RKD genes: a novel transcription factor family involved in the female gametophyte development of *Arabidopsis* and wheat

Dissertation

zur Erlangung des akademischen Grades doctor rerum naturalium (Dr. rer. nat.)

vorgelegt der

Mathematisch-Naturwissenschaftlich-Technischen Fakultät

(mathematisch-naturwissenschaftlicher Bereich)

Martin-Luther-Universität Halle-Wittenberg

von

Dávid Kőszegi

geboren am 20. Februar 1979 in Budapest, Ungarn

Gutachter :
1. Prof. Dr. Gerd Jürgens
2. Prof. Dr. Gunter Reuter
3. Prof. Dr. Ulrich Wobus

Halle (Saale), den 25. September 2008

To my wife and my family

„Mert az őssejtig vagyok minden ős-
Az Ős vagyok, mely sokasodni foszlik:
Apám- s anyámmá válok boldogon,
S apám, anyám maga is kettéoszlik
S én lelkes Eggyé így szaporodom!”

“that I am every parent in the boundless
succession to the primal lonely cell.
till I become my father and mother,
then father splits and mother, procreating
the multiplying me and none other!”

(Dunánál, By the Danube, Attila József,
Hungarian poet, 1905-1937)

Table of contents

List of abbreviations

bp	base pair(s)
CaMV	cauliflower mosaic virus
cDNA	complementary DNA
DAP	days after pollination
DNA	deoxyribonucleic acid
dNTP	deoxy nucleotide triphosphate
EDTA	ethylenediamine tetraacetic acid
EST	expressed sequence tags
FG	female gametophyte
GAPDH	glyceraldehyde 3-phosphate dehydrogenase
GFP	green fluorescence protein
GTP	guanosine triphosphate
GUS	beta-glucuronidase
HPT	hygromycin phospho-transferase
IPTG	Isopropyl β-D-1-thiogalactopyranoside
mRNA	messenger RNA
NBT/BCIP	nitro blue tetrazolium / 5-Bromo-4-chloro-3-indolyl phosphate
OD	optical density
PCR	polymerase chain reaction
RACE	rapid amplification of cDNA ends
RGA	Rho GTPase activator
RNA	ribonucleic acid
RNAi	RNA interference
RT-PCR	reverse transcriptase polymerase chain reaction
SDS-PAGE	sodium dodecyl sulfate polyacrylamide gel electrophoresis
t35S	terminator 35S
T-DNA	transfer DNA
X-gluc	5-bromo-4-chloro-3-indolyl glucuronide
WT	wild type

1. INTRODUCTION

1.1. Alternation of generations – speciality of plants: gametophyte proliferation

The typical life cycle of plants is characterised by the alternation of two generations, the gametophyte and sporophyte. The function of the sporophyte is to produce meiotically reduced micro- and/or macrospores, which develop into the male microgametophyte and the female macrogametophyte, respectively. The gametophytes generate gametes, which - after fertilisation and zygote formation - develop into the sporophyte, closing the cycle (Fig. 1-1). Since the discovery of these alternating plant generations by Wilhelm Hofmeister (Hofmeister, 1851) more than 150 years ago, reproductive structures and processes of heterosporic ferns and seed plants are considered as homologous developmental pathways. This includes homologous pairs like megasporophyll and carpel, megasporangium and nucellus, megaspore mother cell and embryo sac mother cell, megaspore and embryo sac cell, as well as megaprothallium and the primary endosperm of gymnosperms and the developed embryo sac of angiosperms with antipodals, polar cell, egg cell and synergids (egg apparatus). A trend in plant evolution from mosses to ferns, gymnosperms and to angiosperms includes the reduction of the gametophyte generation to three cells in the case of the male gametophyte, and the dominance of the sporophyte generation.

1.2. Development of the female sexual organs

The male gametophyte develops within anthers and consists of one vegetative cell and two sperm cells. The female gametophyte (FG) is embedded within the ovule, which consists of the funiculus, the two integuments and the nucellus. The FG plays essential role in nearly all steps of the sexual reproductive process. (I) Signals from the ovule lead the growing pollen tube into the micropyle (pollen tube guidance) (Higashiyama et al., 2001; Márton et al., 2005). (II) During double fertilisation female cytoskeletal elements help the sperm cells to the egg cell and to the central cell. (III) Embryogenesis takes place in the developing seed originated from the ovule.

The development of the female gametophyte can be divided into two parts: the megasporogenesis and the megagametogenesis (Fig. 1-2). A subepidermal nucellar cell becomes selected to differentiate into the megaspore mother cell (MMC) with enlarged nucleus. The MMC undergoes meiosis generating four haploid megaspores, also called a tetrad. The three megaspores closest to the micropyle undergo programmed cell death, whereas the functional megaspore (FM) survives. The germination of the functional megaspore initiates megagametogenesis which

characteristically includes three haploid mitotic divisions generating eight nuclei. After cellularisation, seven cells can be found in the mature embryo sac: the egg cell flanked by two synergids, three antipodal cells and the central cell with two nuclei fused to a diploid nucleus. The egg cell and the diploid central cell are fertilised independently by one haploid sperm cell each, giving rise to the diploid embryo and triploid endosperm, respectively. This event, called double fertilisation is the speciality of angiosperm sexual reproduction (Friedman, 2001). The described *Polygonum* type of embryo sac development can vary greatly in different species. For instance, the three antipodal cells undergo cell death prior to fertilisation in the model plant *Arabidopsis*, whereas in the antipodals proliferate up to several dozens in *Gramineae.*

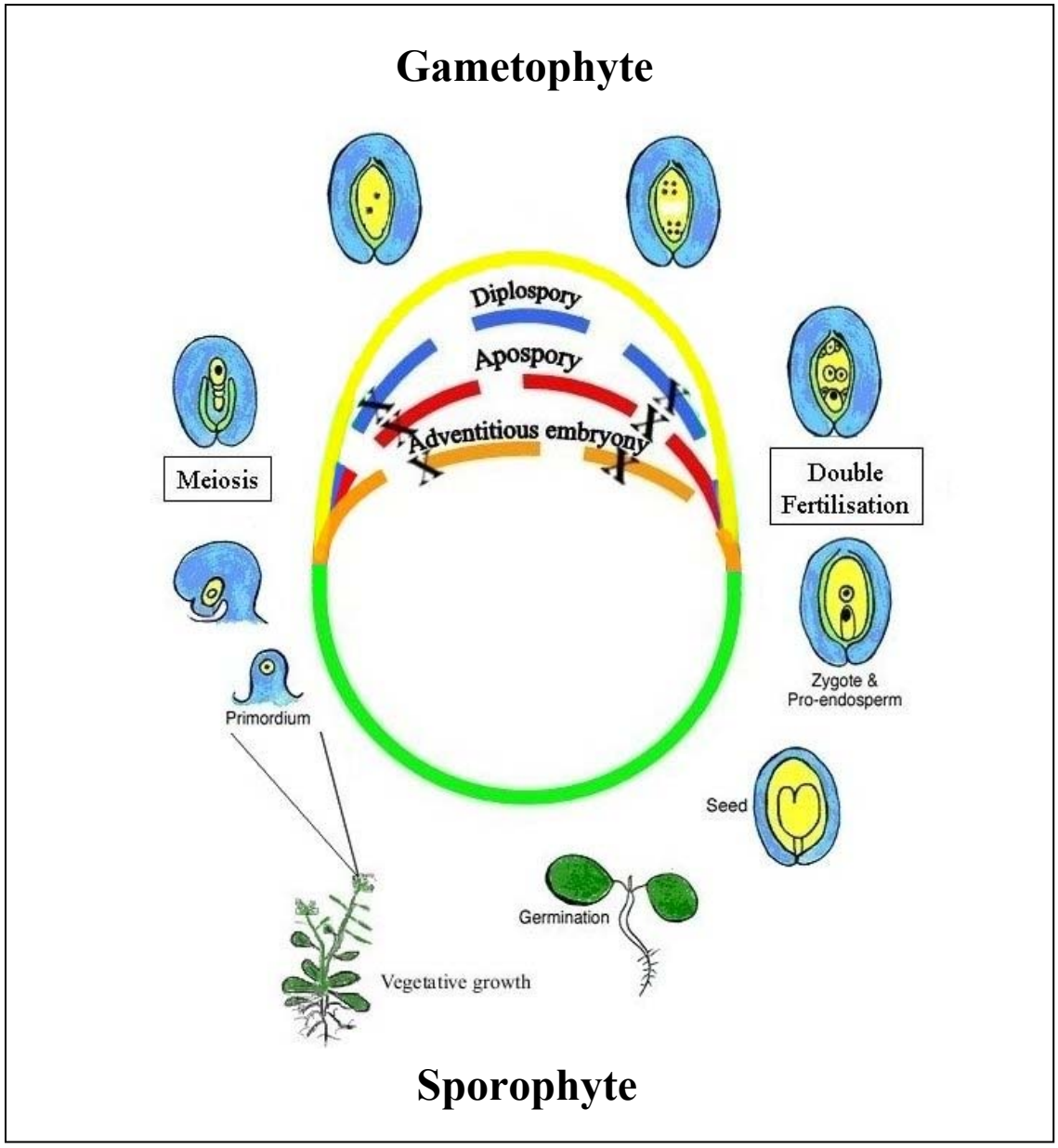


Fig. 1-1

Sexual and apomictic life cycle of plants (modified from Bicknell and Koltunow, 2004) and schematic representation of the female gametopyte development. The continuous green line represents the sporophytic phase, while the gametophytic phase is indicated by a yellow line. Diplospory (blue) and apospory (red) are called gametophytic apomixis. In both cases unreduced embryo sac is formed. Adventitious embryony (orange) is termed sporophytic apomixis because embryos are formed directly from the nucellus or integument cells. X show avoidance either meiosis or fertilisation in different apomictic pathways.

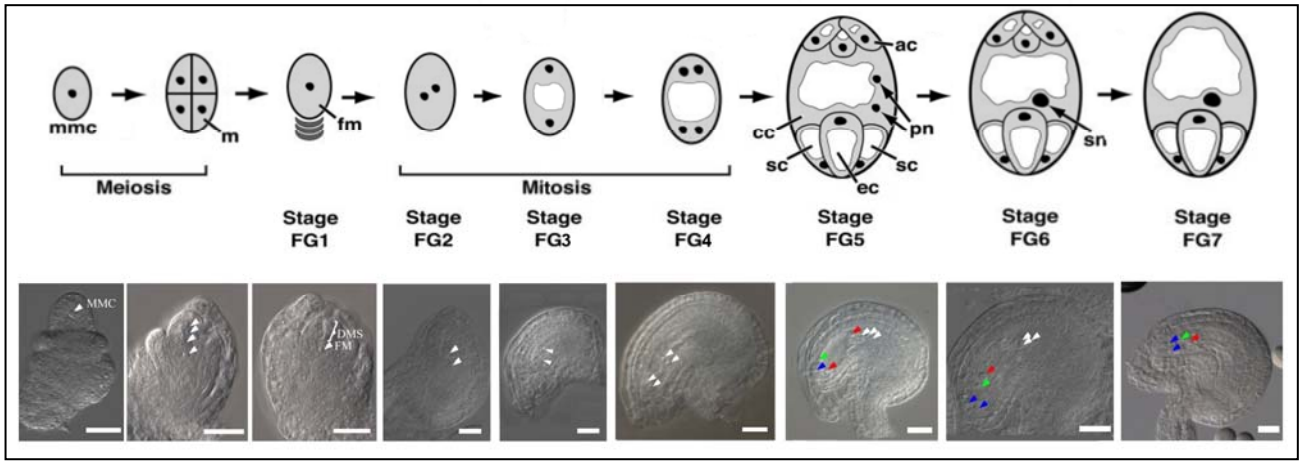


Fig. 1-2

The development of the female gametophyte in *Arabidopsis thaliana*. Upper panel schematic representation (Yadegari and Drews, 2004). Abbreviations m: megaspore pn: polar nuclei, cc: central cell, sc: synergid cell, ec: egg cell, ac: antipodal cell, sn: fused central cell nucleus. Lower panel cleared ovules in stages according to the upper panel. Abbreviations: MMC: megaspore mother cell, DMS degenerating megaspores, FM functional megaspore. The arrowheads indicate the nuclei. From stage FG5: arrows indicating the nuclei. The colours are the following: green - egg cell, blue - synergid, red - central cell, white - antipodal cell. Note that in FG5 only one synergid is visible. The scale bar is 20μm.

1.3. Apomixis - avoidance of sexuality

In contrast to the described pathway of sexual plant reproduction, about 400 flowering plant taxa from more than 40 families reproduce by an alternative asexual pathway, known as apomixis (Mogie, 1992; Carman, 1997), which has been first described for *Alchornea ilicifolia* (Smith, 1841). Nowadays the process of apomixis is synonymous with the term “agamospermous”, and includes all the asexual reproductive processes in the ovule of flowering plants, actually the place where normally sexual processes take place.

Apomictic reproduction is the result of a complex developmental programme including three major components: a) the avoidance of meiotic reduction during embryo sac formation (apomeiosis), b) the fertilisation-independent initiation of embryogenesis (parthenogenesis) and c) the autonomous (fertilisation independent) or pseudogamous (fertilisation dependent) development of the endosperm (Bicknell and Koltunow, 2004; Koltunow, 1993). Conceptionally, it can be considered as triple chain of uncoupling developmental check points: gametophyte formation is uncoupled from meiosis, the initiation of embryogenesis is uncoupled from fertilisation and also endosperm formation is uncoupled from fertilisation at least in species with autonomous endosperm formation. Natural apomicts can be divided into two groups which differ in their

requirement of fertilisation for endosperm development. Whereas apomicts with autonomous endosperm formation (e. g. *Hieracium*, *Taraxacum*) do not require the fertilisation of the central cell, endosperm development depends on fertilisation of the central cell in pseudogamous apomicts (e. g. *Poa pratensis*, *Hypericum*) (Koltunow and Grossniklaus, 2003; Spillane et al., 2001). Normal endosperm development requires a delicate balance between the maternal and paternal genomes maintained by differential epigenetic imprinting (for review see Köhler and Grossniklaus, 2005). On theoretical basis the observations are explained by the parental genome conflict theory (Haig and Westoby, 1989). Distortion of this balance usually results in defective seeds. Obviously, apomicts adopt mechanisms to avoid or tolerate these imbalances between genomes.

At least three different pathways of apomictic seed formation have been described. When embryos develop directly from the nucellus or integuments without any embryo sac formation, the process is called adventitious embryony or sporophytic apomixis. In gametophytic apomixis unreduced embryo sac can be found. Two main subgroups can be distinguished based on the cell type from which the unreduced gametophyte develops. In diplosporous apomicts like *Boechera* the MMC or the cell in the same position gives rise to the unreduced gametophyte. In aposporous apomicts like in species of *Poa*, *Hypericum* and *Hieracium* one or more so called apospore initial cells (AI) develop into an unreduced embryo sac. Aposporous gametophytes can coexist in the same ovule with the reduced sexual gametophyte, although the latter often degenerates. Further categorisations are possible for both aposporous and diplosporous apomixis (Crane, 2001).

Apomixis can be considered as plant cloning *via* seeds leading to offsprings genetically identical to the mother plant. This clonal propagation promises several important advantages for plant production, especially for the fixation of hybrid effects, and is internationally considered as a most favourable goal of plant breeding and production with an economic impact comparable to that of the `Green revolution` (Vielle-Calzada et al., 1996; Spillane et al., 2004).

1.4. Genetic analysis of plant reproduction

1.4.1. The sexual pathway

The molecular processes which are involved in the sexual and asexual reproduction of plants are poorly understood. At least two basic strategies and their combinations have been applied to study apomictic processes. A first approach tries to understand the often complex genetics of natural apomicts, whereas a second approach aims to synthesise an apomictic pathway based on detailed knowledge of the gametophytic and sporophytic development of sexual model plants like *Arabidopsis*. For this purpose forward and reverse genetic approaches have been applied

extensively focusing on sporophytic and gametophytic mutants. Sporophytic mutations affect the sporophytic steps of the megagametophyte development, including the selection of the MMC, meiosis and the influence of the sporophytic tissue on the gametophyte development like the establishment of correct polarisation. Several forward mutants have been identified and described in more detail (Chaudury et al., 1998; Gasser et al., 1998; Grossniklaus and Schneitz, 1998; Schneitz et al., 1998; Schneitz, 1999). Gametophytic mutations affect the development of the haploid embryo sac from the germination of the functional megaspore until the mature embryo sac as well as functions involved in the male and female gametophyte interaction like pollen tube guidance, double fertilisation and the maternal control of the seed development. Several mutants have been described (Christensen et al., 2002; Drews et al., 1998; Page and Grossniklaus, 2002; Pagnussat et al., 2005).

Functional characterisation has been performed for several, mainly *Arabidopsis* genes, which affect the development and/or function of the megagametophyte. These genes can be categorised according to the process they affect. Genes affecting nuclear proliferation are for instance *ARABINOGALACTAN PROTEIN 18* (*AGP18)* (Acosta-Garcia and Vielle-Calzada, 2004), *RETINOBLASTOMA-RELATED (RBR)* (Ebel at al., 2004), *INDERMINATE GAMETOPHYTE 1 (IG1)* (maize) (Evans, 2007), *LACHESIS* (*LIS*) (Groß-Hardt et al., 2007) *CHROMATIN-REMODELING FACTOR 11* (*CHR11)* (Huanca-Mamani et al., 2005), *NOMEGA* (Kwee and Sundaresan, 2003), *SUCCINATE DEHYDROGENASE (SDH)* (León et al., 2007), *SLOW WALKER 1* (*SWA1*) (Shi et al., 2005), *NOZZLE/SPOROCYTELESS* (*SPO*) (Schiefthaler et al., 1999, Yang et al., 1999) and *DYAD* (Siddiqi et al., 2000). Genes affecting cellularisation processes include *MATERNAL GAMETOPHYTIC BASELESS 1* (maize) (Gutierrez-Marcos et al., 2006), *CYTOKININ-INDEPENDENT 1* (*CKI1*) (Hejatko et al., 2003; Pischke et al., 2002), and *LYSOPHOSPHATIDYL ACYLTRANSFERASE* 2 (*LPAT2*) (Kim et al., 2005). Defects in the structure of the mature embryo sac have been described for *GFA2* (Christensen et al., 2002), *MYB98* (Kasahara et al., 2005), *GLUCOSE 6-PHOSPHATE/PHOSPHATE TRANSLOCATOR* (*GPT1*) (Niewiadomski et al., 2005), *AGAMOUS-LIKE 80* (*AGL80*) (Portereiko et al., 2006a), *RPL21M* (Portereiko et al., 2006b), and *DEMETER* (*DME*) (Choi et al., 2002; Gehring et al., 2006). Additional genes are involved in the maternal control of endosperm formation and seed development as for instance *MEDEA* (*MEA*) (Grossniklaus et al., 1998), *MULTICOPY SUPPRESSOR OF IRA 1* (*MSI1)* (Köhler et al., 2003), *FERTILIZATION-INDEPENDENT SEED* (*FIS2)* (Luo et al., 2000), *FERTILIZATION INDEPENDENT ENDOSPERM* (*FIE)* (Ohad et al., 1999) and *FWA* (Kinoshita et al., 2004) as well as pollen tube guidance like *FERONIA* (*FER*) (Huck et al., 2003; Escobar-Retsrepo et al., 2007), *EGG APPARATUS 1* (*EA1*) of maize (Márton et

al., 2005) and *SIRENE* (*SRN*) (Rotman et al., 2003). Recent results also demonstrate the expected involvement of microRNAs in flower and ovule development (Válóczi et al., 2006; Wu et al., 2007).

Besides these mutant approaches expression based strategies have been described to isolate gametophyte specific genes. These kind of experiments meet a series of difficulties mainly due to the fact that the female gametophyte consists of only a few cells deeply embedded in sporophytic tissue. Embryo sac specific genes have been identified in several studies based on comparative expression profiling in *Arabidopsis* (Ahn et al., 2007; Honys and Twell, 2003; Johnston et al., 2007; Steffen et al., 2007; Yu et al., 2005) and in tobacco (Ning et al., 2006). In other species, allowing the isolation of individual cells of the female gametophyte, cell specific cDNA libraries have been generated for example in maize (Cordts et al., 2001; Lê et al., 2005; Yang et al., 2006) and wheat (Kumlehn et al., 2001; Sprunck et al., 2005).

1.4.2. The apomictic pathway

The inheritance of gametophytic apomixis has first been reported to be associated with the transfer of either a single locus or a small number of loci in most of the systems studied to date. In aposporous grasses, like *Pennisetum* (Sherwood et al., 1994), *Panicum* (Savidan, 1983), and *Brachiaria* (Valle et al., 1994), apomixis is reported to be simply inherited, with the trait conferred by the transfer of a single dominant factor. Simple dominant inheritance has also been reported for apospory in the dicotyledonous genera *Ranunculus* (Nogler, 1984) and *Hieracium* (Bicknell et al., 2000). Among the diplosporous apomicts, independent inheritance of diplospory and parthenogenesis have been recorded in *Taraxacum* (van Dijk et al., 1999) and in *Erigeron* (Noyes, 2000; Noyes and Rieseberg, 2000), while Voigt and Burson (1983) have shown the simple dominant inheritance of diplospory in *Eragrostis curvula.* The inheritance of diplospory in *Tripsacum dactyloides* is reported to be simple and dominant (Leblanc et al., 1995). Genetic analyses of the mode of reproduction in *Poa pratensis* identified at least five loci with differences in gene expressivity and penetrance to be required for the control of apomixis (Matzk et al., 2005). There is evidence of segregation ratio distortion in some of these systems, often because the dominant factor(s) associated with apomixis also appears to confer gamete lethality, restricting its transfer to some gamete genotypes (Nogler, 1984; Grimanelli et al., 1998; Roche et al., 2001; Jessup et al., 2002).

1.5. The Salmon system of wheat

The parthenogenetic, fertilisation-independent initiation of embryogenesis from a non-reduced egg cell is one component of apomixis. The wheat Salmon system (Matzk, 1996) provides an excellent experimental model to study genes involved in the control of parthenogenesis (Fig. 1-3).

