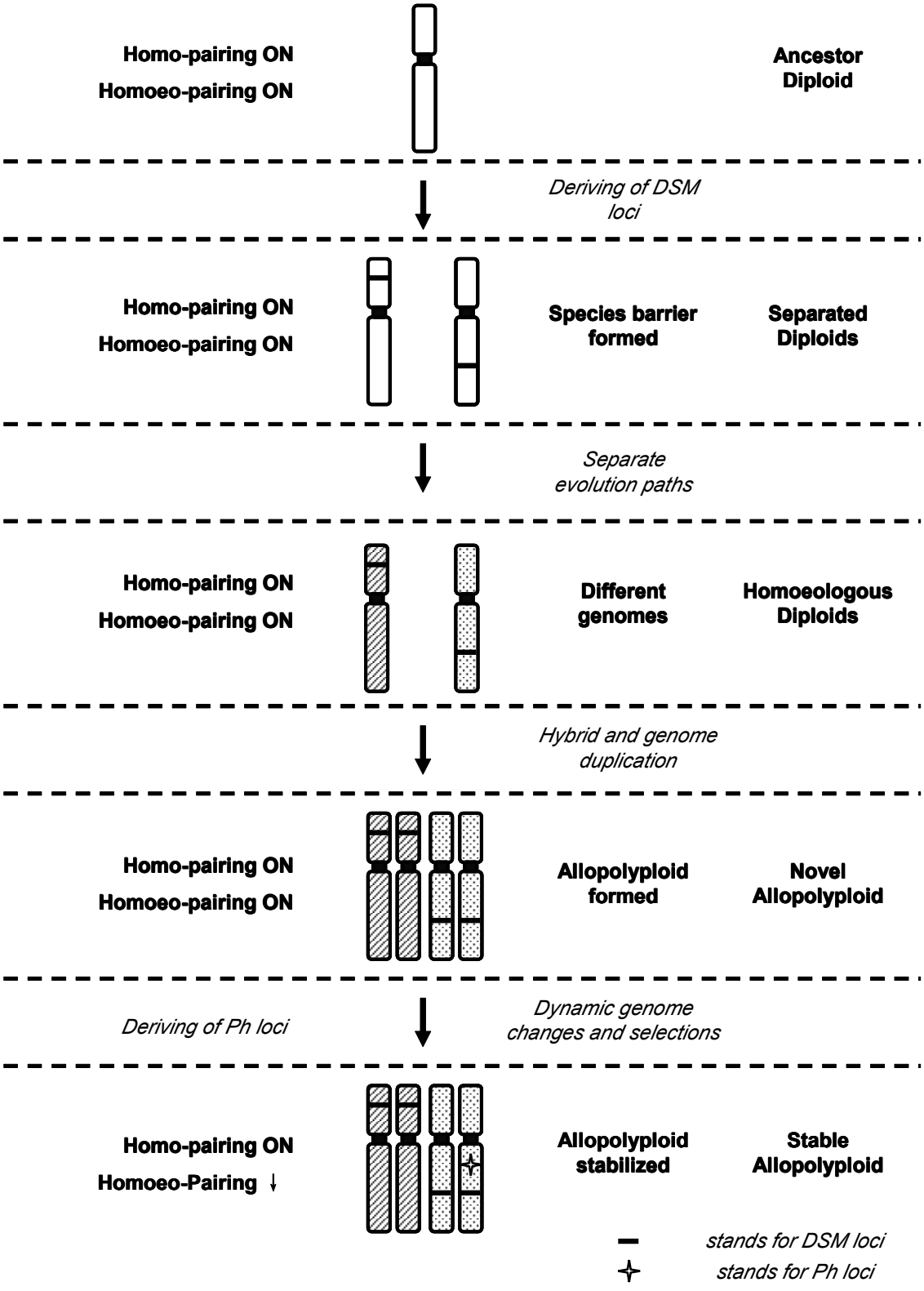


Figure 18 Hypothesis describing the initiation of *Ph1* locus.

(Chapter I, Section 1.2 for details), although chromosomes of the newly formed sub-species were in general homologous. Sub-species went through separate evolutionary paths because of species barriers. Genomic changes built up during the course of separate evolution. After a long period of separation, chromosomes from the two sub-species changed into homoeologous chromosomes. Tetraploid wheat was then formed between these two homoeologous diploid wheat species. At the time that tetraploid wheat was just formed, the genome underwent dynamic rearrangement because of the active homoeologous pairing. Most of the newly formed tetraploid wheat lines were sterile because of the genome instability and homoeologous pairing. The mutated Ph1-CDK, Ph1-B2, was derived from one of the newly formed tetraploid wheat lines. The mutated Ph1-B2 suppressed the activities of functional Ph1-CDKs. The homoeologous pairing level was greatly reduced as a consequence. A diploid like meiosis was therefore established in this line. Hereby, a stable tetraploid wheat line arose.

The homoeologous pairing control gene (HPCG) must be present in all allopolyploid organisms to secure faithful diploid-like chromosome segregation during meiosis. However, it is always under debate how HPCGs were established in polyploids (211). Some argue that HPCGs already existed in the diploid species, since newly formed polyploids can always become stabilised within a few generations, and new gain-of-function genes cannot be produced within such a short period. Some others argue that HPCG should be derived after the formation of polyploid species, since the diploid donors of allopolyploid normally do not show any activity of HPCG. Here the presented hypothesis can embrace facts from both sides. It is unlikely that new functional gene can be gained within several generations of evolution, but mutation of a functional gene, like the frame shift of Ph1-B2, can be reasonably achieved within a short period of time. Such kind of mutation can also happen in the diploid. However, because there is no benefit for diploids to keep such mutated genes (which could even be a disadvantage for diploids), genes of this type are unlikely to spread in diploids like in polyploid species. Therefore, the restriction of homoeologous pairing is not commonly seen in diploids.

Reference

1. Ozkan, H., Brandolini, A., Schäfer-Pregl, R. and Salamini, F. (2002) AFLP analysis of a collection of tetraploid wheats indicates the origin of emmer and hard wheat domestication in southeast Turkey. *Molecular Biology and Evolution*, **19**, 1797.

2. Foote, T., Roberts, M., Kurata, N., Sasaki, T. and Moore, G. (1997) Detailed comparative mapping of cereal chromosome regions corresponding to the Ph1 locus in wheat. *Genetics*, **147**, 801-807.

3. Qu, L.J., Foote, T.N., Roberts, M.A., Money, T.A., Aragon-Alcaide, L., Snape, J.W. and Moore, G. (1998) A simple PCR-based method for scoring the ph1b deletion in wheat. *Theor Appl Genet*, **96**, 371-375.

4. Roberts, M.A., Reader, S.M., Dalgliesh, C., Miller, T.E., Foote, T.N., Fish, L.J., Snape, J.W. and Moore, G. (1999) Induction and characterization of Ph1 wheat mutants. *Genetics*, **153**, 1909-1918.