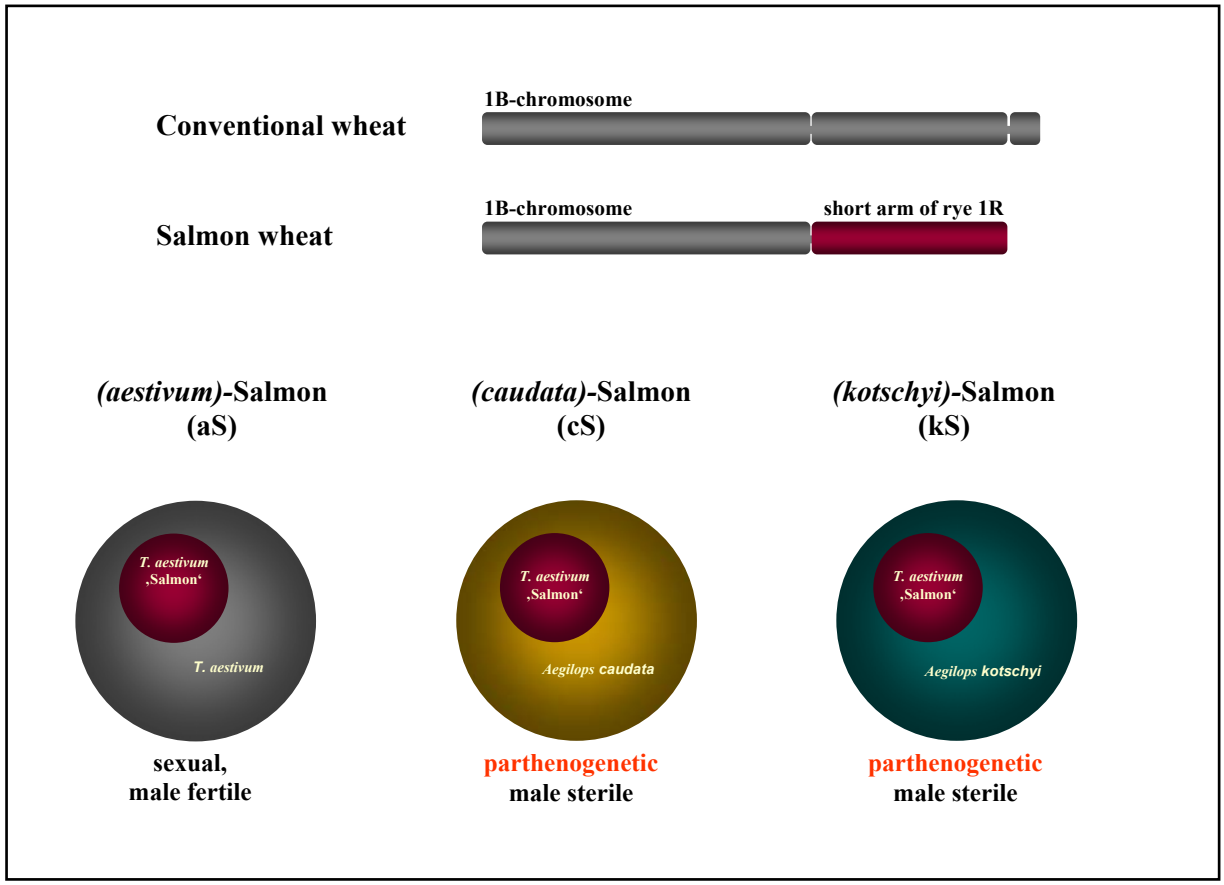


Fig. 1-3
Generation of isogenic wheat lines of the Salmon system (Matzk, 1996). After replacement of the short arm of wheat 1B chromosome with the short arm of rye 1R chromosome, the resulting *aestivum* line was crossed with two different *Aegilops* lines. These lines show parthenogenetic embryo development (*caudata*-Salmon, *kotschyi*-Salmon) and one line shows sexual reproduction (*aestivum*-Salmon).

In the Salmon wheat lines the short arm of chromosome 1B of wheat is replaced by the short arm of rye chromosome 1R. The transfer of this nucleus into different *Aegilops* cytoplasms results in male sterile plants with the capability of autonomous embryo development. Although embryo development is initiated autonomously from the reduced egg cell, seed development depends on fertilisation of the polar nuclei (pseudogamy). Three homozygous isogenic lines have been established after more than 20 backcrosses and the generation of doubled haploids: the

standard line (*aestivum*)-Salmon (aS) with normal sexual reproduction, and the alloplasmic lines (*caudata*)-Salmon (cS) and (*kotschyi*)-Salmon (kS) which exhibit autonomous embryo development with a frequency of about 90% (Matzk, 1996). It is an obvious advantage of the system that gene expression can be directly compared in two parthenogenetic lines (cS and kS) *versus* the isogenic sexual line (aS).

Remarkably, even the isolated egg cell of the parthenogenetic line can initiate embryogenesis autonomously (Kumlehn et al., 2001). This is in contrast to the egg cell of the sexual line and suggests that the fertilisation independent initiation of embryogenesis might be an inherent property of the egg cell itself (Fig. 1-4).

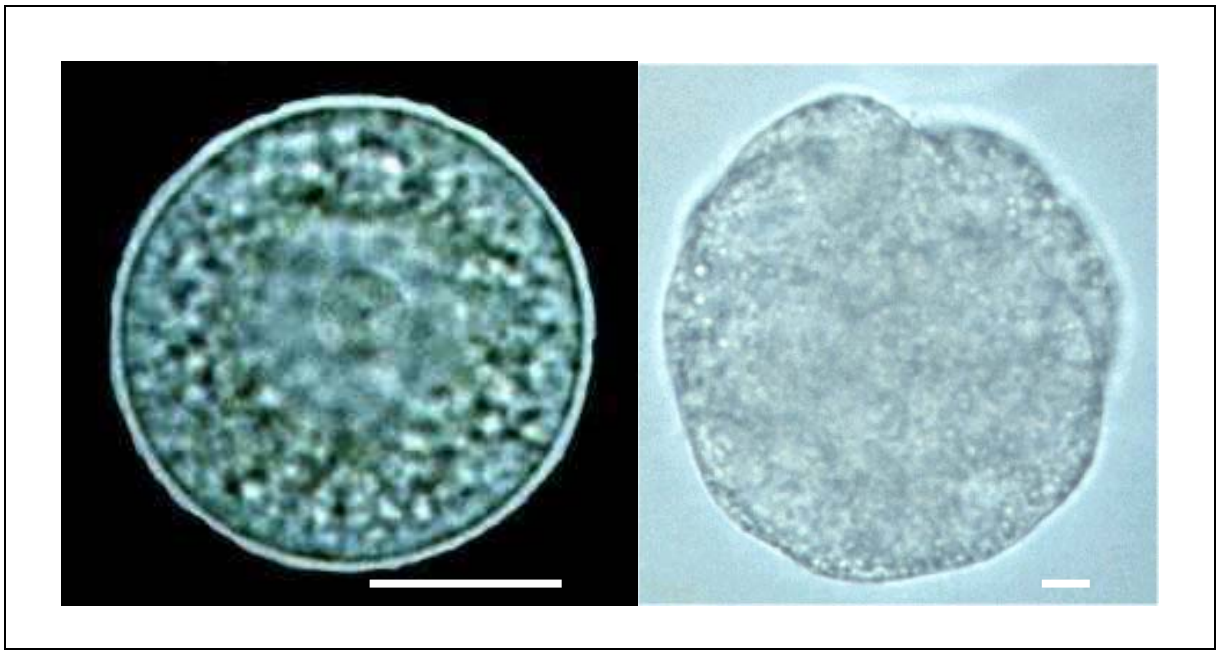


Fig. 1-4

A freshly isolated parthenogenetic (cS) egg cell (left) develops into a globular embryo (right) after *in vitro* culture for three days (Kumlehn et al., 2001). The bars represent 10 µm. (figure courtesy: J. Kumlehn)

1.6. Main objectives

To target the molecular differences between the parthenogenetic and sexual pathway, egg cell-specific cDNA-libraries from the sexual aS and the parthenogenetic cS line have been generated (Kumlehn et al., 2001). From the sexual cDNA library subtractive approaches have been used to remove chloroplast and mitochondria specific sequences as well as clones which represent ubiquitously and abundantly expressed genes. Remaining clones have been sequenced. For the identification of putative egg cell-specifically expressed genes, the egg cells are considered a small

component of floral tissues. Therefore, cDNAs of such genes are expected to be exceedingly rare. Thus, BLAST searches have been performed to identify sequences expressed in the egg cell, which are not yet present in available ESTs of monocotyledonous plants. These sequences are considered egg cell-specifically expressed genes.

One of these candidate genes defines a member of the *RWP-RK* gene family, named according to the characteristic amino acid sequence motif. The wheat egg cell-derived candidate gene shares sequence similarity with the *Chlamydomonas MID* gene (*MINUS DOMINANCE*). *MID* is located in the mating-type locus and determines the development of minus gametes during gametogenesis, induced by nitrogen limitation (Ferris and Goodenough, 1997). Moreover, members of the *RKD* family have been identified in a screen for female gametophyte-specifically expressed genes in *Arabidopsis* (Yu et al., 2005). RWP-RK factors are also required for the formation of infection threads and the initiation of symbiotic nitrogen-fixing root nodule on legumes (Schauser et al., 1999). These authors also suggest a further sub classification of the *RWP-RK* genes into the larger NIN and NIN-like proteins and the smaller RKD subfamily (Schauser et al., 2005).

The current work aims at the functional characterisation of the *RKD* gene family as a novel, strictly plant specific class of transcription factors, expressed in the female gametophyte of wheat and *Arabidopsis*.

2. *MATERIALS AND METHODS*

2.1. Plant material

Two plant species, *Arabidopsis thaliana* (ecotype Columbia-0) and *Triticum aestivum* L. (cv. aestivum Salmon and cv. Certo) were used in the present study. If not stated differently, plants were grown at 23°C with a cycle of 16 hours light and 8 hours dark.

2.2. Cloning methods

Standard molecular techniques (Sambrook et al., 1989) were used for cloning. The GATEWAY™ technology (Invitrogen, CA, USA) was performed according to the manufacturer's protocol. For the cloning work the DH5α strain of *Escherichia coli* was used, if not stated otherwise. The recombinant plasmids were isolated with QIAprep® Spin Miniprep Kit (Qiagen, Germany) and the fragments were isolated back from agarose gel by QIAquick® Gel Extraction Kit (Qiagen, Germany) according to the provided protocol. The oligos were ordered from Metabion AG, Germany, except for the primers with attachment sites, which were produced by Invitrogen, Germany.

2.3. *Arabidopsis* transformation

T-DNA constructs were introduced into the *Agrobacterium tumefaciens* strain GV2260 by freeze-thaw transformation (Chen et al., 1994). The floral dip method (Clough and Bent, 1998) was used for *Arabidopsis* transformation. Transformed progeny was selected by sowing surface sterilised T_0 seeds on 50 μg/mL antibiotic in MS basal medium (Murashige and Skoog, 1962) complemented with 1 % sucrose and 0,8 % agar (Sigma M-9274).

2.4. cDNA array hybridisation

mRNA was isolated from leaf material using Dynabeads from Dynal (Dynabeads m-RNA Direct™ kit (610.11)) according to the protocol provided by the manufacturer. The cDNA was synthetised on beads by RevertAid™ H Minus M-MuLV Reverse Transcriptase (MBI) at 42°C for 60 minutes. After elution, [^{33}P]dCTP was used for random prime labelling (Megaprime DNA Labelling Kit from Amersham). Prehybridisation was performed at 65°C for 2 hours and hybridisation was done at 65°C overnight. After hybridisation the membran was washed twice with 2x SSC/0.1% SDS, twice with 1x SSC/0.1% SDS, and once with 0.5x SSC/0.1% SDS at 65°C for 30 minutes each. The signal was detected by BAS-IP MS 2325 imaging plates from Fujifilm and

was read by a FLA 5100 phosphoimager. The data were analysed with Array Vision™ software (IMAGING Research INC, Brock University, Ontario, Canada) and an analysis module developed by Dr. Lothar Altschmied (IPK).

2.5. RT-PCR

RNA was isolated from different tissues using BIOMOL (kat. No. 51583) solution according to the protocol provided by the producer. 1 μg of RNA was used for cDNA synthesis after DnaseI (2,5 unit) (Roche, Germany) treatment, by RevertAid™ H Minus M-MuLV Reverse Transcriptase (MBI, Germany) at 42°C for 60 minutes. The PCR step was performed using the programme: 94°C, 2 minutes, 35 cycles, at 94°C for 30 seconds, 55°C for 30 seconds and 72°C for 60 seconds with a final elongation step at 72°C for 5 minutes, with Qiagen *Taq* polymerase (Qiagen, Germany). The primers were the following:

TaRKD forward 5'-GGTTTGAGGAGCTACGTACGTATTTCT-3'

TaRKD reverse 5'-AGAACCCAAAGAAACCCATCTGTGGCA-3'

AtRKD1 forward 5'-ATGAAATCGTTTTGCAAGTTGGAGTATGA-3'

AtRKD1 reverse 5'-TTATCTCTCAAAACCCGAAACAGAAGAGTA-3'

AtRKD2 forward 5'-ATGGCTGATCACACAACCAAAGAACAGA-3'

AtRKD2 reverse 5'-TCACAAACCACTAGTAAATTCACTTGAGA-3'

AtRKD3 forward 5'-ATGGCTGATCAAAGACCTCTAATGACCTGGT-3'

AtRKD3 reverse 5'-ACCAGGTCATTAGAGGTCTTTGATCAGCCAT-3'

AtRKD4 forward 5'-ATGAGTTCGTCAAAACATTCCTCTGT-3'

AtRKD4 reverse 5'-TCAATAATAATCATCACCAAGTGA-3'

AtRKD5 forward 5'-CCGAATCCGAGGAATCTGTA-3'

AtRKD5 reverse 5'-TGGCTCTAGAAGCCCTGTGT-3'.

In *Arabidopsis* the gene for the α subunit of elongation factor B was used as a positive control:

AtEFBα forward 5'-AGGAGAGGGAGGCTGCTAAG-3'

AtEFBα reverse 5'-AATCTTGTTGAAAGCGACAATGT-3'.

The wheat GAPDH gene was used as a control in the monocot experiments:

TaGAPDH forward 5'-AGGGTGGTGCCAAGAAGGTCA-3'

TaGAPDH reverse 5'-TATCCCCACTCGTTGTCGTA-3'.

2.6. Single cell RT-PCR

The single cell RT-PCR was performed as described by Richert et al. (1996). The following mixture was pipetted onto the isolated cell: 1X PCR buffer, dNTP-s, RNasin (20U) (MBI), reverse transcriptase (RevertAid^{TM} H Minus M-MuLV, MBI), reverse primer for the gene of interest as well for GAPDH as positive control. The same primers were used in these experiments like in the RT-PCR tests (see above).

The cDNA synthesis was carried out at 50°C for 70 minutes, followed by incubation for 10 minutes at 70°C. The PCR step was made in separate tubes for the gene of interest and for the GAPDH with Platinum *Taq* polymerase (Invitrogen, CA, USA) in 50 µl final volume. The following cycle conditions were used: 38 cycles, at 94°C for 45 seconds, 58°C for 30 seconds and 72°C for 90 seconds.

2.7. 5` RACE

Egg cells of *aestivum*-Salmon wheat were isolated according to Kumlehn et al. (1998). Messenger RNA was extracted from 20 egg cells *via* binding to oligo-dT magnetic beads, then it was reverse transcribed using SMART protocol (Zhu et al., 2001). The PCR reaction was carried out on the resulted cDNA, by an adaptor specific oligo at the 5' end (5'-GCTCTAATACGACTCACTATAGG-3') and a gene specific primer at the 3' end of the coding sequence (5'-AGAACCCAAAGAAACCCATCTGTGGCA-3') with Platinum *Taq* polymerase (Invitrogen, CA, USA) in 50 µl final volume. The following cycle conditions were used: 40 cycles, at 94°C for 90 seconds, 55°C for 30 seconds and 72°C for 120 seconds. The longest products were eluted from agarose gel with QIAquick® Gel Extraction Kit (Qiagen, Germany) according to the provided protocol. The fragments were cloned into pCR2.1 plasmid (Invitrogen, CA, USA) by overnight ligation at 14°C. Eight clones were randomly selected and sequenced. Assembled using the Sequencher 4.6 (Gene Codes Corporation MI, USA) software.

2.8. *In situ* hybridisation

Tissues were embedded in paraplast using the protocol of Kerk et al. (2003). Gene-specific fragments were cloned (see primers below) into pCRII-TOPO (Invitrogen, CA, USA) vector using the protocol of the manufacturer. These plasmids were used as templates for generating digoxygenin-UTP-labeled riboprobes by run-off transcription using T7 and SP6 RNA polymerases according to the manufacturer´s protocol (Roche Diagnostics, Switzerland). *In situ* hybridisation

was performed on 8-10 µm semi-thin paraffin sections as described by Vielle-Calzada et al. (1999). The following primer were used to create the probes for the experiments:

RKD1_F 5'-TGCAAGTTGGAGTATGATCAAG-3'

RKD1_R 5'-AGCCAGAAAGCTTGAGGCA-3'; Tm=52°C.

RKD4_F 5'-TTGAAACCAAATGCAGGGTTG-3'

RKD4_R 5'-CATCTCCAATTTGTCTTGCCTC-3'; Tm=52°C.

At5g40260_F 5'-TGGTGTTGGGTTAGTTATCGA-3'

At5g40260_R 5'-TTCTCCTAAACCCTCTCCGTA-3'; Tm=58°C

At4g30590_F 5'-CCTTGTCTTCCTCTTGTTTGC-3'

At4g30590_R 5'-CCAAACCAACCACGACTGC-3'; Tm=52°C.

2.9. Promoter-*GUS* fusion construction and GUS staining

1303bp long upstream fragment of the *AtRKD1* start codon was cloned into the pMDC163 vector (Curtis and Grossniklaus, 2003) using the following primers with Gateway attachment sites *att*B1 and *att*B2, forward and reverse, respectively:

AtRKD1 forward

5'-*GGGGACAAGTTTGTACAAAAAAGCAGGCT*AGCAAAAGTGGATTATGACGAC-3'

AtRKD1 reverse

5'-*GGGGACCACTTTGTACAAGAAAGCTGGGT*CAACGGGAGAGAGAGAACG-3'.

522 bp, 1315 bp and 436 bp upstream fragments of the start codon of the genes *AtRKD2*, *AtRKD3* and *AtRKD4*, respectively, were cloned into the pBIN19 vector (Bevan, 1984), carrying an intron containing *uidA* gene. The following primers were used (adaptor sequences are given in italic letters, HindIII and BamHI restriction sites are underlined):

AtRKD2 forward 5'-*CCCAAGCTTGGG*TTGTTACTTCATTAATAACT-3'

AtRKD2 reverse 5'-*CGGGATCCCG*TGTAAAGAAAGTGAGAGAGA-3'

AtRKD3 forward 5'-*CCCAAGCTTGGG*ACTGAGAAGGACATGGTCGGT-3'

AtRKD3 reverse 5'-*CGGGATCCCG*TTCGATTGAAAGTGGATATAGAT-3'

AtRKD4 forward 5'-*CCCAAGCTTGGG*TTTGATTTGACTACGTTTTC-3'

AtRKD4 reverse 5'-*CGGGATCCCG*ATATAGTCTCTGTAAAAACA-3'.

For detection GUS activity, the harvested plant material was vacuum infiltrated and incubated overnight at 37°C in the GUS staining solution containing 50 mM Na phosphate buffer pH 7.0, 10 mM EDTA, 0.1% Triton X-100, 2 mM potassium ferrocyanide, 2 mM potassium ferricyanide, 100 µg/ml chloramphenicol and 1 mg/ml X-gluc. The plant material was cleared in 20% lactic acid and

20% glycerol for 15 minutes and was analysed with a light microscope (Axioplan, Zeiss, Germany).

2.10. Transient expression of AtRKD–GFP fusions in protoplasts

The coding sequences, from the start codon until the last codon (the stop codon was excluded) were amplified with the following primers as reverse oligos with *att*B2 sites, shown by italic letters:

AtRKD1 reverse

5'-*GGGGACCACTTTGTACAAGAAAGCTGGGT*TCTCTCAAAACCGAAACAGAAGAGTA-3'

AtRKD2 reverse

5'-*GGGGACCACTTTGTACAAGAAAGCTGGGT*CAAACCACTAGTAAATTCACTTGAGA-3'

AtRKD3 reverse

5'-*GGGGACCACTTTGTACAAGAAAGCTGGGT*TGAAGAAGATGAGAAGAGCCTT-3'

AtRKD4 reverse

5'-*GGGGACCACTTTGTACAAGAAAGCTGGGT*ATAATAATCATCACCAAGTGA-3'.

The forward primers were the same which were used in the RT-PCR experiments, with the *att*B1 sites (not shown). Amplicons were integrated into the Gateway destination vector, pMDC84 (Curtis and Grossniklaus, 2003), which was used for transient expression in tissue culture derived *Arabidopsis* protoplasts. The cell suspension was incubated for 18h in K3 medium at room temperature in dark (Reidt et al., 2000), containing 6-benzylaminopurine, 1-naphthaleneacetic acid and 2,4-dichlorophenoxyacetic acid, in concentration of 4,5 μM, 10 μM and 4.5 μM, respectively. The GFP signal was localised *in vivo* using a confocal laser-scanning microscope (Zeiss, Germany). The eGFP fluorophore was excited at 488 nm by an argon laser and detected between 505 nm and 520 nm.

2.11. Southern hybridisation

15 μg of isolated genomic DNA was digested overnight, extracted with phenol-chlorform, separated on 0,8% agarose gel overnight and blotted onto a Hybond-N^+ nylon membrane (GE Healthcare, NJ, USA). 498 bp long RsaI-ClaI fragment of *TaRKD2* was used as a probe. Labelling was made with a random priming kit (GE Healthcare, NJ, USA) using [α^{32}P]dCTP. Hybridisation was performed at 65°C overnight. The membrane was washed twice with 2x SSC/0.1% SDS, twice with 1x SSC/0.1% SDS, and once with 0.5x SSC/0.1% SDS at 65°C for 15 minutes each. The

signal was detected by BAS-IP MS 2325 imaging plate from Fujifilm and was read by FLA 5100 phosphoimager.

2.12. Wheat transformation

For the transformation of wheat the protocol published by Jones et al. (2005) was used with slight modifications. *Agrobacterium* cultures were grown in 10 ml CPY medium at 28°C with shaking at 200 rpm until optical density (OD) of 0,25 at 600 nm was reached. Immature wheat embryos were isolated under stereo microscope in sterile condition from approximately two weeks old surface sterilised seeds and placed on solid DM media (Altpeter et al., 1996) at 12°C or 24°C in the dark for 5 days. The *Agrobacterium* cell suspension was pipetted onto the immature embryos and left in the dark at room temperature for 1 hour. The bacteria were removed and the embryos were transferred on filter paper on solid co-cultivation media. This is an MS medium, containing 800 mg/l cystein and 0,5 mM acetosyringon. The co-cultivation was carried out in the dark for 3 days at 23°C. Then the embryos were transferred to a solid DM medium containing the antibiotic Timentin (150 mg/l) for 10 days in the dark. In three subsequent cultivation steps Timentin was used with the same concentration as mentioned before (150 mg/l). After selection for one week on solid DM medium containing 20 mg/l Hygromycin, two rounds of regeneration steps for two weeks each, were performed on solid K4N (Kumlehn et al., 2006) medium containing 25 mg/l Hygromycin at 16°C, under 8/16 hours light/dark cycle. Regenerated plantlets were transferred into small glass tubes with antibiotic free K4N media for about two weeks, depending on the growth of the plant. Thereafter the plants were potted into soil and vernalised at 8°C for 8 weeks, transferred into bigger pots and grown to maturity.