5. Griffiths, S., Sharp, R., Foote, T.N., Bertin, I., Wanous, M., Reader, S., Colas, I. and Moore, G. (2006) Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature*, **439**, 749-752.

6. Al-Kaff, N., Knight, E., Bertin, I., Foote, T., Hart, N., Griffiths, S. and Moore, G. (2008) Detailed dissection of the chromosomal region containing the Ph1 locus in wheat Triticum aestivum: with deletion mutants and expression profiling. *Ann Bot*, **101**, 863-872.

7. Bateson, W. (1909) In Seward, A. C. (ed.), *Darwin and Modern Science*. Cambridge University Press, Cambridge, pp. 85-101.

8. Dobzhansky, T. (1937) *Genetics and the Origin of Species.* Columbia University Press, New York.

9. Muller, H.J. (1942), *Biol. Symp.*, Vol. 6, pp. 71-125.

10. Bikard, D., Patel, D., Le Mett, C., Giorgi, V., Camilleri, C., Bennett, M.J. and Loudet, O. (2009) Divergent evolution of duplicate genes leads to genetic incompatibilities within A. thaliana. *Science*, **323**, 623.

11. Yamamoto, E., Takashi, T., Morinaka, Y., Lin, S., Wu, J., Matsumoto, T., Kitano, H., Matsuoka, M. and Ashikari, M. Gain of deleterious function causes an autoimmune response and Bateson-Dobzhansky-Muller incompatibility in rice. *Molecular genetics and genomics: MGG*, **283**, 305.

12. Masly, J.P., Jones, C.D., Noor, M.A.F., Locke, J. and Orr, H.A. (2006) Gene transposition as a cause of hybrid sterility in Drosophila. *Science*, **313**, 1448.

13. Werth, C.R. and Windham, M.D. (1991) A model for divergent, allopatric speciation of polyploid pteridophytes resulting from silencing of duplicate-gene expression. *The American Naturalist*, **137**, 515-526.

14. Schnable, P.S. and Wise, R.P. (1998) The molecular basis of cytoplasmic male sterility and fertility restoration. *Trends in plant science*, **3**, 175-180.

15. Duvick, D.N. (1959) The use of cytoplasmic male-sterility in hybrid seed production. *Economic Botany*, **13**, 167-195.

16. Lee, H.Y., Chou, J.Y., Cheong, L., Chang, N.H., Yang, S.Y. and Leu, J.Y. (2008) Incompatibility of nuclear and mitochondrial genomes causes hybrid sterility between two yeast species. *Cell*, **135**, 1065-1073.

17. Bomblies, K., Lempe, J., Epple, P., Warthmann, N., Lanz, C., Dangl, J.L. and Weigel, D. (2007) Autoimmune response as a mechanism for a Dobzhansky-Muller-type incompatibility syndrome in plants. *PLoS Biol*, **5**, e236.

18. Dangl, J.L., Dietrich, R.A. and Richberg, M.H. (1996) Death don't have no mercy: cell death programs in plant-microbe interactions. *The Plant Cell*, **8**, 1793.

19. Khanna-Chopra, R., Dalal, M., Kumar, G.P. and Laloraya, M. (1998) A genetic system involving superoxide causes F1 necrosis in wheat (T. aestivum L.). *Biochemical and biophysical research communications*, **248**, 712.

20. Marubashi, W., Yamada, T. and Niwa, M. (1999) Apoptosis detected in hybrids between nicotiana glutinosa and N. repanda expressing lethality. *Planta*, **210**, 168.

21. Perez, D.E., Wu, C.I., Johnson, N.A. and Wu, M.L. (1993) Genetics of reproductive isolation in the Drosophila simulans clade: DNA marker-assisted mapping and characterization of a hybrid-male sterility gene, Odysseus (Ods). *Genetics*, **134**, 261-275.

22. Bayes, J.J. and Malik, H.S. (2009) Altered heterochromatin binding by a hybrid sterility protein in Drosophila sibling species. *Science*, **326**, 1538-1541.

23. Ferree, P.M. and Barbash, D.A. (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in Drosophila. *PLoS Biol*, **7**, e1000234.

24. Brideau, N.J., Flores, H.A., Wang, J., Maheshwari, S., Wang, X. and Barbash, D.A. (2006) Two Dobzhansky-Muller genes interact to cause hybrid lethality in Drosophila. *Science*, **314**, 1292-1295.

25. Forejt, J. and Iványi, P. (1974) Genetic studies on male sterility of hybrids between laboratory and wild mice (Mus musculus L.). *Genetical Research*, **24**, 189.

26. Mihola, O., Trachtulec, Z., Vlcek, C., Schimenti, J.C. and Forejt, J. (2009) A mouse speciation gene encodes a meiotic histone H3 methyltransferase. *Science*, **323**, 373-375.

27. Hayashi, K., Yoshida, K. and Matsui, Y. (2005) A histone H3 methyltransferase controls epigenetic events required for meiotic prophase. *Nature*, **438**, 374.

28. Myers, S., Bowden, R., Tumian, A., Bontrop, R.E., Freeman, C., MacFie, T.S., McVean, G. and Donnelly, P. Drive against hotspot motifs in primates implicates the PRDM9 gene in meiotic recombination. *Science*, **327**, 876-879.

29. Baudat, F., Buard, J., Grey, C., Fledel-Alon, A., Ober, C., Przeworski, M., Coop, G. and de Massy, B. PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science*, **327**, 836-840.

30. Smagulova, F., Gregoretti, I.V., Brick, K., Khil, P., Camerini-Otero, R.D. and Petukhova, G.V. Genome-wide analysis reveals novel molecular features of mouse recombination hotspots. *Nature*, **472**, 375-378.

31. Oliver, P.L., Goodstadt, L., Bayes, J.J., Birtle, Z., Roach, K.C., Phadnis, N., Beatson, S.A., Lunter, G., Malik, H.S. and Ponting, C.P. (2009) Accelerated evolution of the Prdm9 speciation gene across diverse metazoan taxa. *PLoS Genet*, **5**, e1000753.

32. Bailey, J.A., Gu, Z., Clark, R.A., Reinert, K., Samonte, R.V., Schwartz, S., Adams, M.D., Myers, E.W., Li, P.W. and Eichler, E.E. (2002) Recent segmental duplications in the human genome. *Science*, **297**, 1003.

33. Kellis, M., Birren, B.W. and Lander, E.S. (2004) Proof and evolutionary analysis of ancient genome duplication in the yeast Saccharomyces cerevisiae. *Nature*, **428**, 617-624.

34. Adams, K.L. and Wendel, J.F. (2005) Polyploidy and genome evolution in plants. *Current Opinion in Plant Biology*, **8**, 135-141.