PCR was used to prove the presence of the *hpt* (hygromycin phosphotransferase) gene in the regenerant plants. The used programme was the following: 94°C, 2 minutes, 35 cycles, at 94°C for 30 seconds, 59°C for 40 seconds and 72°C for 60 seconds with a final elongation step at 72°C for 5 minutes, with Qiagen *Taq* polymerase (Qiagen, Germany). The primers were the following:

HYG F2: 5'-GATTGCGTCGCATCGACCCT-3'

HYG R1: 5'-TATGGATGCGATCGCTGCGG-3'

2.13. Hygromycin leaf assay

The method according to Wang and Waterhouse (1997) was applied for leaf test. Part of leaves were cut and surface sterilised with 70% ethanol. Then the plant material was placed onto a K4N medium (Kumlehn et al., 2006), containing 200 mg/l Hygromycin. The plates were incubated

at under 8/16 hours light/dark cycle for 2 weeks. If the plant genomic DNA contains the *hpt* (hygromycin phosphotransferase) gene, the positive leaf samples remain green after the incubation and show antibiotic resistance, while negative samples and the wild type control tissue turns yellow.

2.14. Scanning electron microscopy

Tissues of *Arabidopsis thaliana* were collected. After fixation with 2 % glutaraldehyde in 50 mM phosphate buffer, pH 7,0 for 2 hours the shoots were washed with buffer and dehydrated in ethanol series. Following, the samples were dried in Bal-Tec critical point dryer (Bal-Tec AG, Balzers, Switzerland). Dried specimens were attached onto carbon coated aluminium sample blocks and gold coated in Edwards S150B sputter coater (Edwards High Vacuum Inc., Crowley, West Sussex, UK). Probes were examined in Hitachi S4100 SEM (Hisco Europe, Ratingen, Germany) at 5 kV acceleration voltage. Digital recordings were made and saved as tif-files.

2.15. Semi-quantitative RT-PCR by GeXP system

RNA was isolated from seedling, pistil (2 days after emasculation), *AtRKD1* and *AtRKD2* proliferating tissue, using the PicoPure™ RNA isolation kit (ARCTURUS, CA, USA) according to the protocol of the manufacturer. The cDNA synthesis and the PCR step was performed as stated in the protocol of the GenomeLab™ GeXP Start Kit, using 1 μg of isolated RNA. The samples were analysed using a capillary sequencer (CEQ 8000), after diluted 10 times with sample loading solution and adding one drop of mineral oil. The Frag-3 protocol was used during the electrophoresis. The data was analysed using the software provided with the GenomeLab GeXP Genetic Analysis System (Beckman Coulter, USA).

At1g02580 forward 5'-AGGTGACACTATAGAATAATGAAGCTAATGAGCGTGGG-3'

At1g02580 reverse 5'-GTACGACTCACTATAGGGATCATCAACTTGGCGTAGCAG-3'

At1g26795 forward 5'-AGGTGACACTATAGAATAAGGTTTCGTGTCAATCTCCG-3'

At1g26795 reverse 5'-GTACGACTCACTATAGGGAGAAAACACTGAGGTTGGCCT-3'

At2g20070 forward 5'-AGGTGACACTATAGAATATCAGAAGCAAGATCGAATGG-3'

At2g20070 reverse 5'-GTACGACTCACTATAGGGATGGAACATCAACAGGTCCAG-3'

At2g35670 forward 5'-AGGTGACACTATAGAATAACGAAAACAAAGGGTGATCG-3'

At2g35670 reverse 5'-GTACGACTCACTATAGGGAATTGTCCTGCTTCATCCGAG-3'

At3g26650 forward 5'-AGGTGACACTATAGAATAGACCAGAGGTTGCTAGACGC-3'

At3g26650 reverse 5'-GTACGACTCACTATAGGGATTGGTGTTGGTACACGGAGA-3'

At4g18770 forward 5'-AGGTGACACTATAGAATACAAGACAGGATTGTGCCTGA-3'

At4g18770 reverse 5'-GTACGACTCACTATAGGGACATCAATGTCAGGCTGAGGA-3'

At4g25530 forward 5'-AGGTGACACTATAGAATATATGCACCAGTGGAAACCAA-3'

At4g25530 reverse 5'-GTACGACTCACTATAGGGAGAGTGCAGCAGTTGGATTGA-3'

At5g40260 forward 5'-AGGTGACACTATAGAATATTACGGTCTCCCTGTGGTTC-3'

At5g40260 reverse 5'-GTACGACTCACTATAGGGATCCCAAAATGTTCCTTCGAT-3'

2.16. Transient expression of *AtRKD*::*GUS* fusions in protoplasts

The AtRKD1 and AtRKD2 promoters were cloned into pGUS1 plasmid (Peleman et al., 1989), after using the following primers for their amplification (adaptor sequences are given in italic letters, PstI, NcoI, HindIII and NcoI restriction sites are underlined, respectively):

AtRKD1p forward 5'-*AACTGCAG*AGCAAAAGTGGATTATGACGACA-3'

AtRKD1p reverse 5'-*CATGCCATGG*CATTATTAATGGGTTTGTGATCACT-3'

AtRKD2p forward 5'-*CCCAAGCTTGGG*TTGTTACTTCATTAATAACT-3'

AtRKD2p reverse 5'-*CATGCCATGG*TGTAAAGAAAGTGAGAGAGA-3'.

To transform the protoplasts, aliquots of 330 µl were heat-shocked (42°C for 5 minutes) before plasmid DNA (5 µg of each plasmid) and carrier DNA (160 µg of calf thymus DNA) were added. To bring the DNA into the protoplasts PEG 6000 (final concentration 20%) was added before transferring them to small Petri-dishes. After 48 hours incubation in the dark at room temperature, protoplasts were harvested, and the GUS activity was determined by a fluorimetric assay (Jefferson, 1987) using the GUS-LightTM Kit (Tropix, Bedford, USA). A control construct consisting of *35SCaMV* promoter in front of the *GUS* reporter gene was efficiently expressed in this system and used to standardize the different experiments.

2.17. Characterisation of the T-DNA lines

The *rkd1-1* (GABI 522C05) and *rkd2-2* alleles (GABI 237C07) were obtained from GABI-Kat T-DNA mutant collection (http://www.mpiz-koeln.mpg.de/GABI-Kat/GABI-Kat_homepage.html (Rosso et al., 2003). The *rkd1-3* allele (SALK 089683) was obtained from the SALK T-DNA insertion collection (http://signal.salk.edu). The *rkd4-1* (FLAG 301B07) and *rkd4-2* (FLAG 331F10) alleles were ordered from the FLAG collection (http://www-ijpb.versailles.inra.fr/en/sgap/equipes/variabilite/crg/index.htm (Brunaud et al., 2002)).To identify plants with the T-DNA insertion in *AtRKD1* and *AtRKD2* PCR analyses were performed. Allele-specific PCR reactions were performed to confirm the T-DNA insertion sites using primers for

GABI lines o8409 (5'-ATATTGACCATCATACTCATTGC-3'), for SALK line Rba3 (5'-CGGCTTGTCCCGCGTCATC-3') or LBb1 (5'-GCGTGGACCGCTTGCTGCAACT-3'). Gene specific primers were the following:

rkd1-1 forward 5'-GGAGATGGGTTCAAGATCAGGTAAAGCCA-3'

rkd1-1 reverse 5'-CGAATGGAAACAGTCTTGAGAAACA-3'

rkd1-2 forward 5'-CAAATACTAAGGAGTATATAATACTTAGGA-3'

rkd1-2 reverse 5'-GCTCTCTAGCCGCTTGAGTTATCGGCA-3'

rkd1-3 forward 5'-ATGAAATCGTTTTGCAAGTTGGAGTATGA-3'

rkd1-3 reverse 5'-TTATCTCTCAAAACCCGAAACAGAAGAGTA-3'

rkd2-2 forward 5'-ATGGCTGATCACACAACCAAAGAACAGA-3'

rkd2-2 reverse 5'-TCACAAACCACTAGTAAATTCACTTGAGA-3'

rkd3-2 forward 5'-AGAAGATGAGAAGAGCCTTCTCCG-3'

rkd3-2 reverse 5'-CAATCACTTTTCGTCGATCCTCACA-3'

2.18. Pollen viability test (Alexander staining)

The method, described by Alexander (1969) was used for pollen viability test. Anthers were dissected from mature flowers and incubated in a few drops of staining solution for at least 15 minutes. The solution consists of 10% ethanol, 0,01% malachite green, 0,05% fuchsin acid, 0,005% orange G, 5% phenol, 5% chloral hydrate, 2% acetic acid, 25% glycerol in water. Pictures were taken under a light microscope equipped with differential interference contrast optics (DIC) (Zeiss, Germany) using Axiocam HRC CCD camera (Zeiss, Germany).

2.19. Whole-mount preparation

Flowers and pistils were harvested at different developmental stages and fixed in ethanol-acetic acid (9:1) overnight at 4°C. Dehydration was performed in two steps, one hour each, in 80% and 70% ethanol. For clearing chloral hydrate:water:glycerol mixture (8:2:1) was used overnight at 4°C. Pistils were dissected and ovules were isolated. Samples were inspected using differential interference contrast optics (DIC) (Zeiss, Germany). Images were recorded with Axiocam HRC CCD camera (Zeiss, Germany).

2.20. Recombinant protein expression in *E. coli* (PAGE and Western blotting)

For the production of recombinant protein in *E. coli* (strain M15) the coding sequence (without the start codon) of *AtRKD1* and *AtRKD2* were amplified using the following primers (adaptor sequences are given in italic letters, BamHI and SalI restriction sites are underlined):

AtRKD1 forward 5'-*CGGGATCC*AAATCGTTTTGCAAGTTGGAGTATGA-3'

AtRKD1 reverse 5'-*ACGCGTCGAC*TTATCTCTCAAAACCCGAAACAG-3'

AtRKD2 forward 5'-*CGGGATCC*GCTGATCACACAACCAAAGAACAGA-3'

AtRKD2 reverse 5'-*ACGCGTCGAC*TCACAAACCACTAGTAAATTCAC-3'.

The fragment was then cloned using BamHI and SalI restriction enzymes, into the pQE30 vector (Qiagen, Germany). This vector contains a His-tag in the N-terminal part of the multiple cloning site (MCS). A colony was inoculated into liquid LB media containing 25 mg/l kanamycin and 50 mg/l ampicillin and was grown until optical density (OD) reached 0,6 at 600nm. After inducing the expression of the recombinant protein with 2 mM IPTG for 4 hours, the *E. coli* (strain M15) cells were harvested by centrifugation (5000 rpm for 5 minutes), resuspended in 1/10 of the starting volume in TE buffer (10 mM Tris, 1mM EDTA, pH 8) and disintegrated by ultrasonic (4 W for 100 seconds on ice) by VC60 machine (Sonics & Materials Inc., Danbury, Connecticut, USA). To separate the protein fragments 12,5% (w/v) SDS-polyacrylamide gels were used. Proteins were visualised by Coomassie blue staining or by Western blotting. For the latter, gel was blotted onto nitrocellulose membrane (PROTRAN® 0,45µm pore size from Schleicher & Schuell) in buffer, containing 0.25 M Tris, 0,1% SDS, 0,192 M glycine and 20% methanol. To prevent unspecific binding, the membrane was incubated in 3% BSA for 3 hours, followed by incubation with the monoclonal anti-polyhistidine (Sigma H1029) (1:3000) for one hour and three washes for 10 minutes each in 0,01 M Tris and 0,15 M NaCl (TBS) solution. The secondary antibody (anti-mouse Alkaline Phosphatase conjugate, Sigma A3562) was diluted 1:1000 and incubated for one hour. After three washes, each lasting for 10 minutes, with TBS the recombinant antibody was detected by a colorimetric reaction using NBT (ROTH 4421.3) / BCIP (ROTH 6368.2).

3. *RESULTS*

Despite the description of several gametophyte-specifically expressed genes, the molecular control of gametophyte development in higher plants is poorly understood until now. In order to identify egg cell-specifically expressed genes, cDNA library has been generated from isolated egg cells of Salmon wheat (Kumlehn et al., 2001).

3.1. The *RKD* gene family in wheat

3.1.1. Isolation of *RKD* genes from wheat egg cells

Cell isolation and construction of cDNA library from sexual Salmon wheat egg cells has been described previously (Kumlehn et al., 2001). The data, as well as unpublished further results obtained in cooperation with Dr. L. Altschmied, (IPK Gatersleben), can be summerised as follows. After quality trimming and vector clipping the sequencing of 682 randomly chosen clones from the cDNA library of unfertilised *aestivum*-Salmon egg cells yielded 1094 high quality sequences with an average length of 381 bases. Clustering these sequences with StackPack v2.2 led to 245 consensus sequences containing 802 ESTs and 292 singleton ESTs. Representative clones of the nine clusters with the largest number of clone members were chosen and their ^{33}P-labelled cDNA inserts were hybridised with 50,000 ordered clones on high-density colony membranes of the cDNA library. cDNA clones with significant hybridisation signal were excluded from the second phase of EST sequencing, similarly to that clones giving hybridisation signal with ^{33}P-labelled cDNA from green leaves of wheat. During this second sequencing phase, 1297 high quality sequences from 1139 pre-selected cDNA clones were obtained with average length of 354 bases. For further analysis the 1094 sequences from randomly chosen cDNA clones were combined with the 1297 sequences from pre-selected cDNA clones and 734 ESTs described, by Sprunck et al. (2005), the only other publicly accessible source of egg cell ESTs from wheat. Clustering this dataset of 3125 ESTs with StackPack v2.2 led to 1203 unique sequences, comprised of 558 consensus sequences containing 2480 ESTs.

To identify novel, potentially egg cell-specific cDNA sequences in this dataset, 1203 unique sequences were compared *via* BLASTN (NCBI) with database, containing all wheat ESTs available in GenBank release 156 (Oct 15, 2006; 854,901 ESTs), from which the 734 egg cell ESTs of Sprunck et al. (2005) had been removed. 129 sequences did not identify any similar sequence in this database and for 92 additional sequences, the score of the best

database hit was below 100. These 221 sequences are comprised of 55 singleton ESTs from Sprunck et al. (2005) and 58 singletons from the research, described here. When compared with annotated cDNAs in the rice genome (TIGR v5) *via* BLASTN, 71 of the 221 novel cDNA sequences hit on a rice gene with a score above 100. 18 of these 71 sequences are singleton ESTs from the project described here, 13 are singleton ESTs, described by Sprunck et al. (2005) and 40 are consensus sequences, 24 of which are solely composed of ESTs from our project, 6 are composed solely of ESTs described by Sprunck et al. (2005) and 10 are composed of ESTs from both projects. 53 different rice genes are hit by these 71 sequences, indicating the existance of several rice corresponding genes in the hexaploid wheat genome, which are simultaneously transcribed in the egg cell, although it cannot be excluded, that StackPack clustering does not put all ESTs into a cluster, which belong to a certain gene.

3.1.2. Identification of *RWP-RK* genes

To prove that subtractive hybridisation of leaf cDNA combined with *in silico* comparisons to wheat ESTs and rice genes is suitable to identify candidates for egg cell-specifically expressed genes, we selected the consensus sequence with the largest number of clone members exclusively derived from pre-selected cDNA clones, for more detailed study of gene function and expression. This 462 bases consensus sequence encompasses 9 clone members and shows high similarity with the predicted rice gene Os01g37100 (BLASTN score: 188) and less significant similarities with the another predicted rice genes Os06g12360 (score: 66.6) and Os02g51090 (score: 58.6) at the nucleotide level. All three rice genes encode proteins containing an RWP-RK domain (Schauser et al. 1999, 2005), which is named after the occurrence of the amino acid sequence RWPXRK within a larger, highly conserved protein sequence motif. So far, members of this gene family have been implicated in nodule formation in *Lotus japonicus* (Schauser et al., 1999) and in gamete differentiation in *Chlamydomonas reinhardtii* (Ferris and Goodenough, 1997). In *Arabidopsis* and rice the protein family has been subdivided by Schauser et al. (2005) in the NIN-like (nodule inducing-like) proteins (NLPs) and the RKD proteins. These two subfamilies can easily be distinguished by the size of the encoded proteins and their RWPXRK sequences. X is serin in the NIN-like proteins, but histidin in the RKD subgroup. The selected wheat cDNA sequence encodes a protein belonging to the RKD subgroup, since it encodes histidin in the RWPXRK sequence motif. All three rice genes identified *via* BLASTN with the wheat cDNA, also encode RWPHRK motifs.

The isolated *TaRKD* clone from the cDNA library was truncated. In order to isolate the full length gene 5'-RACE was carried out. Messenger RNA, extracted from 20 wheat egg cells *via* binding to oligo-dT magnetic beads was reverse transcribed, using SMART protocol (Zhu et al. 2001). PCR reaction was carried out on the resulted cDNA, by an adaptor specific oligo at the 5' end and a gene specific primer at the 3' end of the coding sequence. The longest products were eluted from agarose gel and cloned into pCR2.1 plasmid (Invitrogen, CA, USA). Eight clones were randomly selected and sequenced. Assembling these sequences with Sequencher 4.6 (Gene Codes Corporation, MI, USA), yielded three different cDNAs represented by five, two and one clone, respectively. The cDNA represented by single clone is a splice variant of the cDNA, represented by five clones. The third cDNA clearly represents a different gene. All three assembled cDNAs contain a potential start codon and an open reading frame up to the position of the gene specific primer employed in 5'-RACE.

3.1.3. Structural characterisation of the isolated *aestivum*-Salmon wheat *RKD* genes

The full length amino acid sequence of the wheat RKD proteins, and the alignment of wheat and *Arabidopsis* proteins with the *Chlamydomonas* MINUS DOMINANCE (MID) protein as well as a dendrogram of protein similarities between plant and algae proteins are shown in Fig. 3-1. The MID protein is known to be involved in the minus gamete differentiation (Ferris and Goodenough, 1997). The wheat *RKD* genes encode two proteins, 374 and 367 amino acids in length, respectively. The characteristic RWP-RK domain is located 98 amino acids upstream from the stop codon. It is preceeded by a heptameric pattern of apolar amino acids like leucine, isoleucine and methionine. The lysine residue, which is also found to be conserved in the wheat proteins, is located 10 amino acids upstream of the RWP-RK motif, and it is known to be functionally essential for the MID protein.

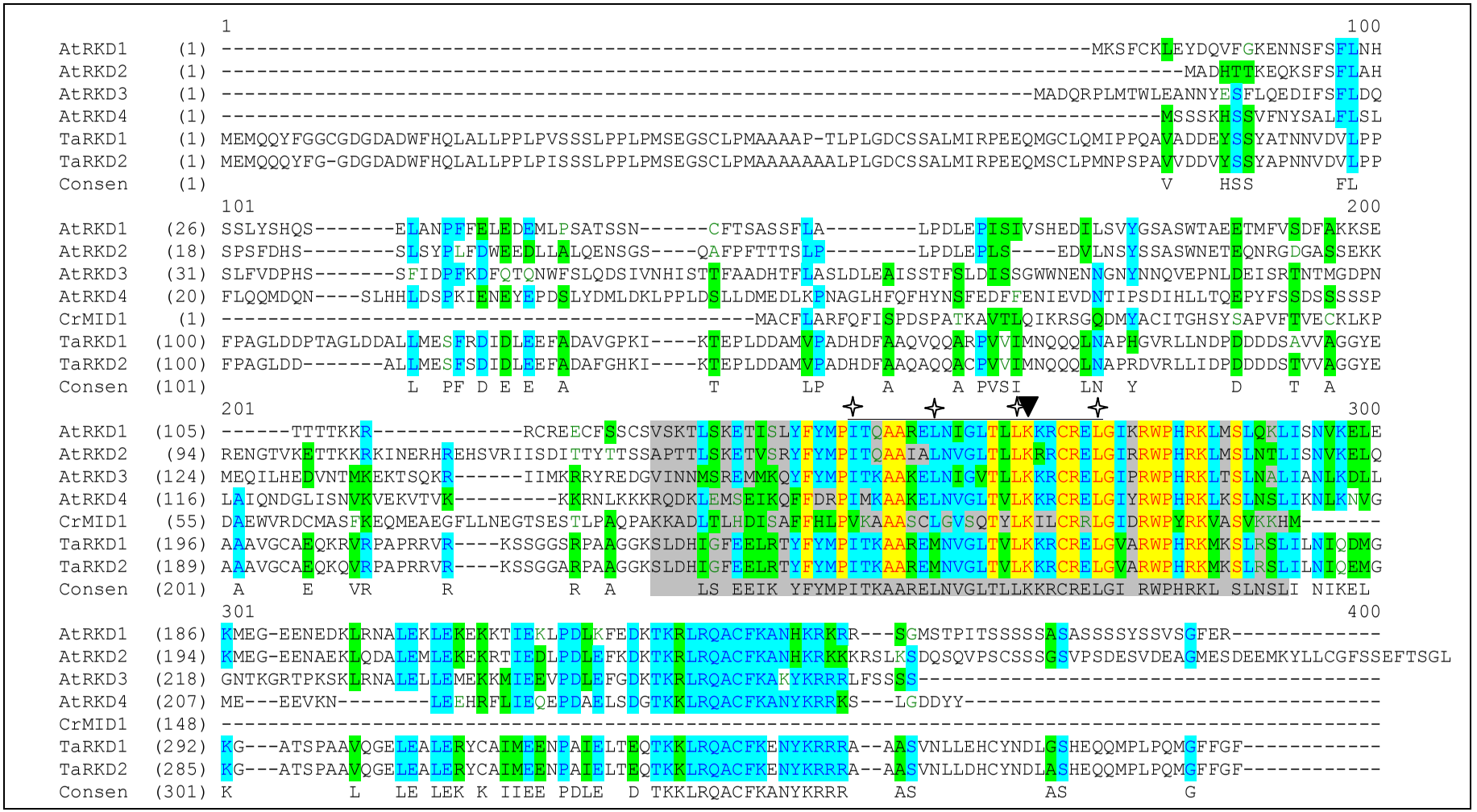


Fig. 3-14

Alignment of RKD proteins of *Arabidopsis* and wheat with the *Chlamydomonas* MID protein. Stars indicate the heptad repeats of large hydrophobic amino acid side chains. The arrowhead indicates the position of the K124 residue of MID, known to be essential for the normal function of the protein (Ferris and Goodenough, 1997). Amino acids, identical between RKDs and MID are shown with a yellow background, conservative amino acid positions are given in blue blocks. Similar residues are shown in green and weakly similar amino acids are shown by green letters. The consensus sequence includes amino acids conserved in two or more proteins. The RWP-RK domain is highlighted with a grey background. The aligment was obtained using the AlignX software (Invitrogen, CA, USA).