35. Dehal, P. and Boore, J.L. (2005) Two Rounds of Whole Genome Duplication in the Ancestral Vertebrate. *PLoS Biology*, **3**.

36. Jenczewski, E., Eber, F., Grimaud, A., Huet, S., Lucas, M.O., Monod, H. and Chevre, A.M. (2003) PrBn, a major gene controlling homeologous pairing in oilseed rape (Brassica napus) haploids. *Genetics*, **164**, 645-653.

37. Attia, T. and Robbelen, G. (1986) Cytogenetic relationship within cultivated Brassica analyzed in amphihaploids from the three diploid ancestors. *Can J Genet Cytol*, **28**, 323-329.

38. Attia, T. and Robbelen, G. (1986) Meiotic pairing in haploids and amphidiploids of spontaneous versus synthetic origin in rape, Brassica napus L. *L. Can. J. Genet. Cytol*, **28**, 330-334.

39. Liu, Z., Adamczyk, K., Manzanares-Dauleux, M., Eber, F., Lucas, M.O., Delourme, R., Chevre, A.M. and Jenczewski, E. (2006) Mapping PrBn and other quantitative trait loci responsible for the control of homeologous chromosome pairing in oilseed rape (Brassica napus L.) haploids. *Genetics*, **174**, 1583-1596.

40. Sears, E.R. (1976) Genetic control of chromosome pairing in wheat. *Annu Rev Genet*, **10**, 31-51.

41. Sears, E.R. (1977) An induced mutant with homoeologous pairing in common wheat. *Can. J. Genet. Cytol*, **19**, 585-593.

42. Sears, E.R. (1982) A wheat mutation conditioning an intermediate level of homoeologous chromosome pairing. *Genome*, **24**, 715-719.

43. Driscoll, C.J. (1972) Genetic suppression of homoeologous chromosome pairing in hexaploid wheat. *Can. J. Genet. Cytol*, **14**, 39-42.

44. Martinez-Perez, E., Shaw, P. and Moore, G. (2001) The Ph1 locus is needed to ensure specific somatic and meiotic centromere association. *Nature*, **411**, 204-207.

45. Rayssiguier, C., Thaler, D.S. and Radman, M. (1989) The barrier to recombination between Escherichia coli and Salmonella typhimurium is disrupted in mismatch-repair mutants. *Nature*, **342**, 396-401.

46. Hunter, N., Chambers, S.R., Louis, E.J. and Borts, R.H. (1996) The mismatch repair system contributes to meiotic sterility in an interspecific yeast hybrid. *EMBO J*, **15**, 1726-1733.

47. de Wind, N., Dekker, M., Berns, A., Radman, M. and te Riele, H. (1995) Inactivation of the mouse Msh2 gene results in mismatch repair deficiency, methylation tolerance, hyperrecombination, and predisposition to cancer. *Cell*, **82**, 321-330.

48. Okamoto, M. (1957) Asynaptic effect of chromosome V. *Wheat Inf Serv*, **5**.

49. Sears, E.R. and Okamoto, M. (1958), Vol. 2, pp. 258-259.

50. Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat.

51. Sears, E.R. (1956), pp. 1-22.

52. Darlington, C.D. (1940) The prime variables of meiosis. *Biological Reviews*, **15**, 307-322.

53. Bennett, M.D. (1977) The time and duration of meiosis. *Philos Trans R Soc Lond B Biol Sci*, **277**, 201-226.

54. Martinez-Perez, E., Shaw, P., Reader, S., Aragon-Alcaide, L., Miller, T. and Moore, G. (1999) Homologous chromosome pairing in wheat. *J Cell Sci*, **112 (Pt 11)**, 1761-1769.

55. Scherthan, H. (2001) A bouquet makes ends meet. *Nat Rev Mol Cell Biol*, **2**, 621-627.

56. Dawe, R.K., Sedat, J.W., Agard, D.A. and Cande, W.Z. (1994) Meiotic chromosome pairing in maize is associated with a novel chromatin organization. *Cell*, **76**, 901-912.

57. Prieto, P., Shaw, P. and Moore, G. (2004) Homologue recognition during meiosis is associated with a change in chromatin conformation. *Nat Cell Biol*, **6**, 906-908.

58. Veiga-Crespo, P., Poza, M., Prieto-Alcedo, M. and Villa, T.G. (2004) Ancient genes of Saccharomyces cerevisiae. *Microbiology*, **150**, 2221-2227.

59. Tsubouchi, T. and Roeder, G.S. (2005). American Association for the Advancement of Science, Vol. 308, pp. 870-873.

60. Martinez-Perez, E., Shaw, P., Aragon-Alcaide, L. and Moore, G. (2003) Chromosomes form into seven groups in hexaploid and tetraploid wheat as a prelude to meiosis. *Plant J*, **36**, 21-29.

61. Loidl, J., Nairz, K. and Klein, F. (1991) Meiotic chromosome synapsis in a haploid yeast. *Chromosoma*, **100**, 221-228.

62. Jong, A.Y. and Ma, J.J. (1991) Saccharomyces cerevisiae nucleoside-diphosphate kinase: purification, characterization, and substrate specificity. *Arch Biochem Biophys*, **291**, 241-246.

63. Sym, M. and Roeder, G.S. (1995) Zip1-induced changes in synaptonemal complex structure and polycomplex assembly. *J Cell Biol*, **128**, 455-466.

64. Loidl, J., Scherthan, H. and Kaback, D.B. (1994) Physical association between nonhomologous chromosomes precedes distributive disjunction in yeast. *Proc Natl Acad Sci U S A*, **91**, 331-334.

65. Sym, M., Engebrecht, J. and Roeder, G.S. (1993) ZIP1 is a synaptonemal complex protein required for meiotic chromosome synapsis. *Cell*, **72**, 365.

66. Hamant, O., Ma, H. and Cande, W.Z. (2006) Genetics of meiotic prophase I in plants. *Annu Rev Plant Biol*, **57**, 267-302.

67. Armstrong, S.J., Caryl, A.P., Jones, G.H. and Franklin, F.C.H. (2002) Asy1, a protein required for meiotic chromosome synapsis, localizes to axis-associated chromatin in Arabidopsis and Brassica. *Journal of Cell Science*, **115**, 3645-3655.

68. Dong, Y., Bogdanova, A., Habermann, B., Zachariae, W. and Ahringer, J. (2007) Identification of the C. elegans anaphase promoting complex subunit Cdc26 by phenotypic profiling and functional rescue in yeast. *BMC Dev Biol*, **7**, 19.

69. Mikhailova, E.I., Phillips, D., Sosnikhina, S.P., Lovtsyus, A.V., Jones, R.N. and Jenkins, G. (2006) Molecular Assembly of Meiotic Proteins Asy1 and Zyp1 and Pairing Promiscuity in Rye (Secale cereale L.) and Its Synaptic Mutant sy10. *Genetics*, **174**, 1247-1258.

70. Higgins, J.D., Sanchez-Moran, E., Armstrong, S.J., Jones, G.H. and Franklin, F.C.H. (2005) The Arabidopsis synaptonemal complex protein ZYP1 is required for chromosome synapsis and normal fidelity of crossing over. *Genes Dev.*, **19**, 2488-2500.

71. Stacey, N.J., Kuromori, T., Azumi, Y., Roberts, G., Breuer, C., Wada, T., Maxwell, A., Roberts, K. and Sugimoto-Shirasu, K. (2006) Arabidopsis SPO11-2 functions with SPO11-1 in meiotic recombination. *Plant J*, **48**, 206-216.

72. Borde, V., Goldman, A.S.H. and Lichten, M. (2000), Vol. 290, pp. 806-809.

73. Bleuyard, J.-Y., Gallego, M.E. and White, C.I. (2004) Meiotic defects in the Arabidopsis rad50 mutant point to conservation of the MRX complex function in early stages of meiotic recombination. *Chromosoma*, **113**, 197-203.

74. Daoudal-Cotterell, S., Gallego, M.E. and White, C.I. (2002) The plant Rad50–Mre11 protein complex. *FEBS Letters*, **516**, 164-166.

75. Sung, P. (1997) Yeast Rad55 and Rad57 proteins form a heterodimer that functions with replication protein A to promote DNA strand exchange by Rad51 recombinase. *Genes & development*, **11**, 1111-1121.

76. Valerie, K. and Povirk, L.F. (2003) Regulation and mechanisms of mammalian double-strand break repair. *Oncogene*, **22**, 5792-5812.

77. Siaud, N., Dray, E., Gy, I., Gérard, E., Takvorian, N. and Doutriaux, M.P. (2004) Brca2 is involved in meiosis in Arabidopsis thaliana as suggested by its interaction with Dmc1. *The EMBO Journal*, **23**, 1392-1401.

78. Bishop, D.K., Park, D., Xu, L. and Kleckner, N. (1992) DMC1: a meiosis-specific yeast homolog of E. coli recA required for recombination, synaptonemal complex formation, and cell cycle progression. *Cell*, **69**, 439-456.

79. Couteau, F., Belzile, F., Horlow, C., Grandjean, O., Vezon, D. and Doutriaux, M.P. (1999) Random chromosome segregation without meiotic arrest in both male and female meiocytes of a dmc1 mutant of Arabidopsis. *Plant Cell*, **11**, 1623-1634.

80. Hong, E.L., Shinohara, A. and Bishop, D.K. (2001) Saccharomyces cerevisiae Dmc1 protein promotes renaturation of single-strand DNA (ssDNA) and assimilation of ssDNA into homologous super-coiled duplex DNA. *J Biol Chem*, **276**, 41906-41912.

81. Sung, P. (1994) Catalysis of ATP-dependent homologous DNA pairing and strand exchange by yeast RAD51 protein. *Science*, **265**, 1241-1243.

82. Sehorn, M.G., Sigurdsson, S., Bussen, W., Unger, V.M. and Sung, P. (2004) Human meiotic recombinase Dmc1 promotes ATP-dependent homologous DNA strand exchange. *Nature*, **429**, 433-437.

83. Bishop, D.K. (1994) RecA homologs Dmc1 and Rad51 interact to form multiple nuclear complexes prior to meiotic chromosome synapsis. *Cell*, **79**, 1081-1092.

84. Franklin, A.E., McElver, J., Sunjevaric, I., Rothstein, R., Bowen, B. and Cande, W.Z. (1999) Three-Dimensional Microscopy of the Rad51 Recombination Protein during Meiotic Prophase. *Plant Cell*, **11**, 809-824.

85. Keeney, S. (2001) Mechanism and control of meiotic recombination initiation. *Curr Top Dev Biol*, **52**, 1-53.

86. Aragon-Alcaide, L., Reader, S., Beven, A., Shaw, P., Miller, T. and Moore, G. (1997) Association of homologous chromosomes during floral development. *Curr Biol*, **7**, 905-908.

87. Martinez-Perez, E., Shaw, P.J. and Moore, G. (2000) Polyploidy induces centromere association. *J Cell Biol*, **148**, 233-238.

88. Lukaszewski, A.J. (1997) The development and meiotic behavior of asymmetrical isochromosomes in wheat. *Genetics*, **145**, 1155.

89. Corredor, E., Lukaszewski, A.J., Pach 髙, P., Allen, D.C. and Naranjo, T. (2007) Terminal regions of wheat chromosomes select their pairing partners in meiosis. *Genetics*, **177**, 699.

90. Stern, H., Westergaard, M. and Von Wettstein, D. (1975) Presynaptic events in meiocytes of Lilium longiflorum and their relation to crossing-over: a preselection hypothesis. *Proceedings of the National Academy of Sciences*, **72**, 961.

91. Dawe, R.K. (1998) Meiotic chromosome organization and segregation in plants. *Annual Review of Plant Biology*, **49**, 371-395.

92. Colas, I., Shaw, P., Prieto, P., Wanous, M., Spielmeyer, W., Mago, R. and Moore, G. (2008) Effective chromosome pairing requires chromatin remodeling at the onset of meiosis. *Proc Natl Acad Sci U S A*, **105**, 6075-6080.

93. Wall, A.M., Riley, R. and Gale, M.D. (1971) The position of a locus on chromosome 5B of *Triticum aestivum* affecting homoeologous meiotic pairing. *Genetical Research*, **18**, 329-339.

94. Sears, E.R. (1977) An induced mutant with homologous pairing in common wheat. *Genome*, **19**, 585-593.

95. Giorgi, B. (1978) A homoeologous pairing mutant isolated in Triticum durum cv Cappelli. *Mutation Breeding Newsletter*, **11**, 4-5.

96. Sidhu, G.K., Rustgi, S., Shafqat, M.N., von Wettstein, D. and Gill, K.S. (2008) Fine structure mapping of a gene-rich region of wheat carrying Ph1, a suppressor of crossing over between homoeologous chromosomes. *Proc Natl Acad Sci U S A*, **105**, 5815-5820.

97. Kawaura, K., Mochida, K. and Ogihara, Y. (2005) Expression profile of two storage-protein gene families in hexaploid wheat revealed by large-scale analysis of expressed sequence tags. *Plant Physiol*, **139**, 1870-1880.

98. Menges, M., De Jager, S.M., Gruissem, W. and Murray, J.A.H. (2005) Global analysis of the core cell cycle regulators of Arabidopsis identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control. *The Plant Journal*, **41**, 546-566.