PCR amplification of genomic fragments from *aestivum*-Salmon wheat, resulted in four different fragments (Fig. 3-2), representing the two different expressed genes identified by 5'-RACE, and two additional ones. The exon-intron structure of *TaRKD1* and *TaRKD2* resembles to the *Arabidopsis* genes *AtRKD1-3* (for summarised structure of *Arabidopsis RKD* genes, see Fig. 3-14). The exon-intron structure of the two other genomic clones is different. One of them lacks the last exon, whereas the other one lacks all introns. Single nucleotide exchanges can be identified respect to *TaRKD1* and *TaRKD2*. Corresponding RACE products were not found, suggesting that these genes are not transcribed at detectable level in the egg cells of wheat. The Southern hybridisation with genomic DNA of *aestivum*-Salmon wheat reveals a small *RKD* gene family, with most likely one gene per genome in the hexaploid wheat (Fig. 3-3).

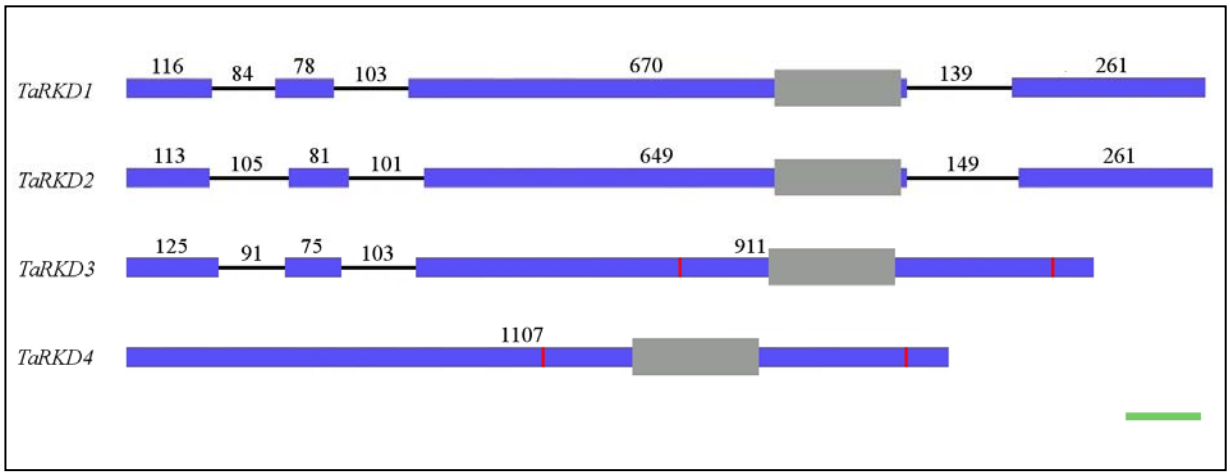


Fig. 3-2

Schematic representation of the structures of genomic *TaRKD* genes of *aestivum*-Salmon wheat. Blue boxes represent the coding region, black lines the introns. The grey boxes are the RWP-RK domains. The numbers indicate the lengths of coding regions and introns in bp. The vertical red lines indicate point mutations at the nucleotide level compared to *TaRKD1* and *TaRKD2*. The green scale bar represents 100 bp.

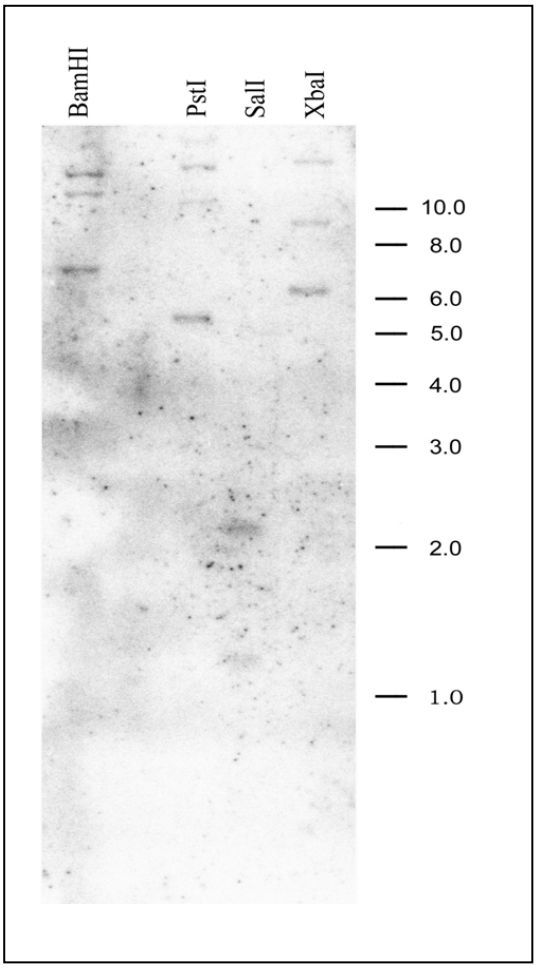


Fig. 3-3

Southern blot analysis of the *aestivum*-Salmon *TaRKD* gene family. Genomic DNA was isolated and digested with BamHI, PstI, SalI, XbaI restriction enzymes. DNA fragments were separated overnight on 0.8% agarose gel, blotted to nylon membrane and hybridised with 498 bp RsaI-ClaI fragment of *TaRKD2*. The sizemarker is given in kbp.

3.1.4. The *TaRKD* genes are specifically expressed in the embryo sac of *aestivum*-Salmon wheat

The expression of the *RKD* genes was analysed by two types of RT-PCR. Standard RT-PCR has been performed on several type of the *aestivum*-Salmon wheat tissues. The expression of the *RKD* genes was not detected in root, seedling, young leaf, stem, lemma, immature carpel, mature carpel (2 days after emasculation), fertilised carpel (2 days after pollination) and anthers in different developmental stages (Fig. 3-4).

In wheat the embryo sac and the gametophytic cells are big enough for isolation by microdissection technique. The diameter of a wheat egg cell is approximately 70-80 μm, while in *Arabidopsis* the whole embryo sac is at same size. This feature of female

gametophytic cells of wheat provides the opportunity to confirm the egg cell specific expression of the wheat *RKD* genes. Single cell RT-PCR has been performed using isolated embryo sac cells (egg cell, synergid, central cell) as well as zygote and two- three- and four-celled proembryos. As shown in Fig. 3-4 the highest expression level was observed in the egg cell. Lower transcript levels were detected in synergids and in the zygote. No transcript could be detected in the central cell and the two-celled proembryo. The data clearly demonstrates the egg apparatus-specific expression of *TaRKD* in the mature embryo sac of wheat. The transcript, found in the zygote, most likely represents remnant transcripts persisting from the egg cell. The expression is rapidly downregulated after fertilisation and - besides the zygote - no expression is detectable in sporophytic tissue.

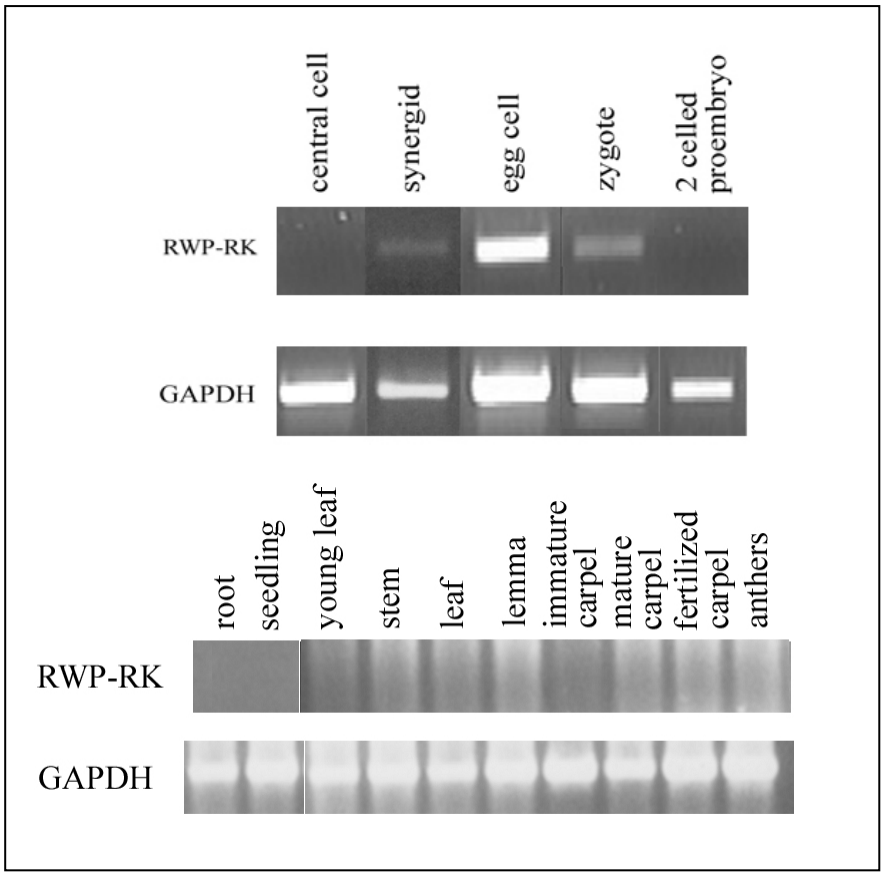


Fig. 3-4

RT-PCR analysis of *TaRKD* expression. *Upper panel:* Expression of *TaRKD* in isolated cells of *aestivum*-Salmon wheat ovules. *GAPDH* gene was used as a control. *Lower panel:* Expression of *TaRKD* in different tissues of the *aestivum*-Salmon wheat line. *GAPDH* gene was used as a control.

3.1.5. Functional characterisation of *TaRKD* genes

To elucidate the function of *TaRKD* genes in wheat, knock-down experiments were performed. Two RNAi constructs have been generated, containing either 242 bp or 253 bp fragment of *TaRKD2* gene. The first fragment includes the RWP-RK domain coding region (experiment WD19), whereas the second one contains the C-terminal part of the coding region (experiment WD20). Both RNAi cassettes are driven by the ubiquitin promoter. The schematic representation of the constructs are shown in Fig. 3-5. Both constructs were used to transfrom wild type wheat plants (cv `Certo`). After transformation 19 plants were obtained for the first construct and 13 plants for the second one. To identify the transgenic plants, PCR reaction was performed to detect the presence of the hygromycin resistance gene. In parallel, hygromycin resistance test was performed as shown in Fig. 3-6. In experiment WD19 three hygromycin sensitive plants were identified by leaf test, while PCR test indicated the lack of the resistance gene in one plant. In experiment WD20, two hygromycin sensitive plants were found by leaf test and three plants were tested negative for the resistance gene by PCR. Plants tested positive for hygromycin resistance assay as well as being confirmed to contain the resistance gene *via* PCR, were selected for further analysis.

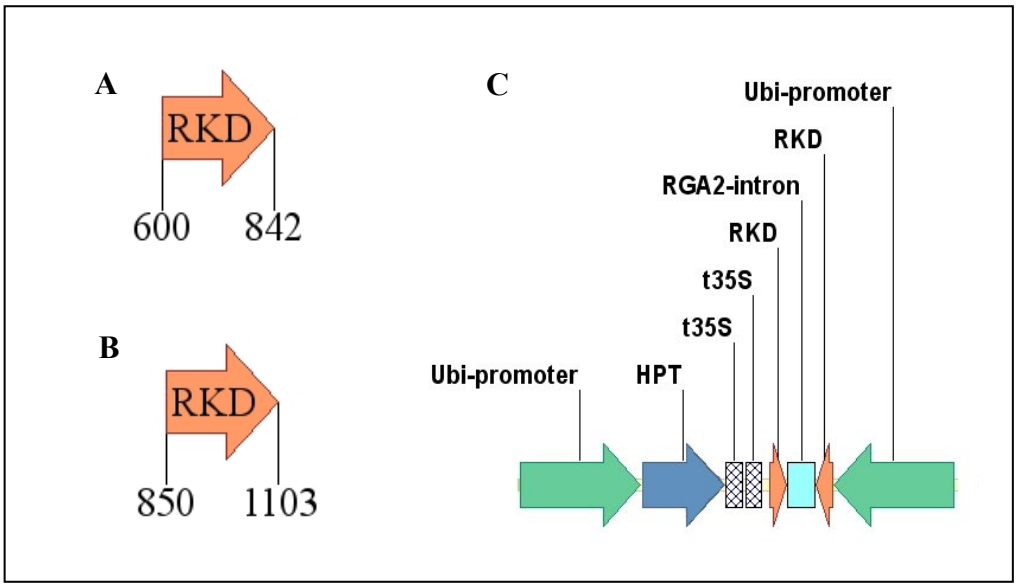


Fig. 3-5

Schematic representation of the constructs used for RNAi experiments. A and B: Arrows indicate the gene fragment used for construction of the RNAi silencing vector. Numbers indicate nucleotide positions with respect to the start codon of the *TaRKD2* mRNA. C: schematic diagram of the cassette used to silence *TaRKD* genes posttranscriptionally. HPT-hygromycin resistance gene, t35S - CaMV35S terminator sequence, RGA2-intron sequence to separate the sense and antisense fragments.

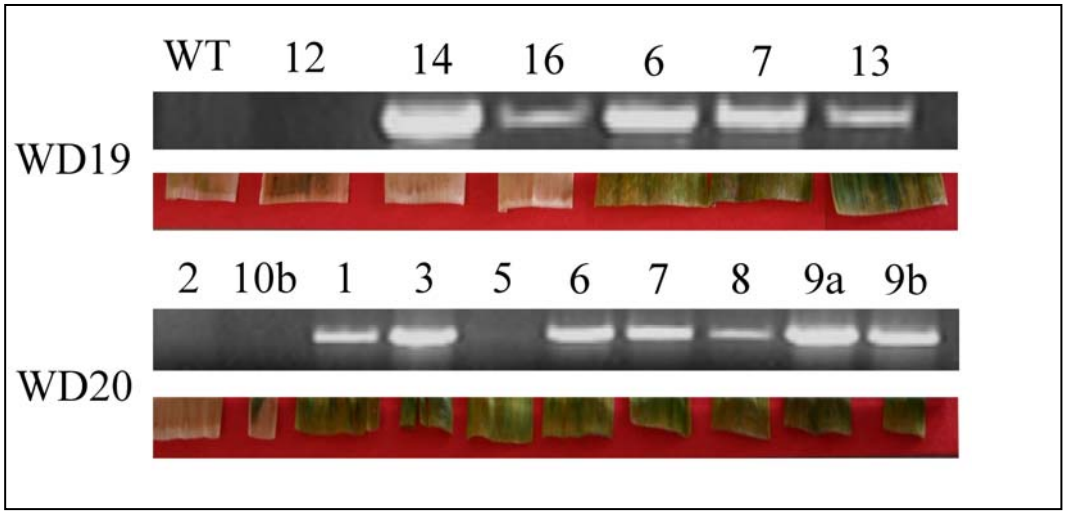


Fig. 3-6

Analysis of transgenic wheat plants in the RNAi experiment. *Upper lanes:* PCR to detect the hygromycin resistance gene using gene specific primers. *Lower lanes:* Hygromycin resistance test using young leaves on K4N medium contained 200 µg/ml hygromycin. Photos were taken of the leaves after two weeks.

If *TaRKD*s are essential for gamete development/function, as in *Chlamydomonas*, in a loss of function situation the female gametophyte development should be disturbed. Seed set was analysed in the first two spikes counting the number of seeds and the number of aborted florets. In wild type plants the first two spikes are the most fertile and the percentage of abortion is lowest. In experiment WD19, plants 12 and 16, which were negative in the hygromycin test, showed 9% and 14% of abortion rate, respectively. Unsuccessfully transformed and transgenic plant spikes can be seen in Fig. 3-7. The 6, 7 and 13 hygromycin resistance plants exhibit abortion rate 65%, 40%, 100%, respectively. In experiment WD20 the hygromycin sensitive plants 2 and 10b showed abortion rates 27% and 18%, respectively, while hygromycin resistant plants show abortion rates ranging from 30% to 96% (Table 3-1). An embryo rescue test was performed to analyse the functionality of the hygromycin resistance gene in the next generation. This test allows to follow the segregation of the hygromycin resistance marker carried by the T-DNA. Segregation and linkage between the seed abortion phenotype and the presence of the RNAi construct is currently tested. Furthermore, we perform histological analyses of ovule development.

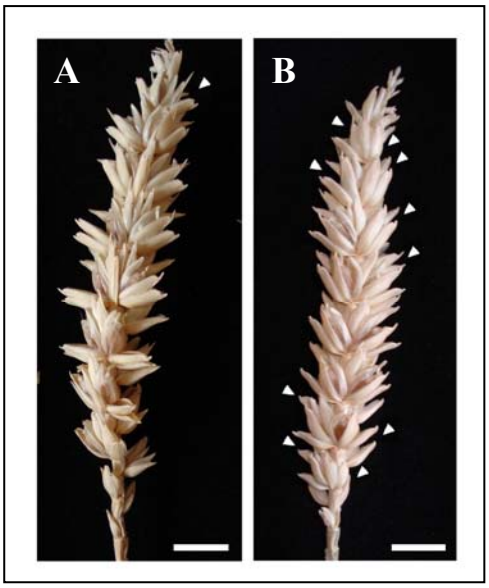


Fig. 3-7

Wheat spikes of control (unsuccesfully transformed) (A) and transgenic (B) (T_0) plants from experiment WD19. The arrows indicate non fertile florets. The scale bar represent 1 cm.

Experiment	Name	Spike 1		Spike 2		Abortion (%)	Hygro test	PCR (Hyg)	Construct
		Seed	Aborted	Seed	Aborted				
WD19	P1A	8	22	2	14	78	resistant	+	RWP-RK domain
WD19	P1B	25	4	17	6	19	resistant	+	
WD19	P2	27	10	2	22	53	?	+	
WD19	P3	22	6	19	9	27	?	+	
WD19	P4	16	6	19	6	26	?	+	
WD19	P5	18	6	14	7	29	?	+	
WD19	P6	14	14	5	21	65	resistant	+	
WD19	P7	14	12	23	13	40	resistant	+	
WD19	P8A	13	5	5	17	55	resistant	+	
WD19	P8B	28	6	30	3	13	?	+	
WD19	P9A	48	1	15	6	10	resistant	+	
WD19	P9B	22	10	20	7	29	resistant	+	
WD19	P10	10	11	12	12	51	?	+	
WD19	P11	27	5	43	10	18	?	+	
WD19	P12	31	1	23	4	9	sensitive	-	
WD19	P13	0	12	0	8	100	resistant	+	
WD19	P14	0	28	0	36	100	sensitive	+	
WD19	P15	25	2	29	3	9	?	+	
WD19	P16	19	4	26	3	14	sensitive	+	
WD20	P1	18	16	19	19	49	resistant	+	coding region of the last exon
WD20	P2	43	3	8	16	27	sensitive	-	
WD20	P3	20	6	22	12	30	resistant	+	
WD20	P4	2	25	0	28	96	?	+	
WD20	P5	20	10	18	10	35	resistant	+	
WD20	P6	8	21	4	20	77	resistant	+	
WD20	P7	7	17	0	22	85	resistant	+	
WD20	P8	6	20	2	18	83	resistant	+	
WD20	P9A	15	23	2	23	73	resistant	+	
WD20	P9B	15	22	9	28	68	resistant	+	
WD20	P10A	12	19	36	2	30	?	+	
WD20	P10B	25	10	34	3	18	sensitive	-	
WD20	P10C	6	16	20	8	48	?	+	
-	WT	34	4	39	3	9	sensitive	-	-

Table 3-1

Summary of the wheat RNAi experiments. On each plant the first two spikes were analysed for seed set. Test for hygromycin resistance was also performed using leaf material of the plants. The presence of the T-DNA was analysed *via* PCR using specific primers for the hygromycin resistance gene.

3.2. The *RKD* gene family of *Arabidopsis thaliana*

The *Arabidopsis* genome contains 14 *RWP-RK* genes (Schauser et al., 2005). The gene family consists of two distinct subfamilies including the NIN-like proteins, 767-1031 amino acids in length and the smaller RKD proteins with 256-370 residues (Fig. 3-8). The two subfamilies differ with respect to the position of the RWP-RK motif. All proteins of the NIN-like subfamily exhibit the RWPSRK motif, whereas the RKD proteins possess the characteristic RWPHRK motif. Furthermore, PB1 domain can be found only in the members

of the NIN-like subfamily. Interestingly, this domain is involved in cytosolic signalling and also found in the cell fate determining transmembrane receptor NOTCH of *Drosophila*. After ligand binding to an extracellular domain, the protein is cleaved and the intracelluar domain is free to act as a transcriptional co-regulator in the nucleus (Schroeter et al., 1998). The *RKD* subfamily consist of five genes in *Arabidopsis*. These genes are *AtRKD1* (At1g18790), *AtRKD2* (At1g74480), *AtRKD3* (At5g66990), *AtRKD4* (At5g53040) and *AtRKD5* (At4g35590).

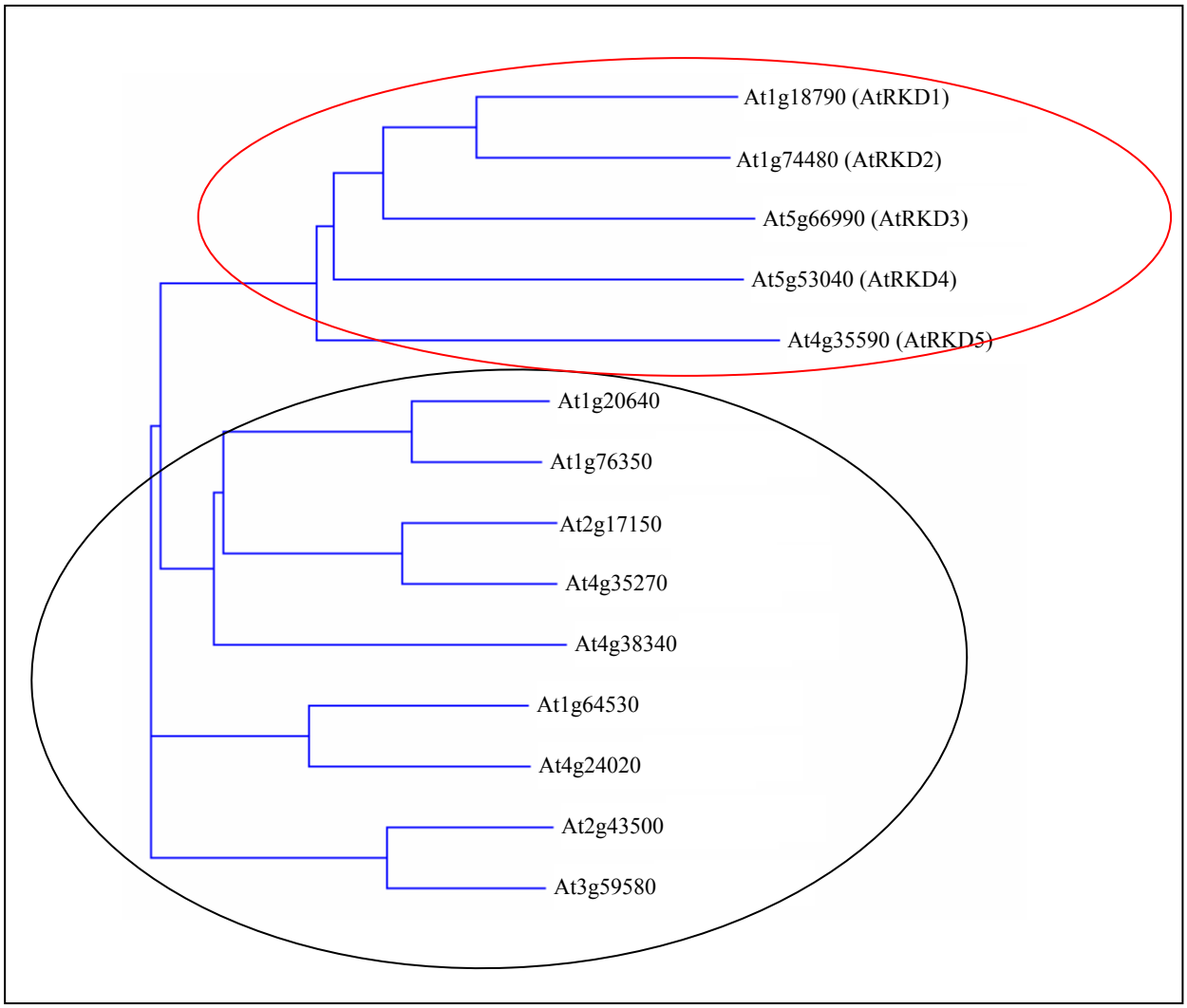


Fig. 3-8

The RWP-RK gene family of *Arabidopsis thaliana*. The red circle indicates the AtRKD subfamily and the black circle labels the NIN-like subfamily (NLP). The tree is based on amino acid sequences and was obtained using the AlignX software (a component of Vector NTI Advance 10.0.1, Invitrogen, CA, USA).

3.2.1. Flower bud specific expression of *Arabidopsis RKD* genes

To analyse the expression pattern of the *RKD* genes RT-PCR was applied using RNA isolated from different organs. As shown in Fig. 3-9 transcripts of *AtRKD1*, *AtRKD2*, *AtRKD3* and *AtRKD4* could only be detected in flower buds. No transcripts are detectable in open flowers, rosette leaf, stem, young siliques, roots and cauline leaf. These data well correlate with the expression pattern of *TaRKD*s in wheat. The more distantly related gene *AtRKD5* showed less specific expression pattern with more or less ubiquitious expression and has therefore been omitted from further analysis.

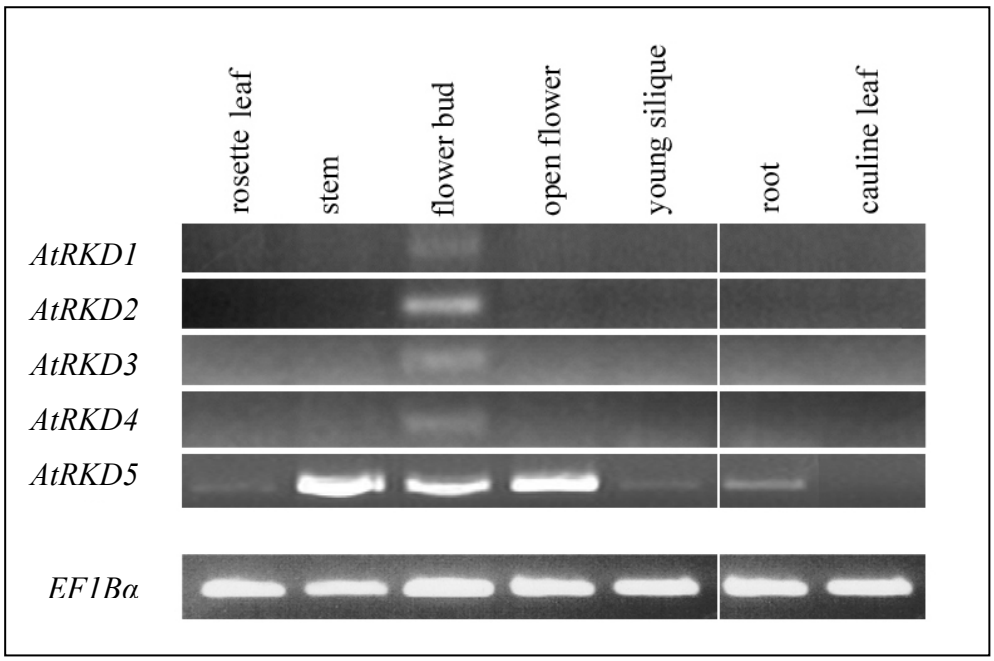


Fig. 3-9

RT-PCR analysis of *Arabidopsis RKD* genes in different organs. The constitutively expressed gene for the α subunit of elongation factor B (EF1Bα) was used as control. 35 cycles were performed using gene specific primers of At*RKD1, AtRKD2, AtRKD3, AtRKD4, AtRKD5* and *EF1Bα* on RNA isolated from different tissues.

3.2.2. Gametophyte specific expression of *AtRKD1* and *AtRKD4*

For more precise localisation of *AtRKD* gene transcripts at the cellular level, *in situ* hybridisation was performed in cooperation with Amal Johnston and Ueli Grossniklaus (University of Zürich, Switzerland). Hybridisation signal could be detected in the egg cell and the synergids of the mature embryo sac for *AtRKD1* and *AtRKD4* using the antisense probe. The specificity of the signal has been controlled by using sense probe in control experiment (Fig. 3-10). No hybridisation signal could be detected for *AtRKD3*, indicating very low transcript abundance. These experiments proved that *AtRKD1* and *AtRKD4* are specifically

expressed in the egg apparatus. These results support the hypothesis that *TaRKD1*, *TaRKD2* and *AtRKD1*, *AtRKD4* are orthologous genes with similar expression data.

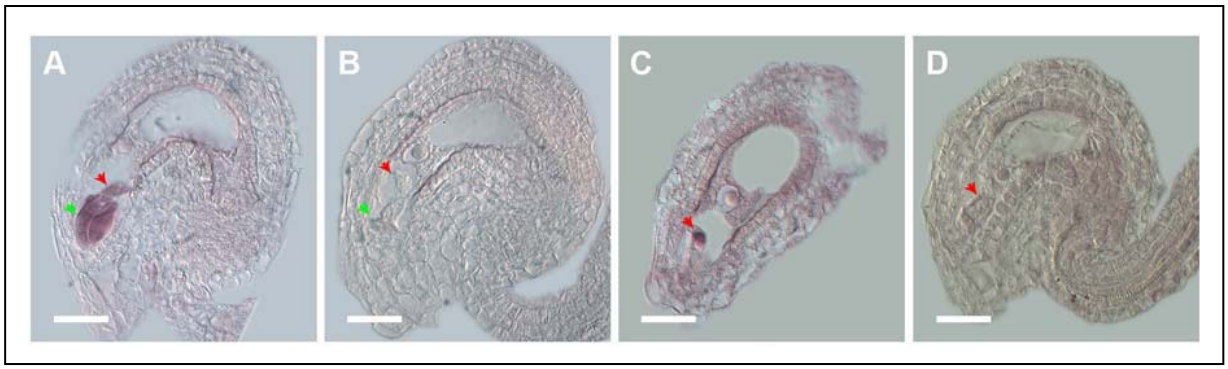


Fig. 3-10

In situ hybridisation of *AtRKD1* and *AtRKD4* in the mature embyro sac. A, B: *AtRKD1*, C, D: *AtRKD4*. Green arrows indicate the position of the synergids, red arrows label the egg cell. A, C: antisense, B, D: sense probe as control. The scale bars represent 20 µm. (figure courtesy: A. Johnston)

3.2.3. Egg apparatus-specific activity of *RKD* gene promoters

In order to support the *in situ* hybridisation results, promoter::*GUS* fusions were generated and *Arabidopsis* plants were transformed with. In the first series of experiments, the vector pMDC163 (Curtis and Grossniklaus, 2003) was used for *AtRKD1* and *AtRKD2* promoters (in cooperation with Amal Johnston and Ueli Grossniklaus, University of Zürich, Switzerland). The hygromycin resistance marker *(hpt)* of this vector is driven by the *CaMV35S* gene promoter. The *AtRKD2* promoter *GUS* construct was found to be active in nearly all plant organs like stem, rosette leaf, cauline leaf and flower. Obviosuly, this result did not correspond to the results of the RT-PCR and *in situ* hybridisation experiments. We assume that the strong constitutive *CaMV35S* enhancer can influence the activity of the *AtRKD2* promoter in this construct. Similar observation have been reported by Yoo et al. (2005). In the second series of experiments pBIN19 (Bevan, 1984) vector was used for *AtRKD2* promoter. The kanamycin resistance gene of this vector is driven by the *nopalin synthase* (*nos*) gene promoter. After transformation 13 plants were obtained. 6 plants out of the 13 showed embryo sac specific GUS activity. As it is shown in Fig. 3-11 the GUS activity was detectable from stage FG4 (Christensen et al., 1998.) on. In the mature embryo sac the signal was mainly restricted to the egg cell and the synergids. Moreover, 24 and 48 hours after pollination the staining was still found in the zygote and the first globular embryo stage. This might reflect the activity of the *AtRKD2* gene promoter also after fertilisation, however,

interpretation is favoured, that the post-fertilisation GUS activity is probably due to the high stability of the GUS protein (H. Bäumlein, A. Johnston, pers. comm.).

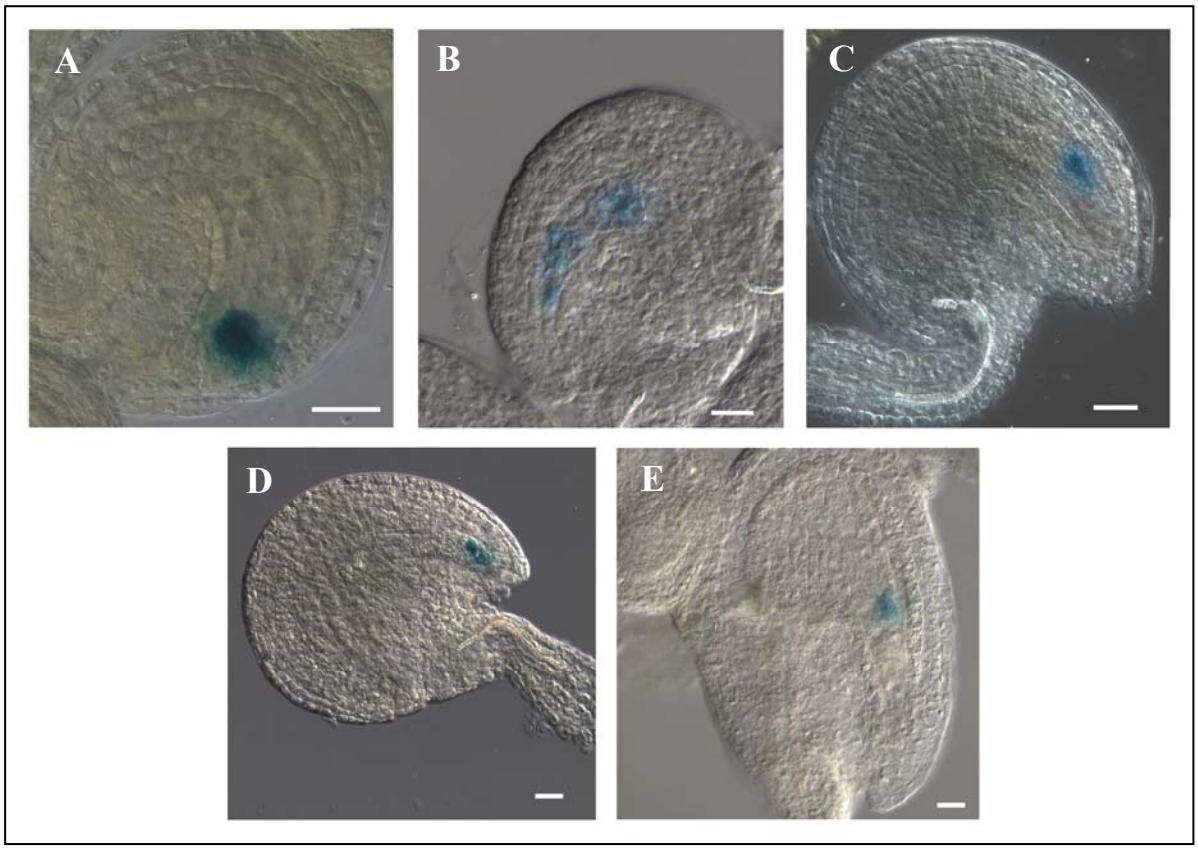


Fig. 3-11

Activity of *AtRKD1* and *AtRKD2* promoter *GUS* constructs in transgenic *Arabidopsis*. A: *AtRKD1*::*GUS* construct shows egg apparatus-specific activity in the mature embryo sac. B, C, D, E: Activity of an *AtRKD2*::*GUS* construct in transgenic *Arabidopsis*. B: GUS activity at the FG4 stage. C: the GUS signal is mainly restricted to the egg cell in the mature embryo sac, two days after emasculation. D: One day after pollination the promoter activity can be detected in the zygote. E: the GUS signal is detectable in early stage of the globular embryo, two days after pollination. The scale bar represent 20 μm.

3.2.4. Identification of a common *cis*-motif in the promoters of egg cell expressed genes

Both, *in situ* hybridisation and promoter GUS experiments demonstrate the embryo sac specific expression of *AtRKD* genes. This led to search for common *cis*-motifs in promoters of these genes, putatively involved in embryo sac-, egg apparatus- or egg cell-specific expression. With the help of the newly developed software CoMo-Finder (M. Mohr, I. Grosse, pers. comm.) a 9 bp *cis*-motif was identified in the promoters of all four *AtRKD* genes. The positions of this motif are 154, 197, 191 and 55 bp upstream of the start codon of

AtRKD1, *2*, *3*, *4*, respectively. The motif is not found in the promoter of the *AtRKD5* gene, which is not specifically expressed in the gametophyte (Fig. 3-9). A similar motif was identified in the promoters of other egg cell-specifically expressed genes. For example in the gene At2g21740 has been described as egg cell-specifically expressed gene by promoter::*GFP* fusion (Steffen et al., 2007), the motif is found 58 bp upstream of the start codon. Although the sequence motif needs extensive experimental confirmation, the data suggest the existence of a *cis*-element required for egg cell-specific gene expression. The functionality of the motif needs to be tested for instance by mutational analysis.

Name	Description	Strand	Position	P-value	Sequence
AtRKD1	RWP-RK domain containing protein	+	154	2.83e-06	CTCGTAACT
AtRKD2	RWP-RK domain containing protein	+	197	2.83e-06	CTCGTAACT
AtRKD3	RWP-RK domain containing protein	+	191	2.83e-06	CTCGTAACT
AtRKD4	RWP-RK domain containing protein	+	55	2.83e-06	CTCGTAACT
DD45 (At2g21740)	Expressed protein	+	54	2.83e-06	CTCGTAACT

Table 3-2

A supposed *cis*-motif in the promoters of egg cell-specifically expressed genes. Properties of the identified motifs. The numbers in the position column indicate the distance between the start codon and the last base pair of the element.

3.2.5. Nitrogen response of the *AtRKD1* and *AtRKD2* gene promoter

The *Chlamydomonas MID* gene is sequence homologue to the *RKD* genes. The gamete formation in *Chlamydomonas* is under control of the nitrogen level (Sager and Granick, 1954). Assuming phylogenetic conservation of this regulation, the *AtRKD* gene promoters activity were analysed at reduced nitrogen levels. The *AtRKD1* and *AtRKD2* gene promoters were cloned in front of the *GUS* reporter gene in the pGUS1 plasmid (Peleman et al., 1989). The constructs were transiently expressed in *Arabidopsis* protoplasts. The preparation of the cells and the transformation was done as described by Reidt et al. (2000). Empty pGUS1 plasmid (Peleman et al., 1989) was used as negative control and a *CaMV35S* promoter driven construct was used as positive control. Three different nitrogen levels were used: full 26 mM (24 mM NH_4^+, and 2 mM NO_3^-) nitrogen, 2 mM (2 mM NO_3^-) concentration and no nitrogen. The GUS activity was measured after 2 days incubation at room temperature in dark. Without any nitrogen supply the *AtRKD2* gene promoter was found to be ~37%

induced (Fig. 3-12) in comparison to the full nitrogen level. The induction rates of 4%, 8%, -7% have been measured for the *CaMV35S* promoter, the empty vector and the *AtRKD1* gene promoter, respectively. Up to now, the experiment was repeated twice. Current work focuses on the generation of transgenic *AtRKD1*::*GUS* and *AtRKD2*::*GUS* lines to further analyse the *AtRKD* gene promoter response on nitrogen supply.

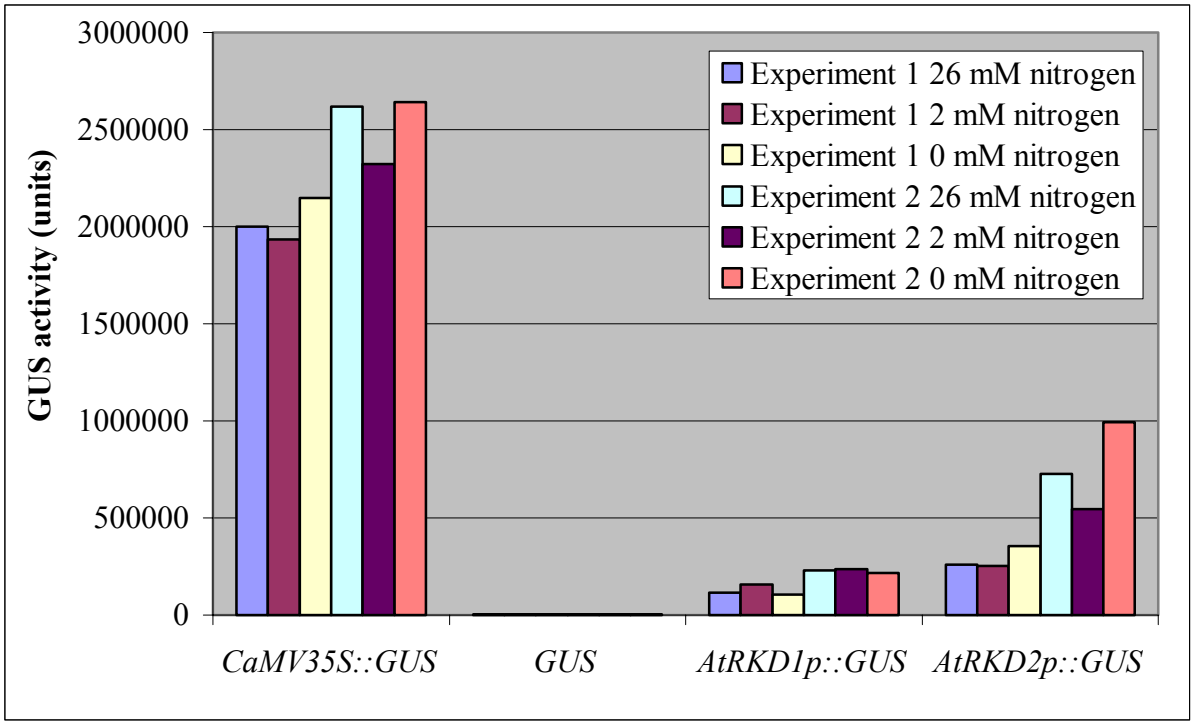


Fig. 3-12

Assay of promoter *GUS* constructs in *Arabidopsis* protoplasts under nitrogen depletion. The cells were transformed with different promoter *GUS* constructs: *CaMV35S* promoter, empty vector, *AtRKD1* promoter and *AtRKD2* promoter. The GUS activity was measured 2 days after transformation. A *CaMV35S*::*GUS* construct serves as a positive control and the empty GUS plasmid was used as negative control.

3.2.6. Nuclear localisation of RKD proteins

Although RKD proteins have been suggested to act as transcription factors, no experimental data are available to support this fact or to show their nuclear localisation. Therefore translational fusions were constructed of the coding regions of *AtRKD1*, *AtRKD2*, *AtRKD3* and *AtRKD4* with the *GREEN FLUORESCENCE PROTEIN* (*GFP*) gene in the pMDC84 plasmid (Curtis and Grossniklaus, 2003). In the fusion the coding region between the start and last codon was used, with the exculsion of the stop codon. The expression of the fusion products was driven by double *CaMV35S* promoter. The chimaeric genes were transformed into *Arabidopsis* protoplasts and transiently expressed. The unfused *GFP* gene

was used as a control. As shown in Fig. 3-13 all four fusion proteins can be localised in the nucleus of the cells, whereas the non fused *GFP* gene product is also detectable in the cytoplasm. These results demonstrate that all four AtRKD proteins are localised in the nucleus.

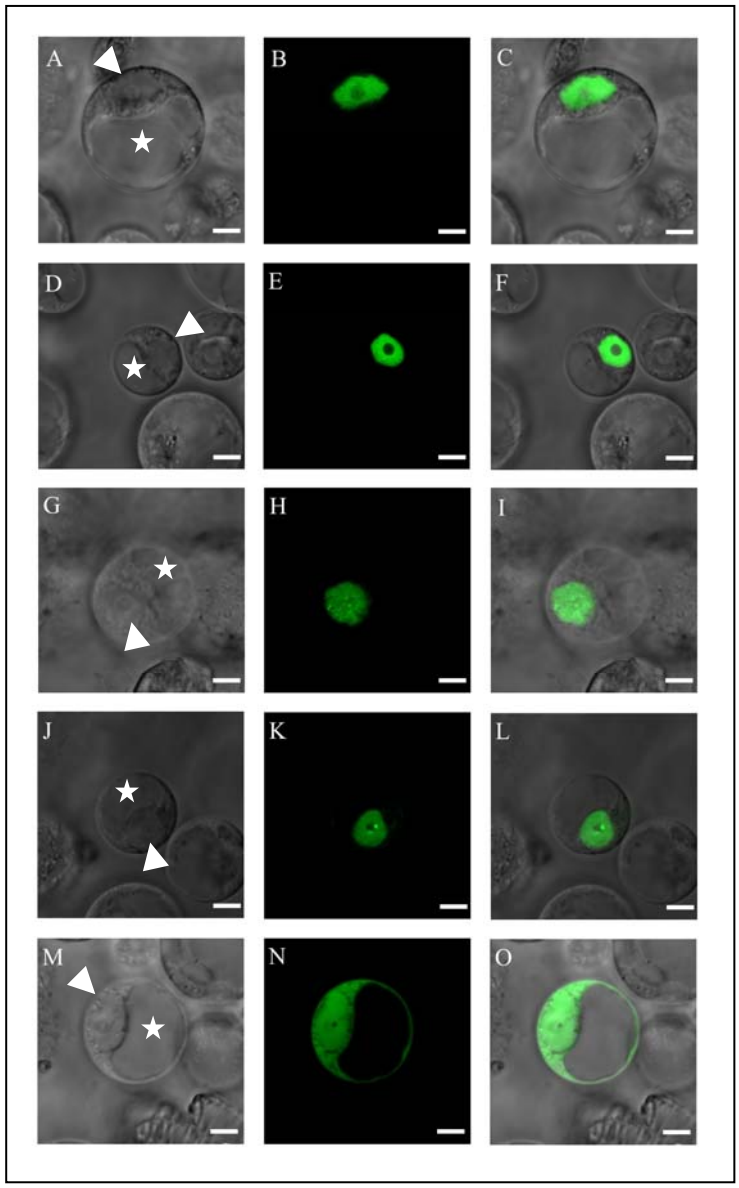


Fig. 3-13

Subcellular localisation of AtRKD-GFP fusion proteins expressed by double *CaMV35S* promoter in *Arabidopsis* protoplasts. Images were taken by laser scanning microscopy. White stars indicate vacuole in the protoplasts. White arrowheads show the nucleus of the protoplasts. A, B, C: AtRKD1-GFP. D, E, F: AtRKD2-GFP. G, H, I: AtRKD3-GFP. J, K, L: AtRKD4-GFP. M, N, O: GFP control. A, D, G, J, M: white light. B, E, H, K, N: UV-light (excitation/emission wavelength: 488 nm and between 505 nm and 520 nm, respectively). C, F, I, L, O: merged images. Scale bars represent 5 µm.