99. Santamaría, D., Barrière, C., Cerqueira, A., Hunt, S., Tardy, C., Newton, K., Cáceres, J.F., Dubus, P., Malumbres, M. and Barbacid, M. (2007) Cdk1 is sufficient to drive the mammalian cell cycle. *Nature*, **448**, 811-815.

100. Inzé, D. and De Veylder, L. (2006) Cell Cycle Regulation in Plant Development 1. *Annu. Rev. Genet.*, **40**, 77-105.

101. Connell-Crowley, L., Solomon, M., Wei, N. and Harper, J. (1993) Phosphorylation independent activation of human cyclin-dependent kinase 2 by cyclin A in vitro. *Molecular biology of the cell*, **4**, 79.

102. Russo, A.A., Jeffrey, P.D. and Pavletich, N.P. (1996) Structural basis of cyclin-dependent kinase activation by phosphorylation. *Nature Structural & Molecular Biology*, **3**, 696-700.

103. Shimotohno, A., Matsubayashi, S., Yamaguchi, M., Uchimiya, H. and Umeda, M. (2003) Differential phosphorylation activities of CDK-activating kinases in Arabidopsis thaliana. *FEBS Letters*, **534**, 69-74.

104. Shimotohno, A., Umeda-Hara, C., Bisova, K., Uchimiya, H. and Umeda, M. (2004) The plant-specific kinase CDKF; 1 is involved in activating phosphorylation of cyclin-dependent kinase-activating kinases in Arabidopsis. *The Plant Cell Online*, **16**, 2954-2966.

105. Takatsuka, H., Ohno, R. and Umeda, M. (2009) The Arabidopsis cyclin‐dependent kinase‐activating kinase CDKF; 1 is a major regulator of cell proliferation and cell expansion but is dispensable for CDKA activation. *The Plant Journal*, **59**, 475-487.

106. Morgan, D.O. (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annual review of cell and developmental biology*, **13**, 261-291.

107. Sherr, C.J. and Roberts, J.M. (1999) CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes & development*, **13**, 1501.

108. Verkest, A., Weinl, C., Inzé, D., De Veylder, L. and Schnittger, A. (2005) Switching the cell cycle. Kip-related proteins in plant cell cycle control. *Plant Physiology*, **139**, 1099.

109. De Schutter, K., Joubès, J., Cools, T., Verkest, A., Corellou, F., Babiychuk, E., Van Der Schueren, E., Beeckman, T., Kushnir, S. and Inzé, D. (2007) Arabidopsis WEE1 kinase controls cell cycle arrest in response to activation of the DNA integrity checkpoint. *The Plant Cell Online*, **19**, 211-225.

110. Dissmeyer, N., Weimer, A.K., Pusch, S., De Schutter, K., Kamei, C.L.A., Nowack, M.K., Novak, B., Duan, G.L., Zhu, Y.G. and De Veylder, L. (2009) Control of cell proliferation, organ growth, and DNA damage response operate independently of dephosphorylation of the Arabidopsis Cdk1 homolog CDKA; 1. *The Plant Cell Online*, **21**, 3641-3654.

111. Tyers, M., Tokiwa, G. and Futcher, B. (1993) Comparison of the Saccharomyces cerevisiae G1 cyclins: Cln3 may be an upstream activator of Cln1, Cln2 and other cyclins. *The EMBO Journal*, **12**, 1955.

112. Nasmyth, K. and Dirick, L. (1991) The role of SWI4 and SWI6 in the activity of G1 cyclins in yeast. *Cell*, **66**, 995-1013.

113. Amon, A., Tyers, M., Futcher, B. and Nasmyth, K. (1993) Mechanisms that help the yeast cell cycle clock tick: G2 cyclins transcriptionally activate G2 cyclins and repress G1 cyclins. *Cell*, **74**, 993-1007.

114. Schwob, E. and Nasmyth, K. (1993) CLB5 and CLB6, a new pair of B cyclins involved in DNA replication in Saccharomyces cerevisiae. *Genes & development*, **7**, 1160.

115. Zachariae, W., Schwab, M., Nasmyth, K. and Seufert, W. (1998) Control of cyclin ubiquitination by CDK-regulated binding of Hct1 to the anaphase promoting complex. *Science*, **282**, 1721.

116. Yoshida, S., Asakawa, K. and Toh-e, A. (2002) Mitotic exit network controls the localization of Cdc14 to the spindle pole body in Saccharomyces cerevisiae. *Current biology*, **12**, 944-950.

117. Bardin, A.J. and Amon, A. (2001) MEN and SIN: what's the difference? *Nature Reviews Molecular Cell Biology*, **2**, 815-826.

118. Champion, A., Kreis, M., Mockaitis, K., Picaud, A. and Henry, Y. (2004) Arabidopsis kinome: after the casting. *Functional & integrative genomics*, **4**, 163-187.

119. Dardick, C., Chen, J., Richter, T., Ouyang, S. and Ronald, P. (2007) The rice kinase database. A phylogenomic database for the rice kinome. *Plant Physiology*, **143**, 579.

120. Kobe, B., Kampmann, T., Forwood, J.K., Listwan, P. and Brinkworth, R.I. (2005) Substrate specificity of protein kinases and computational prediction of substrates. *Biochimica et Biophysica Acta (BBA)-Proteins & Proteomics*, **1754**, 200-209.

121. Brinkworth, R.I., Breinl, R.A. and Kobe, B. (2003) Structural basis and prediction of substrate specificity in protein serine/threonine kinases. *Proceedings of the National Academy of Sciences of the United States of America*, **100**, 74.

122. Edgar, R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, **32**, 1792.

123. Katoh, K. and Toh, H. (2008) Recent developments in the MAFFT multiple sequence alignment program. *Briefings in Bioinformatics*, **9**, 286.

124. Castresana, J. (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Molecular Biology and Evolution*, **17**, 540.

125. Guindon, S., Dufayard, J.F., Hordijk, W., Lefort, V. and Gascuel, O. (2009) PhyML: fast and accurate phylogeny reconstruction by maximum likelihood. *Infection Genetics and Evolution*, **9**, 384-385.

126. Anisimova, M. and Gascuel, O. (2006) Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Systematic Biology*, **55**, 539.

127. Stover, B. and Muller, K. TreeGraph 2: Combining and visualizing evidence from different phylogenetic analyses. *BMC Bioinformatics*, **11**, 7.

128. Durbin, R. (1998) *Biological sequence analysis: Probabilistic models of proteins and nucleic acids*. Cambridge Univ Pr.

129. Schuster-Böckler, B., Schultz, J. and Rahmann, S. (2004) HMM Logos for visualization of protein families. *BMC Bioinformatics*, **5**, 7.

130. Longtine, M.S., McKenzie, A., 3rd, Demarini, D.J., Shah, N.G., Wach, A., Brachat, A., Philippsen, P. and Pringle, J.R. (1998) Additional modules for versatile and economical PCR-based gene deletion and modification in Saccharomyces cerevisiae. *Yeast*, **14**, 953-961.