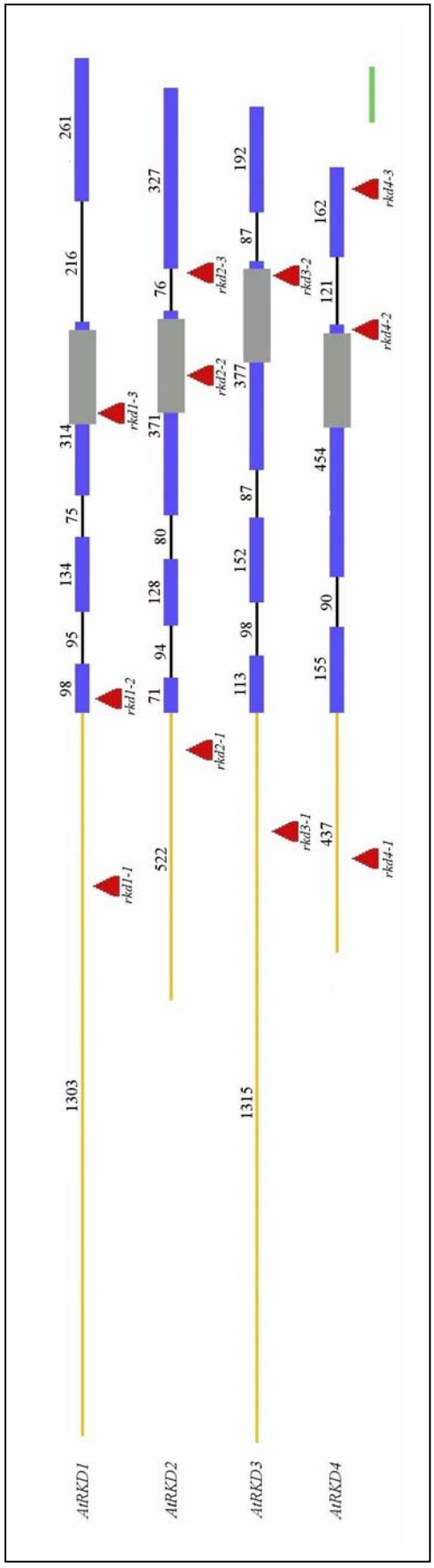
3.2.7. "Loss of function" mutant analysis

For a functional characterisation of the *AtRKD* gene subfamily eleven T-DNA insertion lines have been analysed, summarised in Table 3-3 and Fig. 3-14. Line carrying single T-DNA insertion, the offspring of a mother heterozygous for the insertion, should segregate for resistant/sensitive at 3:1. A 1:1 ratio suggests a defect in either male or female gametophyte. In this case, no homozygous plant for the insertion exist. Segregation ratio lower than 1:1 suggests the reduced viability of both gametophytes. None of the analysed single mutants showed detectable morphological difference during gametophyte development in comparison to wild type. In line SALK 133716 no T-DNA insertion has been found in the *AtRKD2* gene. Homozygous mutant plants have been identified for the following alleles: *rkd1-1*, *rkd1-2*, *rkd1-3, rkd2-2* and *rkd3-2* indicating that these mutant alleles are not lethal for the gametophyte, most likely due to a functional redundancy among the members of the *RKD* gene family.

Gene	Catalogue number/ T-DNA allele	Location of the T-DNA	Genotype	Phenotype
RKD1	GABI 522C05 (*rkd1-1*)	5′-UTR	homozygous	none
	SALK 061813 (*rkd1-2*)	coding region	homozygous	none
	SALK 089683 (*rkd1-3*)	coding region	homozygous	none
RKD2	SALK 133716 (*rkd2-1*)	5′-UTR	-	none
	GABI 237C07 (*rkd2-2*)	coding region	homozygous	none
	GABI 116G12 (*rkd2-3*)	intron	in progress	in progress
RKD3	SALK 142532 (*rkd3-1*)	5′-UTR	in progress	none
	SALK 020897 (*rkd3-2*)	coding region	homozygous	none
RKD4	FLAG 301B07 (*rkd4-1*)	5′-UTR	in progress	none
	FLAG 331F10 (*rkd4-2*)	coding region	in progress	none
	GABI 730H01 (*rkd4-3*)	coding region	in progress	in progress

Table 3-3

T-DNA mutant alleles. Mutant lines were identified from the SALK T-DNA Express database and seeds were received from the Nottingham *Arabidopsis* Stock Centre (NASC) (http://signal.salk.edu (Alonso et al., 2003)) from the GABI-Kat resource (http://www.mpiz-koeln.mpg.de/GABI-Kat/GABI-Kat_homepage.html (Rosso et al., 2003)) and from the FLAG collection (http://www-ijpb.versailles.inra.fr/en/sgap/equipes/variabilite/crg/index.htm (Brunaud et al., 2002)).


AtRKD1
AtRKD2
AtRKD3
AtRKD4
1303
1315
522
437
98
95
134
75
314
216
261
71
94
128
80
371
76
327
113
98
152
87
377
87
192
155
90
454
121
162
rkd1-1
rkd1-2
rkd1-3
rkd2-1
rkd2-2
rkd2-3
rkd3-1
rkd3-2
rkd4-1
rkd4-2
rkd4-3

Because the *AtRKD1* and *AtRKD2* genes are the most similar genes in the *AtRKD* subfamily they are probably functionally redundant. The goal was to create a double mutant containing T-DNA insertions in these two genes. In the *rkd1-1* allele a single T-DNA insertion was confirmed at position 303 bp in front of the translation start codon. In the *rkd2-2* allele the T-DNA insertion is present 621 bp downstream of the translation start. These mutants were crossed to create *rkd1-1*/*rkd2-2* double mutants. The siliques of F1 generation show 40% (n=536) abortion. The development is interrupted at the functional megaspore stage (Fig. 3-15). The pollen viability assay (Alexander, 1969) demonstrated, that the male gametophyte is not affected in the double mutant. F1 plants have been crossed reciprocally with wild type and the offspring was genotyped. Unexpectedly, homozygous double mutants could be found in the F2 generation. Currently a second double mutant *rkd1-3/rkd2-2* was analysed. The F1 plants again exhibit an abortion rate of about 45% (n=326). These observations might indicate that the observed phenotype is due to other effects like T-DNA rearrangements or growth conditions. Alternatively, the two remaining genes *AtRKD3* and *AtRKD4* might be able to complement the function *of AtRKD1* and *AtRKD2* due to functional redundancy. Therefore, T-DNA insertion mutants in *AtRKD3* and *AtRKD4* is currently being characterised.

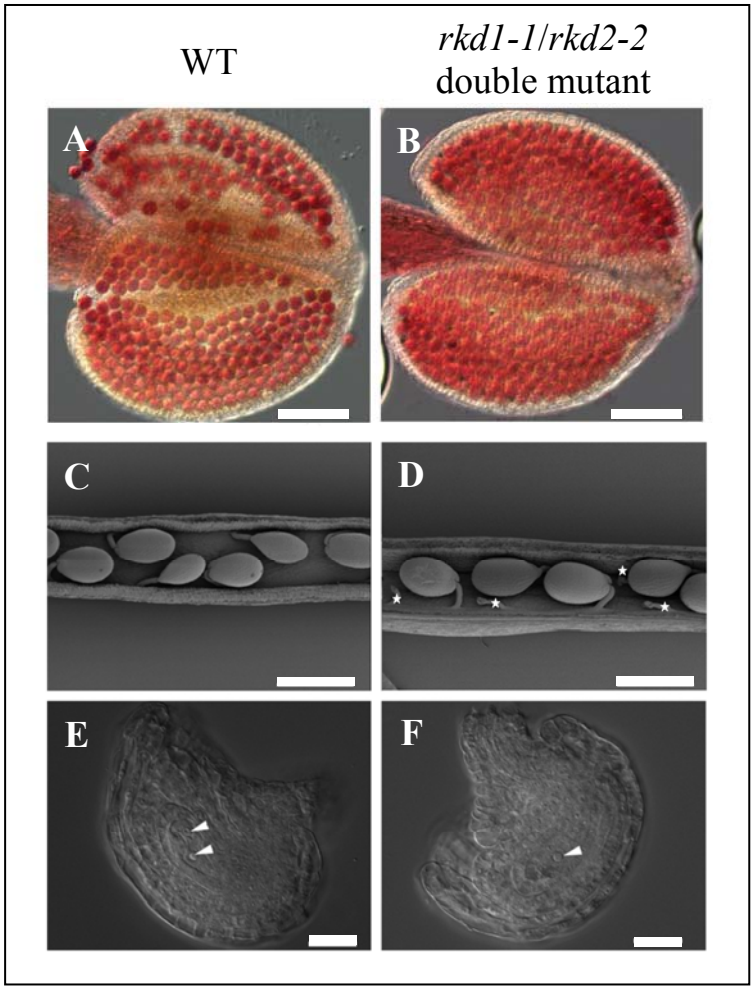


Fig. 3-15

Phenotype of *rkd1-1/rkd2-2* double mutant, compared to wild type (Columbia-0). A, B: Alexander staining of anthers. C, D: siliques with developing and aborted seeds. E, F: cleared ovules. Arrows indicating the nuclei in the developing embryo sac. The arrest at the megaspore stage means that in the gametophyte only one nucleus can be seen. The stars show aborted ovules in the mutant siliques. The scale bars represent 100 µm (A, B), 700 µm (C, D) and 20 µm (E, F), respectively.

3.2.8. Posttranscriptional silencing of *AtRKD* genes, using RNAi constructs

In addition to the T-DNA mutants, RNAi experiments have been performed, to have additional knock out alleles of *AtRKD1* and *AtRKD2*. Two different constructs have been used (Fig. 3-16). For optimal RNAi effect, 500 bp length fragment is thought to be sufficient. Since the *AtRKD2* gene is a rather short gene, with only 897 bp of coding region, two overlapping fragments have been applied. One construct includes the RWP-RK motif coding sequence

(Fig. 3-16, A) and the other one excludes it (Fig. 3-16, B), but still contains the N-teminal part of the RWP-RK domain. *In silico* analysis (http://bioinfo2.noble.org/) suggested the used constructs would downregulate both the *AtRKD1* and the *AtRKD2* gene.

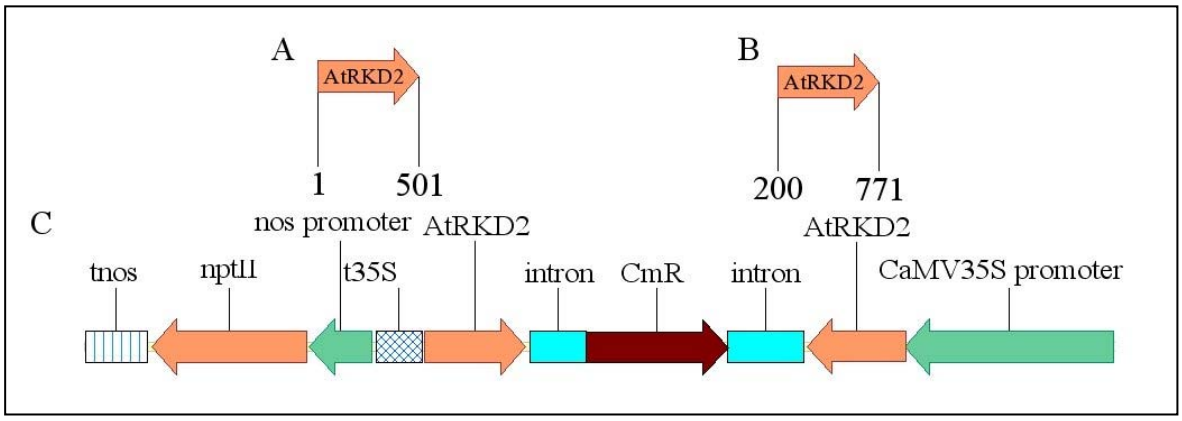


Fig. 3-16

Schematic structure of the RNAi cassette. A, B: the arrows indicate the sequences cloned in the RNAi silencing vector. Numbers indicate nucleotide positions with respect to start codon of the *AtRKD2* mRNA. C: schematic diagram of the cassette, used for posttranscriptional silencing of *AtRKD*s.

Whereas no stable transformants were obtained with the fragment, reaching from 1-501 bp, in total 24 transformants were obtained for the construct, containing the fragment 200-771. Different degrees of reduced seed set were observed in 8 primary transformants (Tables 3-4 and 3-5). The abortion is the consequence of the undeveloped gametophyte. The gametophyte development is found to be interrupted at the stage of the functional megaspore. For detailed analysis of the phenotype, pistils of different plants were investigated for female gametophyte development (Tables 3-4 and 3-5). In wild type *Arabidopsis* the development of the female gametophyte in one flower is synchronous. However, in one pistil the ovules can be found at different (2-5) stages (Table 3-4). In the mutant plants the FG1 stage was found to be more frequent than in the wild type. The analysis is currently repeated in the next generation.

Number of pistil	Number of female gametophytes at different developmental stages									
	FG1	FG2	FG3	FG4	FG5	FG6	FG7	FG8	Zyg	Σ
P1	4	20								24
P2	2	15	3	8						28
P3	1	6	6	21	5					39
P4			3	13	8	4				28
P5					1	4	6	3	2	16
P6					2	3	3	7	5	20
P7					3	1	11	8	5	28

Table 3-4

Female gametophyte development in *Arabidopsis* Columbia-0. The pistils were fixed as described in Materials and methods. Pistils of seven different plants were analysed. The female gametophyte (FG) stage for each ovule was determined and counted. The stages were determined, according to Christensen et al., (1998).

Number of pistil	Number of female gametophytes at different developmental stages										Abortion (%)
	FG1	FG2	FG3	FG4	FG5	FG6	FG7	FG8	Zyg	Σ	
P1A	18	3		4	4	1				30	60
P1B	26		1	1	4					32	81
P2A	27							9	1	37	73
P2B	25							8	2	35	71
P3A	13					5	10			28	46
P3B	20						17	6	1	44	45
P4A	8						8			16	50
P4B	6				3	4	3		11	26	23
P5A	11				1	1			20	33	33
P5B	18				1		4		12	35	51
P6A	8					2	9		7	26	31
P6B	10				1	1	12		12	36	28
P7A	20				5	1	4		1	31	65
P7B	17				4	2	6		3	32	53
P8A	12					5	7		4	28	43
P8B	14				2	4	8			28	50

Table 3-5

Female gametophyte development in *Arabidopsis* RNAi lines. The pistils were fixed as described in Materials and methods. Two pistils (A, B) of different mutant lines were analysed. The female gametophyte (FG) stage for each ovule was determined and counted. The stages were determined, according to Christensen et al., (1998).

3.2.9. Ectopic expression of *AtRKD1*, *AtRKD2*, *AtRKD3* and *AtRKD4* genes

For more detailed characterisation of the gene subfamily, "gain of function" experiments were carried out. The coding regions of all four genes have been cloned into the vector pMDC32 (Curtis and Grossniklaus, 2003). The expression is controlled by the double *CaMV35S* promoter. These plasmids were used to transform wild type *Arabidopsis* (Columbia-0) plants. The ectopic expression of *AtRKD1* and *AtRKD2* genes causes severe distortion in somatic growth (Fig. 3-17). Nevertheless, 24 less distorted plants could be raised out of about 35,000 seeds in case of *AtRKD1*. Using other constructs usually 80-90 plants can be obtained. The *AtRKD1* and *AtRKD2* overexpression plants are bushy with curly stems and mainly sterile flowers (Fig. 3-17 B, C). In case of *AtRKD2* only 2 less distorted plants were obtained, one of them did not express the transgene (data not shown), indicating, that the effect of *AtRKD2* ectopic overexpression might be stronger than that of *AtRKD1*. Three plants did not express the transgene. One of these plants exhibits a rather normal phenotype, shown in Fig. 3-17 B. For molecular analysis of *AtRKD1* misexpression effects, macroarray experiment was performed, using the plant with a rather weak phenotype (see chapter 3.2.10.). Some of the induced transcription factor genes have been found to be expressed in gametophytes. Triggered by this observation, quantitative RT-PCR (see chapter 3.2.11.) and *in situ* hybridisation experiments have been performed on white, proliferating part of the *AtRKD1* overexpressing tissue.

Similar growth distortions as described above for *AtRKD1* and *AtRKD2* misexpression could be detected in plant lines with ectopic expression of *AtRKD1-GFP* and *AtRKD2-GFP* fusion constructs (Fig. 3-18). The resulting fusion can easily be followed at the protein level with the help of the GFP reporter protein, fused to the protein of interest. The growth distorted tissue can be divided into a differentiating green part and an undifferentiated, colourless, proliferating outgrowth. No GFP signal is detectable in the differentiating green parts. To show that this is not an effect of different GFP detectablity between both parts, the GFP was detected in differentiated green tissue under the same conditions in the control double *CaMV35S*::*GFP* plant (Fig 3-18). The control plants were similar to the wild type plants, showing normal development. It is not yet clear, what causes the loss of *AtRKD*::*GFP* expression in the differentiating green part of the tissue. One explanation for this finding is that the transgene might be downregulated by the silencing machinery. Anyway, the two types of tissue allows to conclude that the presence of the AtRKD-GFP does positively correlate

with the undifferentiated state of the tissue. This leads to the preliminary suggestion that RKD factors are involved in the maintenance of the undifferentiated stadium of gametophytic cells.

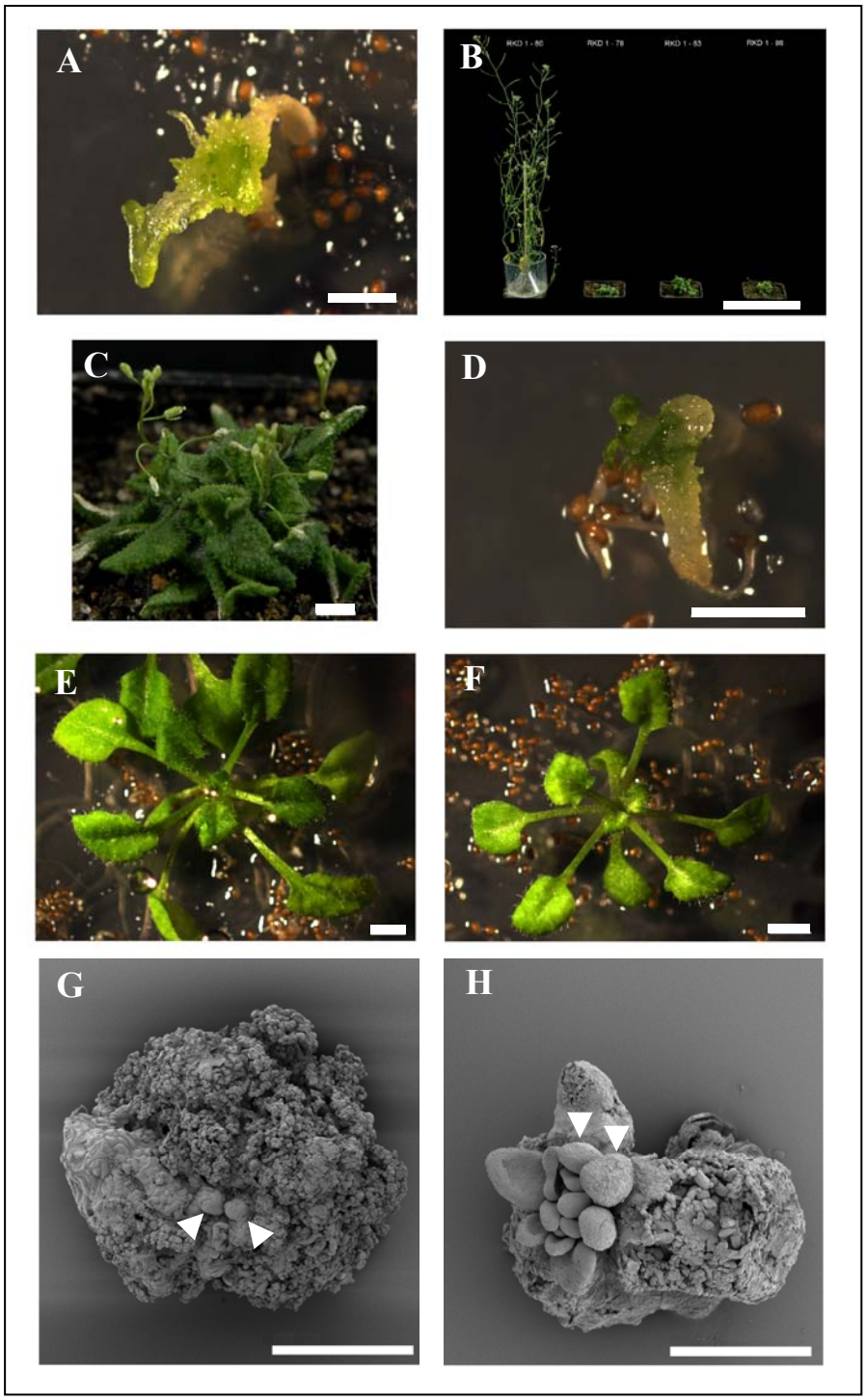


Fig. 3-17

Ectopic expression of the genes *AtRKD1*, *AtRKD2*, *AtRKD3* and *AtRKD4* under the control of the double *CaMV35S* promoter. A, B, C: *AtRKD1*, D: *AtRKD2*, E: *AtRKD3*, F: *AtRKD4*. G, H: scanning electron microscopic pictures of *AtRKD1* and *AtRKD2* overexpression tissues, respectively. The white arrows indicate somatic embryo-like sturctures. Scale bars represent in A, D – H 2 mm, in B 10 cm and in C 0,5 cm, respectively.