131. Hirt, H., Pay, A., Gyorgyey, J., Bako, L., Nemeth, K., Bogre, L., Schweyen, R.J., Heberle-Bors, E. and Dudits, D. (1991) Complementation of a yeast cell cycle mutant by an alfalfa cDNA encoding a protein kinase homologous to p34cdc2. *Proc Natl Acad Sci U S A*, **88**, 1636-1640.

132. Hashimoto, J., Hirabayashi, T., Hayano, Y., Hata, S., Ohashi, Y., Suzuka, I., Utsugi, T., Toh-e, A. and Kikuchi, Y. (1992) Isolation and characterization of cDNA clones encoding cdc2 homologues from Oryza sativa: a functional homologue and cognate variants. *Mol Gen Genet*, **233**, 10-16.

133. Chu, S., DeRisi, J., Eisen, M., Mulholland, J., Botstein, D., Brown, P.O. and Herskowitz, I. (1998) The transcriptional program of sporulation in budding yeast. *Science*, **282**, 699-705.

134. Primig, M., Williams, R.M., Winzeler, E.A., Tevzadze, G.G., Conway, A.R., Hwang, S.Y., Davis, R.W. and Esposito, R.E. (2000) The core meiotic transcriptome in budding yeasts. *Nat Genet*, **26**, 415-423.

135. Vershon, A.K., Hollingsworth, N.M. and Johnson, A.D. (1992) Meiotic induction of the yeast HOP1 gene is controlled by positive and negative regulatory sites. *Mol Cell Biol*, **12**, 3706-3714.

136. Williams, R.M., Primig, M., Washburn, B.K., Winzeler, E.A., Bellis, M., Sarrauste de Menthiere, C., Davis, R.W. and Esposito, R.E. (2002) The Ume6 regulon coordinates metabolic and meiotic gene expression in yeast. *Proc Natl Acad Sci U S A*, **99**, 13431-13436.

137. Hui, C.M., Campistrous, A. and Stuart, D.T. (2002) Purification and some properties of Saccharomyces cerevisiae meiosis-specific protein kinase Ime2. *Protein Expr Purif*, **26**, 416-424.

138. Sherr, C.J. and Roberts, J.M. (1999) CDK inhibitors: positive and negative regulators of G1-phase progression. *Genes Dev*, **13**, 1501 - 1512.

139. Koff, A., Giordano, A., Desai, D., Yamashita, K., Harper, J.W., Elledge, S., Nishimoto, T., Morgan, D.O., Franza, B.R. and Roberts, J.M. (1992) Formation and activation of a cyclin E-cdk2 complex during the G1 phase of the human cell cycle. *Science*, **257**, 1689-1694.

140. Dulic, V., Lees, E. and Reed, S.I. (1992) Association of human cyclin E with a periodic G1-S phase protein kinase. *Science*, **257**, 1958-1961.

141. Satyanarayana, A., Berthet, C., Lopez-Molina, J., Coppola, V., Tessarollo, L. and Kaldis, P. (2008) Genetic substitution of Cdk1 by Cdk2 leads to embryonic lethality and loss of meiotic function of Cdk2. *Development*, **135**, 3389-3400.

142. Szwarcwort-Cohen, M., Kasulin-Boneh, Z., Sagee, S. and Kassir, Y. (2009) Human Cdk2 is a functional homolog of budding yeast Ime2, the meiosis-specific Cdk-like kinase. *Cell Cycle*, **8**, 647-654.

143. Mao-Draayer, Y., Galbraith, A.M., Pittman, D.L., Cool, M. and Malone, R.E. (1996) Analysis of meiotic recombination pathways in the yeast Saccharomyces cerevisiae. *Genetics*, **144**, 71-86.

144. Niu, H., Li, X., Job, E., Park, C., Moazed, D., Gygi, S.P. and Hollingsworth, N.M. (2007) Mek1 kinase is regulated to suppress double-strand break repair between sister chromatids during budding yeast meiosis. *Mol Cell Biol*, **27**, 5456-5467.

145. Niu, H., Wan, L., Baumgartner, B., Schaefer, D., Loidl, J. and Hollingsworth, N.M. (2005) Partner choice during meiosis is regulated by Hop1-promoted dimerization of Mek1. *Mol Biol Cell*, **16**, 5804-5818.

146. Loidl, J., Klein, F. and Scherthan, H. (1994) Homologous pairing is reduced but not abolished in asynaptic mutants of yeast. *J Cell Biol*, **125**, 1191-1200.

147. Hollingsworth, N.M. and Byers, B. (1989) HOP1: a yeast meiotic pairing gene. *Genetics*, **121**, 445-462.

148. Carballo, J.A., Johnson, A.L., Sedgwick, S.G. and Cha, R.S. (2008) Phosphorylation of the axial element protein Hop1 by Mec1/Tel1 ensures meiotic interhomolog recombination. *Cell*, **132**, 758-770.

149. Boden, S.A., Langridge, P., Spangenberg, G. and Able, J.A. (2009) TaASY1 promotes homologous chromosome interactions and is affected by deletion of Ph1. *Plant J*, **57**, 487-497.

150. Elledge, S.J. and Spottswood, M.R. (1991) A new human p34 protein kinase, CDK2, identified by complementation of a cdc28 mutation in Saccharomyces cerevisiae, is a homolog of Xenopus Eg1. *EMBO J*, **10**, 2653-2659.

151. Ninomiya-Tsuji, J., Nomoto, S., Yasuda, H., Reed, S.I. and Matsumoto, K. (1991) Cloning of a human cDNA encoding a CDC2-related kinase by complementation of a budding yeast cdc28 mutation. *Proc Natl Acad Sci U S A*, **88**, 9006-9010.

152. Colasanti, J., Tyers, M. and Sundaresan, V. (1991) Isolation and characterization of cDNA clones encoding a functional p34cdc2 homologue from Zea mays. *Proc Natl Acad Sci U S A*, **88**, 3377-3381.

153. Hirt, H., Pay, A., Bogre, L., Meskiene, I. and Heberle-Bors, E. (1993) cdc2MsB, a cognate cdc2 gene from alfalfa, complements the G1/S but not the G2/M transition of budding yeast cdc28 mutants. *Plant J*, **4**, 61-69.

154. Teste, M.A., Duquenne, M., Francois, J.M. and Parrou, J.L. (2009) Validation of reference genes for quantitative expression analysis by real-time RT-PCR in Saccharomyces cerevisiae. *BMC Mol Biol*, **10**, 99.

155. Smith, H.E. and Mitchell, A.P. (1989) A transcriptional cascade governs entry into meiosis in Saccharomyces cerevisiae. *Mol Cell Biol*, **9**, 2142-2152.

156. Ruijter, J.M., Ramakers, C., Hoogaars, W.M.H., Karlen, Y., Bakker, O., Van Den Hoff, M.J.B. and Moorman, A.F.M. (2009) Amplification efficiency: linking baseline and bias in the analysis of quantitative PCR data. *Nucleic acids research*, **37**, e45.

157. Kew, R.B.G., Gardens, V.K. and Wakehurst, V.

158. Draper, J., Mur, L.A., Jenkins, G., Ghosh-Biswas, G.C., Bablak, P., Hasterok, R. and Routledge, A.P. (2001) Brachypodium distachyon. A new model system for functional genomics in grasses. *Plant Physiol*, **127**, 1539-1555.

159. Pacurar, D.I., Thordal-Christensen, H., Nielsen, K.K. and Lenk, I. (2008) A high-throughput Agrobacterium-mediated transformation system for the grass model species Brachypodium distachyon L. *Transgenic Res*, **17**, 965-975.

160. Vogel, J. and Hill, T. (2008) High-efficiency Agrobacterium-mediated transformation of Brachypodium distachyon inbred line Bd21-3. *Plant Cell Rep*, **27**, 471-478.

161. Foote, T.N., Griffiths, S., Allouis, S. and Moore, G. (2004) Construction and analysis of a BAC library in the grass Brachypodium sylvaticum: its

use as a tool to bridge the gap between rice and wheat in elucidating gene content. *Funct Integr Genomics*, **4**, 26-33.

162. Huo, N., Gu, Y.Q., Lazo, G.R., Vogel, J.P., Coleman-Derr, D., Luo, M.C., Thilmony, R., Garvin, D.F. and Anderson, O.D. (2006) Construction and characterization of two BAC libraries from Brachypodium distachyon, a new model for grass genomics. *Genome*, **49**, 1099-1108.

163. Farrar, K. and Donnison, I.S. (2007) Construction and screening of BAC libraries made from Brachypodium genomic DNA. *Nat Protoc*, **2**, 1661-1674.

164. Garvin, D.F., McKenzie, N., Vogel, J.P., Mockler, T.C., Blankenheim, Z.J., Wright, J., Cheema, J.J., Dicks, J., Huo, N., Hayden, D.M. *et al.* An SSR-based genetic linkage map of the model grass Brachypodium distachyon. *Genome*, **53**, 1-13.

165. Febrer, M., Goicoechea, J.L., Wright, J., McKenzie, N., Song, X., Lin, J., Collura, K., Wissotski, M., Yu, Y. and Ammiraju, J.S.S. An Integrated Physical, Genetic and Cytogenetic Map of Brachypodium distachyon, a Model System for Grass Research. *PLoS One*, **5**.

166. Thole, V., Worland, B., Wright, J., Bevan, M.W. and Vain, P. Distribution and characterization of more than 1000 T-DNA tags in the genome of Brachypodium distachyon community standard line Bd21. *Plant Biotechnol J*, **8**, 734-747.

167. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature*, **463**, 763-768.

168. Aragon-Alcaide, L., Miller, T., Schwarzacher, T., Reader, S. and Moore, G. (1996) A cereal centromeric sequence. *Chromosoma*, **105**, 261-268.

169. Jiang, J., Nasuda, S., Dong, F., Scherrer, C.W., Woo, S.S., Wing, R.A., Gill, B.S. and Ward, D.C. (1996) A conserved repetitive DNA element located in the centromeres of cereal chromosomes. *Proceedings of the National Academy of Sciences of the United States of America*, **93**, 14210.

170. Miller, J.T., Dong, F., Jackson, S.A., Song, J. and Jiang, J. (1998) Retrotransposon-Related DNA Sequences in the Centromeres of Grass Chromosomes. *Genetics*, **150**, 1615-1623.

171. Presting, G.G., Malysheva, L., Fuchs, J. and Schubert, I. (1998) A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. *The Plant journal: for cell and molecular biology*, **16**, 721.

172. Nagaki, K., Song, J., Stupar, R.M., Parokonny, A.S., Yuan, Q., Ouyang, S., Liu, J., Hsiao, J., Jones, K.M., Dawe, R.K. *et al.* (2003) Molecular and cytological analyses of large tracks of centromeric DNA reveal the structure and evolutionary dynamics of maize centromeres. *Genetics*, **163**, 759-770.

173. Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R. and Jiang, J. (2002) Functional Rice Centromeres Are Marked by a Satellite Repeat and a Centromere-Specific Retrotransposon. *The Plant Cell Online*, **14**, 1691.

174. Hass, B.L., Pires, J.C., Porter, R., Phillips, R.L. and Jackson, S.A. (2003) Comparative genetics at the gene and chromosome levels between rice (Oryza sativa) and wildrice (Zizania palustris). *TAG. Theoretical and applied genetics. Theoretische und angewandte Genetik*, **107**, 773.

175. Bao, W., Zhang, W., Yang, Q., Zhang, Y., Han, B., Gu, M., Xue, Y. and Cheng, Z. (2006) Diversity of centromeric repeats in two closely related wild rice species, Oryza officinalis and Oryza rhizomatis. *Mol Genet Genomics*, **275**, 421-430.

176. Zhang, W., Yi, C., Bao, W., Liu, B., Cui, J., Yu, H., Cao, X., Gu, M., Liu, M. and Cheng, Z. (2005) The transcribed 165-bp CentO satellite is the major functional centromeric element in the wild rice species Oryza punctata. *Plant Physiol*, **139**, 306-315.

177. Kamm, A., Galasso, I., Schmidt, T. and Heslop-Harrison, J.S. (1995) Analysis of a repetitive DNA family from Arabidopsis arenosa and relationships between Arabidopsis species. *Plant molecular biology*, **27**, 853.

178. Lim, K.B., Yang, T.J., Hwang, Y.J., Kim, J.S., Park, J.Y., Kwon, S.J., Kim, J.A., Choi, B.S., Lim, M.H. and Jin, M. (2007) Characterization of the centromere and peri centromere retrotransposons in Brassica rapa and their distribution in related Brassica species. *The Plant Journal*, **49**, 173-183.

179. Ansari, H.A., Ellison, N.W., Griffiths, A.G. and Williams, W.M. (2004) A lineage-specific centromeric satellite sequence in the genus Trifolium. *Chromosome Research*, **12**, 357-367.

180. Han, Y., Wang, G., Liu, Z., Liu, J., Yue, W., Song, R., Zhang, X. and Jin, W. Divergence in centromere structure distinguishes related genomes in Coix lacryma-jobi and its wild relative. *Chromosoma*, **119**, 89-98.

181. Zhang, Y., Huang, Y., Zhang, L., Li, Y., Lu, T., Lu, Y., Feng, Q., Zhao, Q., Cheng, Z., Xue, Y. *et al.* (2004) Structural features of the rice chromosome 4 centromere. *Nucleic Acids Res*, **32**, 2023-2030.