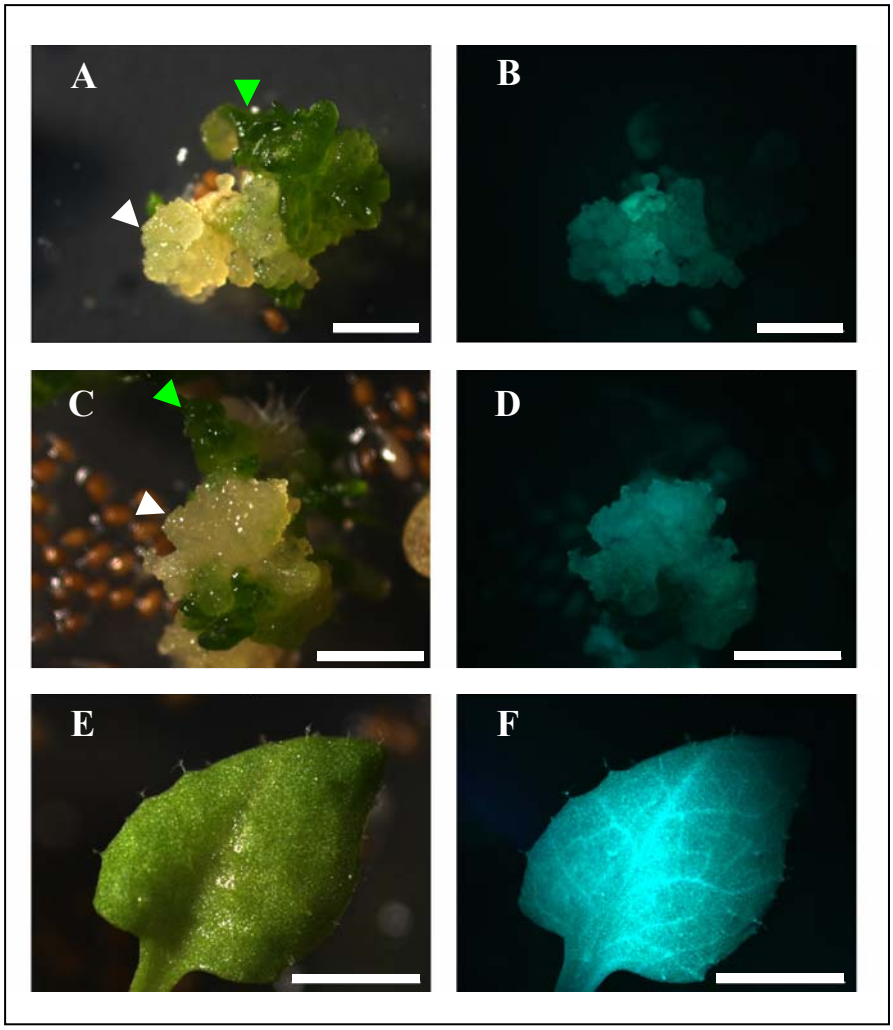


Fig. 3-18

Ectopic expression of *AtRKD-GFP* gene fusion constructs controlled by the double *CaMV35S* promoter. The green arrowheads indicate the green part. The white arrowheads show the white, proliferating tissue. A, B: *AtRKD1-GFP*. C, D: *AtRKD2-GFP*. E, F: control *CaMV35S*::*GFP*. A, C, E: white light images. B, D, F: UV pictures. Scale bars represent 2 mm.

3.2.10. Identification of transcription factors induced by *AtRKD* overexpression

To analyse in more detail the phenotype of transgenic lines, overexpressing the genes *AtRKD1* and *AtRKD2*, transcript profiles were determined. An array-containing gene fragments for 1,200 transcription factors was used (REGIA filter, Paz-Ares et al., 2002). From the 24 less distorted *AtRKD1* expressing plants, four plants showing comparably strong effect, were selected (Plant 78, 85, 85 and 88). One so called escape plant (an unsuccessfully transformed plant) (Plant 80) and two wild type plants were used for comparison. mRNA was

isolated from all seven plants. After synthesis of cDNA and labelling of the samples with [^{33}P], the second strand cDNAs were hybridized with the REGIA transcription factor array. Up- or downregulated transcription factor genes were identified (see Appendix). Genes were selected on the basis of the fold change of >2.0 and <0.5. Several of the induced transcription factor genes have been described as gametophytically expressed (Yu et al., 2005; Johnston et al., 2007; Steffen et al., 2007). These authors have identified in total 15 embryo sac specifically expressed transcription factors, 8 of those are represented on the REGIA macroarray. 4 of them were found to be upregulated in the *AtRKD1* overexpressing plants (Table 3-4). This data suggest that AtRKD1 might play a role to establish a gametophytic cell fate.

Gene ID	Description	induced	Yu et al. (2005)	Steffen et al. (2007)	Johnston et al. (2007)
At1g75430	homeodomain-containing protein	-	-	-	+
At2g01500	homeobox-leucine zipper protein	+	-	-	+
At2g37260	TRANSPARENT TESTA GLABRA 2 (WRKY44)	-	-	+	+
At2g45650	MADS-box protein (AGL6)	-	+	-	-
At3g29020	Myb family protein (MYB110)	-	-	-	+
At4g18770	Myb family protein (MYB98)	+	+	+	+
At5g01380	Expressed protein	+	+	-	-
At5g53950	No apical meristem (NAM) family protein	+	+	-	-
Σ	8	4			

Table 3-6

Upregulated transcription factors in the AtRKD1 overexpressing plants compared to published embryo sac specifically expressed transcription factors as described by Yu et al. (2005), Steffen et al. (2007) and Johnston et al. (2007).

3.2.11. Identification of gametophytic genes by quantitative RT-PCR platform (GeXP system) and *in situ* hybridisation from the proliferating white tissue

To further confirm the expression of gametophytic marker genes in the *AtRKD1* and *AtRKD2* induced proliferating tissue 8 genes were analysed by the semi-quantitative GeXP system (Beckman Coulter, CA, USA). The genes have been chosen as gametophytic marker genes, described by Johnston et al., (2007), Steffen et al., (2007), Yu et al., (2005): At1g02580 (Polycomb group protein, MEDEA), At1g18770 (*MYB98*), At1g26795 (self-incompatibility protein-related), At2g20070 (hypothetical protein), At2g35670 (Fertilisation independent seed, FIS2), At4g25530 (homeobox protein, FWA), At4g30590 (plastocyanin-like domain-containing protein) and At5g40260 (nodulin MtN3 family protein). The constitutively expressed gene *GAPDH* (At3g26650) was used as a control. Seedlings at the cotyledon stage and pistils before fertilisation have been used as a controls. As shown in Fig. 3-18 the following genes are upregulated in the proliferating tissue in comparison to seedlings used as a control: At1g18770 (*MYB98*), At1g26795 (self-incompatibility protein-related), At2g20070 (hypothetical protein), At4g25530 (homeobox protein, FWA), and At5g40260 (nodulin MtN3 family protein). In total five gametophytic marker genes were found to be induced in tissue overexpressing either *AtRKD1* and/or *AtRKD2*. These data support the above mentioned hypothesis that *AtRKD* genes act as triggers of female gametophytic development.

The array- and the RT-PCR data have been further tested by i*n situ* hybridisation (in cooperation with A. Johnston and U. Grossniklaus, Zürich University). As shown in Fig. 3-20 three, gametophytically expressed marker genes, *TD1* gene (T. Dresselhaus, unpubl. result), At5g40260 (nodulin MtN3 family protein), and At4g30590 (plastocyanin-like domain-containing protein), can be detected in the AtRKD1 induced white proliferating tissue. Taken together, the macroarray data, the quantitative RT-PCR experiments and *in situ* hybridisation data suggest the hyptothesis that AtRKD proteins are involved in the induction of a gametophytic pathway.

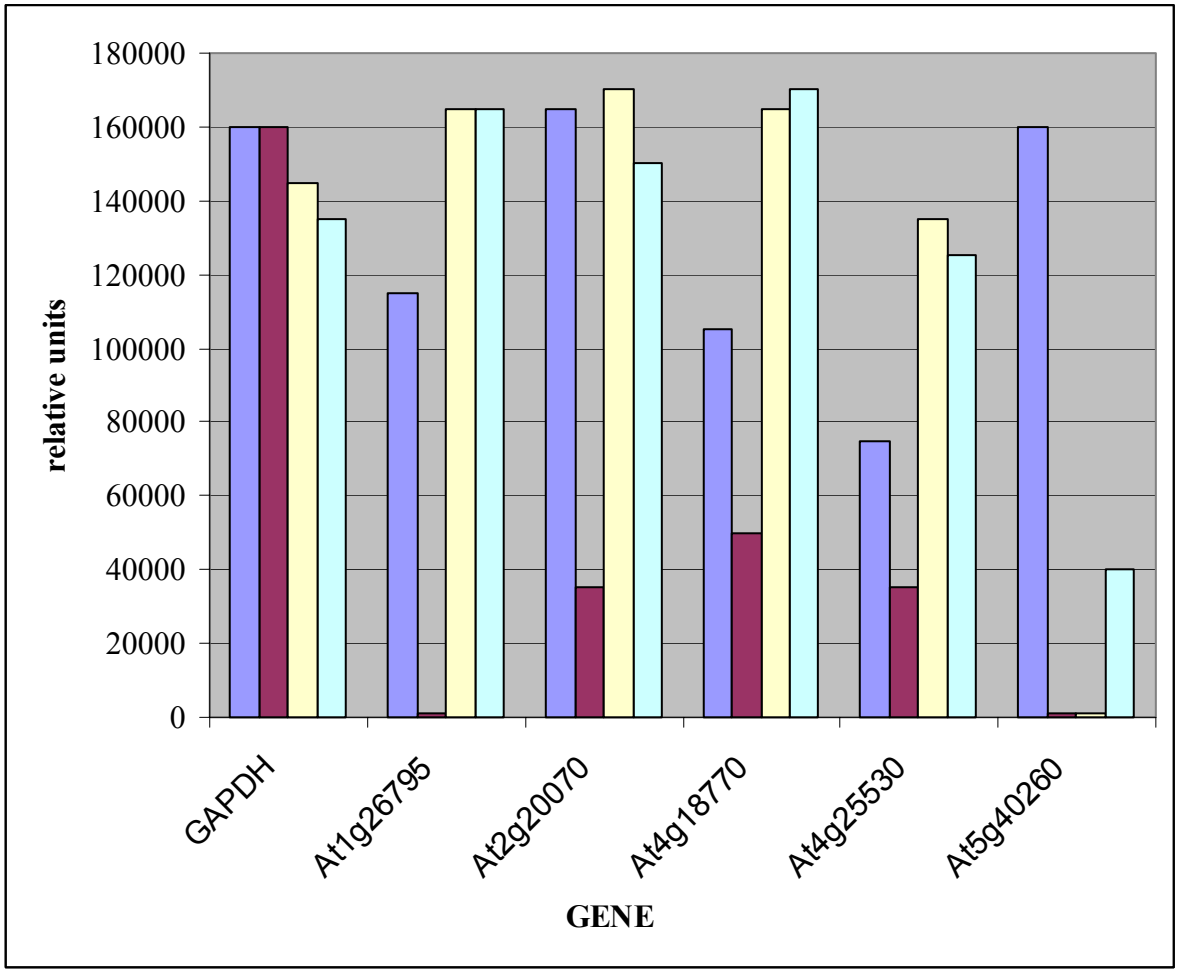


Fig. 3-19

Semi-quantitative RT-PCR analysis using the GeXP system. Gene IDs are indicated and signal intensity in relative units is given. Blue, purple, yellow and light blue indicates pistil, seedling, AtRKD1 and AtRKD2 white proliferating tissue, respectively.

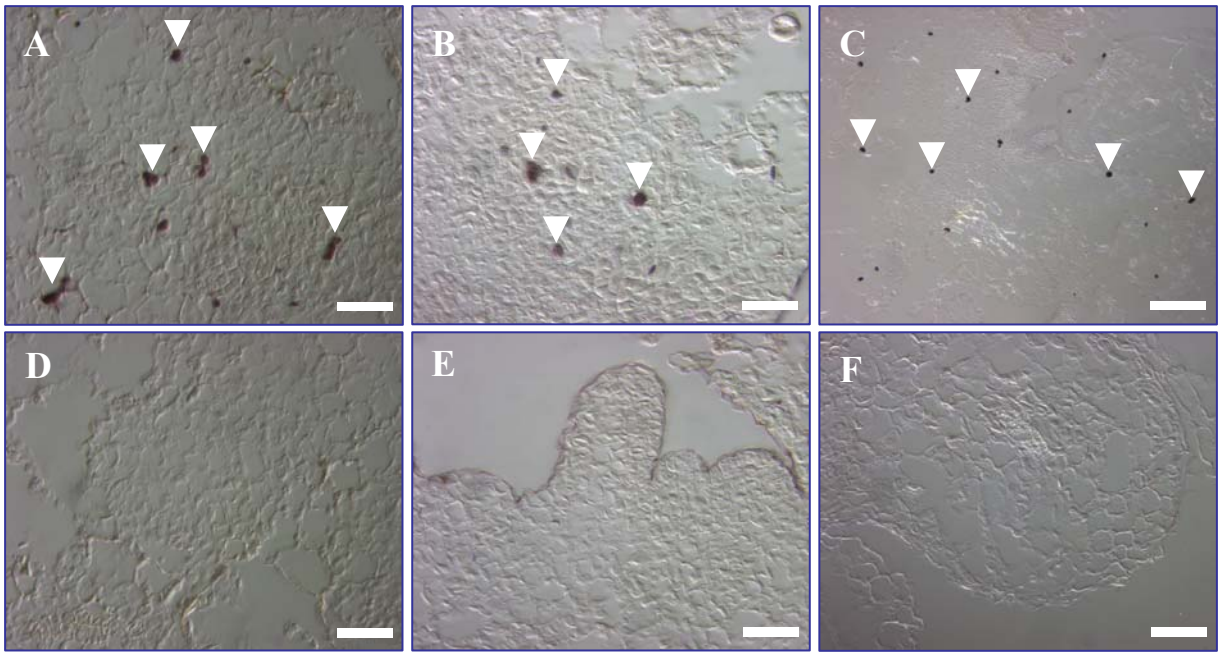


Fig. 3-20

In situ hybridisation assay on the *AtRKD1* overexpressing tissue using gametophytic marker genes. The white arrowheads show positive signal of the hybridisation. A, E: an egg cell specifically expressed gene (T. Dresselhaus, unpubl. result) B, D: At5g40260 (nodulin MtN3 family protein) C, F: At4g30590 (plastocyanin-like domain-containing protein). A, B, C: antisense E, D, F: sense control. Scale bar represent 20 μm.

3.2.12. Recombinant AtRKD1 and AtRKD2 proteins synthesised in bacteria

For further molecular characterisation, for instance binding studies and antibody generation, the *AtRKD1* and *AtRKD2* coding regions were expressed in *E. coli*, aiming to synthesise a soluble, recombinant protein under native conditions. The coding regions of *AtRKD1* and *AtRKD2* was cloned into the pQE30 plasmid (Invitrogen, CA, USA) and transformed into *E. coli* (strain M15). In this plasmid a His-tag can be found at the N-terminal part of the recombinant protein. To improve the solubility of the synthesised protein, induction was also done at lower temperatures of 28°C and 30°C. The majority of recombinant protein was found in the supernatant at induction temperature of 28°C (Fig. 3-21). The expected size of His-tagged AtRKD1 and AtRKD2 is 31,9 kDa and 34,3 kDa, respectively. The observed size of the obtained recombinant proteins are bigger than expected (37 kDa and 43 kD for AtRKD1 and AtRKD2, respectively). This phenomenon has been observed for other proteins as well (G. Mönke, personal communication).

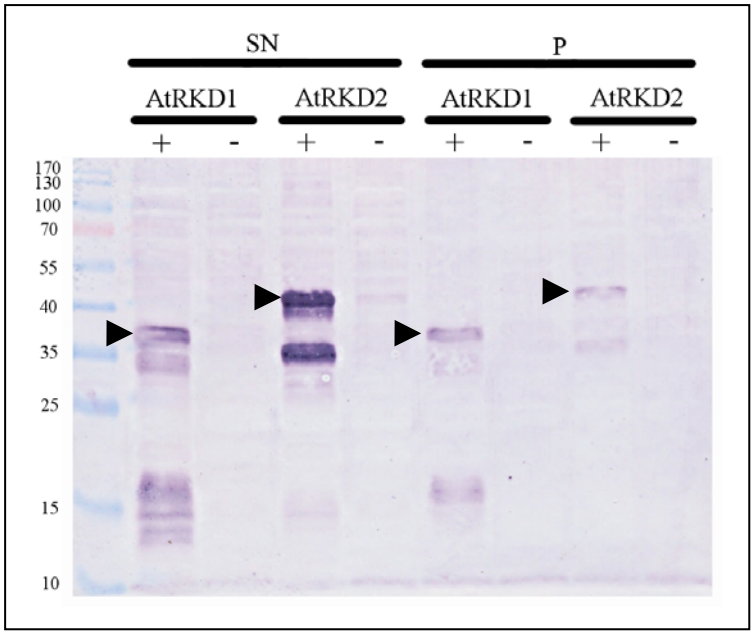


Fig. 3-21

Western blot analysis of recombinant AtRKD1 and AtRKD2 proteins synthesised in *E. coli* (strain M15) at 28°C (+ induced, - non induced samples). The black arrowheads indicate the full length His-tagged proteins. The protein synthesis was induced by IPTG at a final concentration of 2 mM. The induction was performed for 4 hours. Afterwards the cells were harvested by centrifugation and lysed by ultrasonic treatment. 12,5% (w/v) SDS-polyacrylamide gels were used to separate the proteins. Then the gel was blotted to a nitrocellulose membrane and the His-tagged proteins were detected by a colorimetric method (NBT/BCIP). SN-supernatant, P-pellet. The size marker is given in kDa.

4. *DISCUSSION*

4.1. Main observations and conclusions

RKD genes of wheat and *Arabidopsis* are described as novel, plant specific, evolutionarily conserved family of transcription factors with embryo sac-specific expression. A structurally related gene, called *MID* controls gamete differentiation in *Chlamydomonas.* The subcellular localisation in the nucleus supports the function of RKDs as transcription factors. T-DNA insertion mutants exhibit distortions in the development of the gametophytic generation. Constitutive misexpression of *RKD* genes causes severe defects in differentiation and the generation of proliferating tissue. The proliferating tissue can be divided into a differentiating green part, without detectable *RKD* gene expression and white part expressing *RKD* genes. This white, proliferating tissue expresses gametophyte specific marker genes. Preliminary data suggest, that the RNAi mediated posttranscriptional silencing of *RKD* genes in transgenic wheat and *Arabidopsis* causes seed abortions, further supporting an essential function of *RKD* genes in the developing ovule.

Taken together, expression patterns, evolutionary conservation, loss of function mutants, misexpression phenotypes as well as the expression of RNAi constructs in transgenic wheat and *Arabidopsis* lead to the hypothesis that *RKD* genes are involved in the control of gametophyte development, a basic biological process first described by Wilhelm Hofmeister more than 150 years ago.

4.2. A subtractive approach to isolate egg cell specific genes - a proof of concept

cDNA libraries from egg cells of sexual and parthenogenetic Salmon wheat lines have been generated (Kumlehn et al., 2001). To identify egg cell specifically expressed genes, an experimental and virtual subtraction strategy has been applied (Kumlehn et al. 2001; Altschmied et al., unpubl.). This approach is based on the assumption that egg cell specifically expressed genes are of low abundance or absent at all in EST collections from conventional cDNA libraries due to the dilution of egg cells within other tissues and cells. Up to now 17 egg cell specific candidate genes have been identified. The *RKD* gene family is described in more detail in this thesis. *RKD* genes are specifically expressed in the egg apparatus as demonstrated both for wheat and *Arabidopsis*. cDNA array hybridisation revealed, that in the parthenogenetic egg cell cDNA library only 9 (0,016%) RWP-RK clones can be found, while in the sexual 168 (0,305%) out of 55.000 clones (D. Koszegi et al.,

unpubl.). This shows the suitability of the subtraction approach and provides a first proof of concept. In addition to the *RKD* family, embryo sac specificity has been confirmed for two additional candidate genes as demonstrated by single cell RT-PCR (Koszegi et al., in prep.).

4.3. *RKD* genes as transcription factors

RKD genes have been suggested to function as transcription factors regulating downstream genes (Schauser et al., 2005). This suggestion is mainly based on the faint similarity of RKD factors to members of the bZIP transcription factor family characterised by a short DNA binding motif consisting of several basic amino acids and the presence of an adjacent heptameric array of leucine residues potentially involved in protein-protein interaction. In this view, the highly conserved RWP-RK motif of RKD proteins is considered to be functionally analogous to the basic motif of bZIP transcription factors, and a heptameric array of hydrophobic amino acids, like leucine, isoleucine, valine or methionine, found in RKD factors is thought to function as protein protein interacting domain in RKD factors.

Up to now no experimental evidence has been published supporting the theory that RKD proteins indeed act as transcription factors. Therefore, the subcellular localisation of RKD proteins using GFP fusions have been analysed. As expected for transcription factors, all four *Arabidopsis* RKD proteins (AtRKD1-4) are localised in the nucleus, supporting that they might act as transcriptional regulators of downstream genes.

The presence of a regular heptameric array of hydrophobic amino acids N-terminal of the RWP-RK motif suggested that RKD functions require the formation of homo- or heterodimers. Preliminary results based on protein-protein interaction experiments in plant protoplasts however indicate that at least AtRKD1 and AtRKD2 form neither homo- nor heterodimers (M. Michael and D. Koszegi et al., in prep.). Although this observation needs further confirmation, it might demonstrate functional differences between RKD proteins and bZIP transcription factors.

Additional indication for the function of *RKD* factors as transcriptional regulators comes from multiparalell expression analysis using cDNA arrays. Plants with ectopic overexpression of the *AtRKD1* gene, show a dramatically changed expression of several transcription factor genes, about 46 transcription factor genes up- and about 39 transcription factor genes downregulated. Ongoing experiments are based on glucocorticoid regulated expression or estrogen-regulated modulation of the subcellular localisation of *RKD* genes and

will analyse changes in the genome wide transcriptome of *Arabidopsis* (M. Michael and D. Koszegi, et al. in prep.).

RKD proteins are easily recognised by the presence of the highly conserved RWP-RK motif. Using this motif for extensive database searches reveals that RWP-RK motif containing gene products are restricted to plant genomes, suggesting that they are involved in the regulation of plant specific processes. Perhaps *RKD* genes have been lost in the animal kingdom.