182. Wu, J., Yamagata, H., Hayashi-Tsugane, M., Hijishita, S., Fujisawa, M., Shibata, M., Ito, Y., Nakamura, M., Sakaguchi, M., Yoshihara, R. *et al.* (2004) Composition and Structure of the Centromeric Region of Rice Chromosome 8. *The Plant Cell Online*, **16**, 967-976.

183. Fransz, P.F., Armstrong, S., de Jong, J.H., Parnell, L.D., van Drunen, C., Dean, C., Zabel, P., Bisseling, T. and Jones, G.H. (2000) Integrated Cytogenetic Map of Chromosome Arm 4S of A. thaliana:: Structural Organization of Heterochromatic Knob and Centromere Region. *Cell*, **100**, 367-376.

184. Kumekawa, N., Hosouchi, T., Tsuruoka, H. and Kotani, H. (2001) The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 4. *DNA research*, **8**, 285.

185. Nagaki, K., Talbert, P.B., Zhong, C.X., Dawe, R.K., Henikoff, S. and Jiang, J. (2003) Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of Arabidopsis thaliana centromeres. *Genetics*, **163**, 1221-1225.

186. Nagaki, K., Cheng, Z., Ouyang, S., Talbert, P.B., Kim, M., Jones, K.M., Henikoff, S., Buell, C.R. and Jiang, J. (2004) Sequencing of a rice centromere uncovers active genes. *Nature genetics*, **36**, 138-145.

187. Steiner, N.C. and Clarke, L. (1994) A novel epigenetic effect can alter centromere function in fission yeast. *Cell*, **79**, 865-874.

188. Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K. and Willard, H.F. (1997) Formation of de novo centromeres and construction of first-generation human artificial microchromosomes. *Nature genetics*, **15**, 345-355.

189. Takahashi, K., Chen, E.S. and Yanagida, M. (2000) Requirement of Mis6 centromere connector for localizing a CENP-A-like protein in fission yeast. *Science*, **288**, 2215.

190. Blower, M.D. and Karpen, G.H. (2001) The role of Drosophila CID in kinetochore formation, cell-cycle progression and heterochromatin interactions. *Nature cell biology*, **3**, 730-739.

191. Barry, A.E., Howman, E.V., Cancilla, M.R., Saffery, R. and Andy Choo, K.H. (1999) Sequence analysis of an 80 kb human neocentromere. *Human molecular genetics*, **8**, 217.

192. Sullivan, B.A. and Schwartz, S. (1995) Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. *Human molecular genetics*, **4**, 2189.

193. Agudo, M., Abad, J.P., Molina, I., Losada, A., Ripoll, P. and Villasante, A. (2000) A dicentric chromosome of Drosophila melanogaster showing alternate centromere inactivation. *Chromosoma*, **109**, 190-196.

194. van den Engh, G., Sachs, R. and Trask, B.J. (1992) Estimating genomic distance from DNA sequence location in cell nuclei by a random walk model. *Science*, **257**, 1410.

195. Wegel, E., Koumproglou, R., Shaw, P. and Osbourn, A. (2009) Cell Type-Specific Chromatin Decondensation of a Metabolic Gene Cluster in Oats. *The Plant Cell*, **21**, 3926.

196. Chambeyron, S. and Bickmore, W.A. (2004) Chromatin decondensation and nuclear reorganization of the HoxB locus upon induction of transcription. *Genes Dev*, **18**, 1119-1130.

197. Jenkins, G. and Hasterok, R. (2007) BAC 'landing' on chromosomes of Brachypodium distachyon for comparative genome alignment. *Nat Protoc*, **2**, 88-98.

198. Aragon Alcaide, L., Beven, A., Moore, G. and Shaw, P. (1998) The use of vibratome sections of cereal spikelets to study anther development and meiosis. *The Plant Journal*, **14**, 503-508.

199. Prieto, P., Moore, G. and Shaw, P. (2007) Fluorescence in situ hybridization on vibratome sections of plant tissues. *Nat Protoc*, **2**, 1831-1838.

200. Murray, F., Brettell, R., Matthews, P., Bishop, D. and Jacobsen, J. (2004) Comparison of Agrobacterium-mediated transformation of four barley cultivars using the GFP and GUS reporter genes. *Plant cell reports*, **22**, 397-402.

201. Ijdo, J.W., Wells, R.A., Baldini, A. and Reeders, S.T. (1991) Improved telomere detection using a telomere repeat probe (TTAGGG) n generated by PCR. *Nucleic acids research*, **19**, 4780.

202. Berr, A. and Schubert, I. (2006) Direct labelling of BAC-DNA by rolling-circle amplification. *The Plant Journal*, **45**, 857-862.

203. Alves, S.C., Worland, B., Thole, V., Snape, J.W., Bevan, M.W. and Vain, P. (2009) A protocol for Agrobacterium-mediated transformation of Brachypodium distachyon community standard line Bd21. *Nat Protoc*, **4**, 638-649.

204. Tzfira, T., Kozlovsky, S.V. and Citovsky, V. (2007) Advanced expression vector systems: new weapons for plant research and biotechnology. *Plant Physiology*, **145**, 1087.

205. Weigel, D. and Glazebrook, J. Transformation of Agrobacterium using the freeze-thaw method. *Cold Spring Harbor Protocols*, **2006**, pdb. prot4666.

206. Yousafzai, F.K., Al-Kaff, N. and Moore, G. (2010) Structural and functional relationship between the Ph1 locus protein 5B2 in wheat and CDK2 in mammals. *Funct Integr Genomics*, **10**, 157-166.

207. Harashima, H., Shinmyo, A. and Sekine, M. (2007) Phosphorylation of threonine 161 in plant cyclin dependent kinase A is required for cell division by activation of its associated kinase. *The Plant Journal*, **52**, 435-448.

208. Wang, G., Kong, H., Sun, Y., Zhang, X., Zhang, W., Altman, N., dePamphilis, C.W. and Ma, H. (2004) Genome-Wide Analysis of the Cyclin Family in Arabidopsis and Comparative Phylogenetic Analysis of Plant Cyclin-Like Proteins. *Plant Physiol.*, **135**, 1084-1099.

209. Moore, G. (2002) Meiosis in allopolyploids-the importance of 'Teflon' chromosomes. *Trends in genetics*, **18**, 456-463.

210. Sanchez-Moran, E., Benavente, E. and Orellana, J. (2001) Analysis of karyotypic stability of homoeologous-pairing (ph) mutants in allopolyploid wheats. *Chromosoma*, **110**, 371-377.

211. Jenczewski, E. and Alix, K. (2004) From diploids to allopolyploids: the emergence of efficient pairing control genes in plants. *Critical reviews in plant sciences*, **23**, 21-45.