4.4. Expression pattern of *RKD* genes and "loss of function" mutants

Single cell RT-PCR, *in situ* hybridisation, and the activity of promoter reporter gene fusions suggest that *TaRKD* as well as *AtRKD1* and *AtRKD2* are specifically expressed in the mature egg cell of wheat and *Arabidopsis*, respectively. This late gametophytic expression is in contrast to the early distortion of the gametophytic development in the described *rkd1-1*/*rkd2-2* and *rkd1-2*/*rkd2-2* double mutants of *Arabidopsis*, which stop gametophytic development at the stage of the functional megaspore. Possible explanations of this discrepancy: (i) the *RKD* gene expression level in the functional megaspore might be to low to be detectable by *in situ* hybridisation. (ii) the transcriptional promoter GUS fusion constructs might exclude essential regulatory elements for instance in 5'-UTR, coding region and/or intron. Current experiments with translational fusions are performed to test this hypothesis. (iii) the mutant phenotype might be misinterpreted, since other research groups also reported megaspore arrest in T-DNA insertion lines, which was unlinked to the T-DNA (A. Johnston, pers. comm.). Especially the latter observation requires further attention, since genotyping experiments on *rkd1*/*rkd2* double mutant reveal, that genotype and phenotype are not coupled tightly. Therefore, ongoing experiments aim to elucidate this further and include the following approaches. (i) the single T-DNA insertion lines are currently genetically purified by backcrossing to wild type plants. (ii) due to the possible functional redundancy among the four *AtRKD* genes, current experiments aim to the silencing of all four *AtRKD* genes together using RNAi constructs, but by under the control of the authentic *RKD* gene promoter. Using female gametophyte specific promoter should provide at least two advantages: a) the corresponding transcript(s) can be specifically knocked down in the female gametophyte and b) the promoter will be definitely active in the embryo sac. This is not precisely known for the currently used *CaMV35S* promoter (A. Johnston, pers. comm.). Additionally the genes *AtRKD3* and *AtRKD4* will be included to create triple and/or quadruple mutants to obtain

complete "loss of function" situation. Such mutant might be favourable compared to the above suggested RNAi strategy for the following reasons. (i) in ideal knock-out mutant the full length transcript is totally lost, whereas in RNAi transgenic plants the loss of transcript can vary between 0-100%. (ii) easier to follow the genetics of mutants and the segregating ratios can be more informative for gametophytic effects. In case of single, female gametophytically important gene the segregation ratio is expected to be 1:1 instead of the normal 3:1. Such segregation can not be shown in RNAi lines, because of the dominant effect of silencing. (iii) it is experimentally difficult to prove an RNAi mediated effect at the transcript level, especially for weakly, embryo sac specifically expressed genes like *AtRKD*s.

4.5. *RKD* misexpression induced phenotype

The phenotypes of plants overexpressing the *RKD* genes *AtRKD1-4* can be divided into two groups. The ectopic expression of the genes *AtRKD1* and *AtRKD2* leads to the generation of a proliferating tissue similar to that induced by the ectopic expression of genes like *WUSCHEL* (Zuo et al., 2002), *LEAFY COTYLEDON1* (Lotan et al., 1998), and *BABY BOOM* (Boutilier et al., 2002). The phenotype caused by the overexpression of the genes *AtRKD3* and *AtRKD4* is less severe. Slight reduction of the seed set, due to abnormalities in ovule development has been observed in *AtRKD3* overexpressing plants. No obvious deviation from wild type could be detected in plants overexpressing the gene *AtRKD4*. Functional subspecialisation or different interaction partners might be explanations for the observed differences between the four genes.

4.6. *RKD* overexpression induces the expression of gametophytic marker genes

Constitutive misexpression of *RKD* genes causes severe defects in differentiation and the generation of proliferating tissue. The proliferating tissue can be divided into a differentiating green part, without detectable *RKD* gene expression and white part, expressing *RKD* genes. Array experiments, RT-PCR analysis and *in situ* hybridisation approaches reveal, that *RKD* gene expression causes the expression of several female gametophyte-specific marker genes. Among them are for instance the mainly female gametophyte-specific expressed genes At1g18770 (*MYB98*), At1g26795 (self-incompatibility protein-related), At2g20070 (hypothetical protein), At4g25530 (homeobox protein, FWA), and At5g40260 (nodulin MtN3 family protein). The RKD-induced expression of these marker genes leads to

the conclusion, that *RKD* genes redirect developmental pathways towards gametophytic direction. Obviously, this conclusion depends on the gametophyte specificity of the indicated marker genes. However, at least for one of them doubts may be raised. Thus, the synergid expressed gene *AtMYB98* also expressed after fertilisation as for instance in the trichomes of leaves (Kasahara et al., 2005).

Taken together, our data suggest a model in which *RKD* genes are involved in the induction of a gametophytic cell fate. However, further analysis is required to confirm, that the expression of these female gametophytic markers is specific for AtRKD induced proliferating tissue and not found in any other proliferating tissue. To test this, currently an auxin induced callus tissue is being analysed as negative control. This can be further extended by the use of additional transcription factor-induced proliferating tissues, like *BBM* (Boutilier et al., 2002) and/or *WUS* (Zuo et al., 2002) overexpressing tissue.

4.7. RKD is plant specific transcription factor

RKD genes are exclusively found in plant genomes, like for instance: resurrection plant (*Selaginella* sp.), grape vine (*Vitis vinifera*), poplar (*Populus* sp.), rice (*Oryza sativa*), maise (*Zea mays*), barley (*Hordeum vulgare*), etc. (L. Altschmied, pers. comm.). This finding strongly argues for function of RKD proteins in plant specific developmental process. By comparing reproductive processes between plants and animals, it becomes obvious that in both systems one cell is selected to undergo meiosis. This cell is the megaspore mother cell in plants and the oocyte in animals. Plants and animals differ in egg cell differentiation. The surviving meiotic product of animals does not divide further. In contrast, the surviving meiotic product of plants, the functional megaspore, undergoes further mitotic divisions generating the proliferating and differentiating gametophyte. In mosses and ferns for instance, the gametophyte can form separate multicellular organism, independent from the sporophyte. Thus, the female gametophyte is often considered as the “forgotten generation” with its own developmental programme. Therefore, this finding suggests, that RKD proteins are involved in a plant specific process of female gametophyte development. This hypothesis is supported by the following observations: a) the absence of *RKD* genes in the genomes of animals, which lack female gametophytic proliferation, b) the female gametophyte-specific expression of *RKD* genes and c) the expression of female gametophytic marker genes in the RKD protein-induced proliferating tissue.

4.8. Possible applications of *RKD* genes and promoters to manipulate components of apomixis

It is conceivable to exploit a gene-construct, consisting of a nucellus specific gene promoter fused to an *RKD* gene to induce additional, non-reduced and therefore aposporic embryo sacs in addition to the sexual one. Another possible application of *RKD* genes can be the induction of pollen embryogenesis. This could be achieved for example by the use of tapetum specific promoter TA-29 (Koltunow et al., 1990). The female gametophyte-specific activity of *RKD* gene promoters in the egg apparatus both in monocot wheat and dicot *Arabidopsis* challenges their application to manipulate reproductive processes in plants. Moreover, the egg apparatus specific activity of the *AtRKD1* and *AtRKD2* gene promoters could be used to misexpress suitable candidate genes to induce fertilisation-independent parthenogenetic egg cell activation. Candidate genes could be recently isolated genes with the capacity to induce somatic embryo formation, such as *WUSCHEL* (Zuo et al., 2002), *LEAFY COTYLEDON1* (Lotan et al., 1998), and *BABY BOOM* (Boutilier et al., 2002).

4.9. A putative egg cell-specific *cis*-regulatory element

Several sequence motifs directing promoter-specificity are known in plants. Regulatory elements involved in gametophyte-specific gene expression have not been identified, except recently described egg-apparatus-specific element (Yang et al., 2005). A novel, putative *cis*-motif was identified, which is present in promoters of six egg cell-expressed genes. The 9 bp motif is rather long in comparison with other *cis*-elements. Work is in progress to analyse the functional importance of the identified motif using promoter deletion and mutation series. If the motif would turn out to be crucial for egg cell-specific gene expression, “one-hybrid screen” (Li and Herskowitz, 1993) might also be suitable to identify corresponding regulatory protein(s).

4.10. Evolutionary conserved developmental pathways of gamete formation?

Both the isolated *TaRKD* genes of wheat and the described *RKD* genes of *Arabidopsis* exhibit sequence similarity to other plant genes encoding the conserved RWP-RK domain, including for instance the *NIN* (nodule inception) genes of *Lotus japonicus* and *Pisum sativum*. Moreover, there are at least 13 RWP-RK genes in the genome of *Chlamydomonas.*

One of them is the gene *MINUS DOMINANCE* (*MID)* (Ferris and Goodenough, 1997). It is most intriguing, that the *MID* gene product is necessary and sufficient for the minus gamete development (Fig. 4-1). In *Chlamydomonas* the gametogenesis can be triggered by deprivation of nitrogen (Sager and Granick, 1954). The higher plant NIN proteins have been shown also to be involved in the formation of infection threads and the initiation of symbiotic nitrogen-fixing root nodules. This suggests that at least one subclass of RWP-RK proteins is involved in nitrogen dependent pathways.

The *MID* gene is also expressed at low level in vegetative cells and induced by nitrogen limitation (Lin and Goodenough, 2007). This phenomenon raises the question, whether *RKD* genes of higher plants are also regulated by nitrogen limitation. Preliminary data from expression analysis in transiently transformed *Arabidopsis* protoplasts using *AtRKD2* promoter *GUS* reporter construct show, that the promoter becomes slightly activated under limited nitrogen supply. Although this finding needs further confirmation, it indicates, that an evolutionarily old regulatory interconnection between nitrogen metabolism and gamete development still exist in higher plants.

It is also interesting to speculate, that the *MID* gene and *RKD* genes are functionally homologues and whether the mt- gametes of *Chlamydomonas* can be considered as an ancestor of a higher plant egg cell. As mentioned above the expression pattern of the *MID* gene is mainly restricted to the minus gametes. Similarly, the expression of *RKD* genes is detectable in the egg apparatus. In addition to the shared gamete-specific expression, the phenotype of the *Arabidopsis rkd* double mutant showing developmental arrest at the stage of the functional megaspore also suggests similar function of RKD and MID in gametogenesis. Thus, both *RWP-RK* gene products most likely share similar functions in the evolutionarily old process of gamete formation. The mt+ gamete is considered to be maternal and the mt- gamete as paternal (Birky, 1995). This is mainly based on the origin dependent degradation of organelles. Obviously, this is in contradiction to the expression data described above, suggesting that the mt- and the higher plant egg cell share the expression of *RWP-RK* genes. However, there is no single mechanism for uniparental inheritance and the donor of the subcellular organelle can vary (Birky, 1995). In mosses, ferns and higher plants the chloroplasts are maternally inherited *via* the egg cell, whereas in *Chlamydomonas* the mt- cpDNA is degraded after the fusion of both gametes (Birky, 1995). In contrast, the mitochondria are paternally inherited in higher plants. Thus, although no final conclusion can be drawn, the expression pattern shared by the *MID* gene and higher plant *RKD* genes in mt-

gametes and the egg cell, respectively, suggest that mt- gametes might be considered as ancestors of the higher plant egg cell.

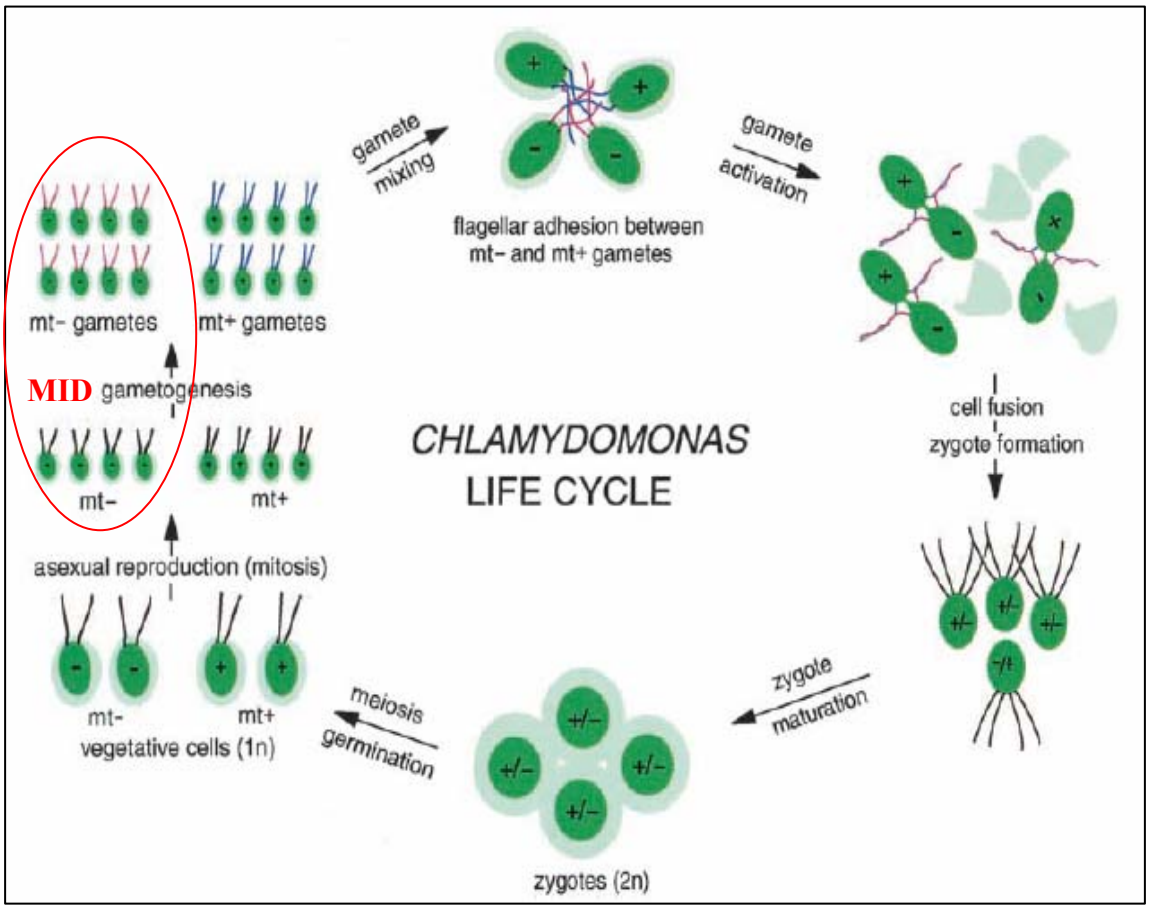


Fig. 4-1

Schematic representation of the *Chlamydomonas* life cycle according to Zhao et al. (2001). Gametes are formed during the gametogenesis from haploid vegetative cells. The gamete formation is induced by nitrogen deprivation. The *MID* gene function is required for the development of minus gametes. When gametes of different mating type meet, they adhere *via* agglutinins. After gamete activation, zygotes are formed. Under sufficient nutrient supply zygotes undergo meiosis. Haploid vegetative cells can propagate *via* asexual reproduction. Gametogenesis is induced under low nitrogen levels completing the life cycle.

4.11. Are gametophytic cells stem cells?

Stem cell research is one of the most booming fields in developmental biology, because of the basic properties of stem cells and their putative therapeutical importance. Stem cells are undifferentiated cells having the capacity of self-renewal and regeneration of the whole organism (totipotent stem cells) or at least several tissues or cell types (pluripotent stem cells). The classical stem cell concept is mainly based on animal experimental systems. Here, the zygote is considered as the totipotent cell in the animal life cycle, since it can differentiate

into all cells, tissues and organs of the organism. From pluripotent stem cells, most, but not all types of cells and tissues can be derived. Animal stem cells are localised in special so called stem cell niches, as for instance the bone-marrow.

In plants, meristems share two properties of stem cells, the self-renewal and the ability to develop into different cells and tissues. Based on this observation the shoot and root apical meristems should contain stem cells, if the stem cell concept is applied to plants (Laux, 2003; Lenhard and Laux, 2003). Obviously, establishing the stem cell concept in plant biology has been a long lasting process, most likely due to the well known fact that many plant cells can in principle regenerate into a whole plant.

The *WUSCHEL-CLAVATA3* system represents the first identified genes, involved in the maintenance of the stem cell niche in the plant shoot apical meristem (Mayer et al., 1998; Schoof et al., 2000). A similar mechanism including related genes (*WOX5*, *CLE40*), has been identified in the root apical meristem (Haecker et al., 2004; Hobe et al., 2003). The stem cell concept has been extended further to embryonic cells (Verdeil et al., 2007). Embryogenic cells are able to form somatic embryos without additional external stimulus. Following this view, the embryonic cells resembles the zygote, the totipotent cell in the life cycle of animals and plants. Although several genes have been described to induce the formation of somatic embryos, the molecular mechanism how cell develops into embryonic cell is poorly understood. Genes which are involved in this process are for instance: *SOMATIC EMBRYOGENESIS RECEPTOR-LIKE KINASE* (*SERK*) (Schmidt et al., 1997), *WUSCHEL* (*WUS*) (Zuo et al., 2002), *LEAFY COTYLEDON1* (*LEC1*) (Lotan et al., 1998), *LEAFY COTYLEDON2* (*LEC2*) (Stone et al., 2001), *BABY BOOM* (*BBM*) (Boutilier et al., 2002) and *AGAMOUS-LIKE15* (*AGL15*) (Harding et al., 2003). The ectopic expression of the genes *AtRKD1* and *AtRKD2* also results in the formation of proliferating tissues and somatic embryos. As discussed above, the proliferating tissue can be divided into differentiated green and non differentiated white part. Embryos are exclusively seen on the green part of the outgrowth. In contrast, the white part represents proliferating tissue, which can be maintained under sterile conditions for up to six months. The described observation, that the fluorescence signal of a RKD-GFP fusion protein cannot be detected in the green tissue, suggests that *RKD* expression might suppress the ability to differentiate. Currently, it is not clear why the expression of the fusion construct is lost in the differentiating green part, although transcriptional or posttranscriptional silencing might be an explanation.

The phenotype and the expression pattern of the *RKD* expressing tissue let to consider the stem cell character of gametophytic cells, especially of the egg cell. At the beginning of

the gametophyte development (the sporophyte-gametophyte transition) cells should adopt the new gametophytic developmental programme leading to the fully differentiated cells types of the embryo sac like antipodal cells and synergids. In the mature gametophyte there are only two cell types, the egg cell and the central cell, which will give rise to newly developing tissues. The central cell is omnipotent or pluripotent (but not totipotent), since it differentiates into the endosperm only. According to this view, the zygote alone would be a totipotent cell, since it develops into an embryo and a whole plant.

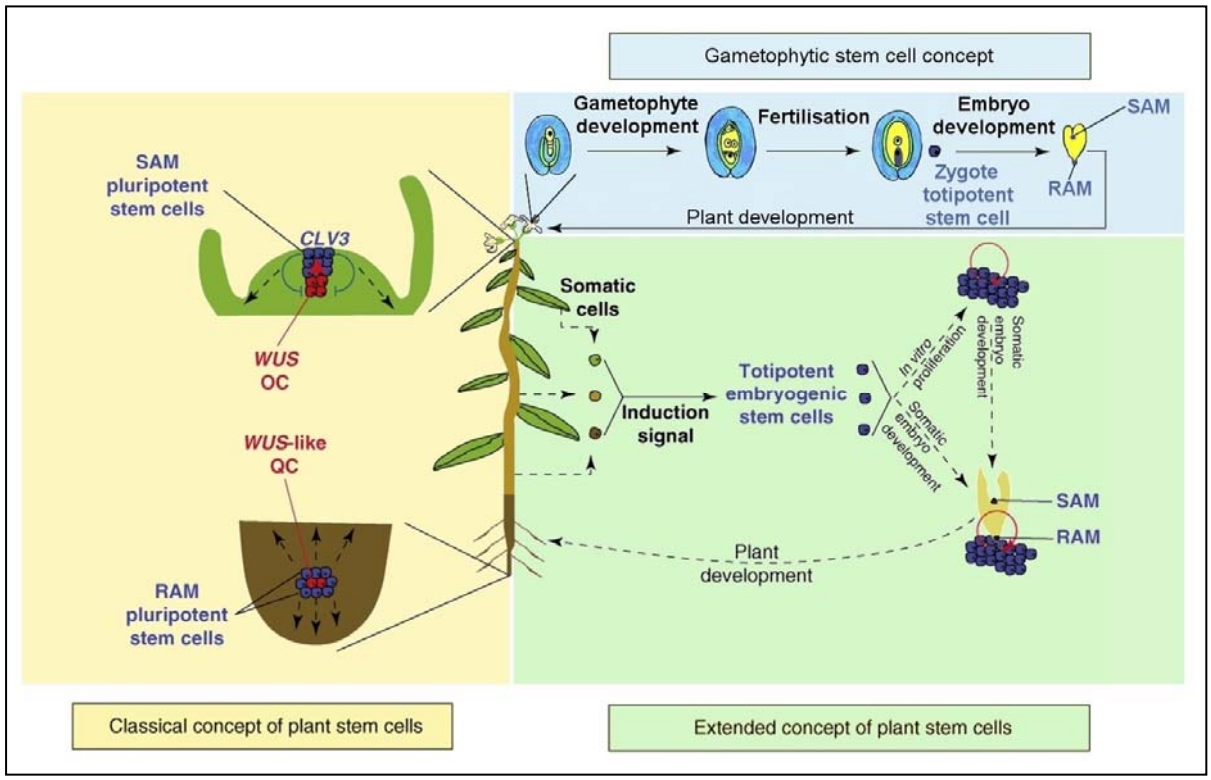


Fig. 4-2

Schematic representation of stem cell concepts in plant biology (modified from Verdeil et al., 2007). The yellow and green part is adapted from Verdeil et al. (2007). During vegetative growth the plant body is derived from stem cells found in the shoot (SAM) and root (RAM) apical meristems. Both, SAM and RAM are generated during embryogenesis. The origin of embryo development is the zygote. An alternative pathway to generate a new clonal plant is the formation of somatic embryos from totipotent embryogenic stem cells. The organizing centre (OC) and the quiescent centre (QC) are represented in red, stem cells are shown in blue.

5. *ABSTRACT*

A typical plant life cycle consists of two, alternating generations, the sporophyte and the gametophyte. Usually the gametophyte generates gametes and the sporophyte produces spores. The female gametophyte development starts with the germination of the meiotically reduced functional megaspore. The sporophyte development begins after fertilisation with the first division of the zygote.

Several plant species deviate from this sexual reproduction process and follow an apomictic pathway. Apomictic reproduction includes three major components: a) the avoidance of meiotic reduction (apomeiosis), b) the fertilisation-independent initiation of embryogenesis (parthenogenesis) and c) the autonomous or pseudogamous development of the endosperm.

Two cDNA libraries have been generated from sexual and parthenogenetic egg cells of isogenic wheat lines of the Salmon system. A strategy which combines experimental and virtual subtraction approaches identifies several putative egg cell-specifically expressed genes. Among them, there is one gene with sequence similarity to *MINUS DOMINANCE* (*MID*), a gene, known to be essential for minus gamete development in the green algae *Chlamydomonas reinhardtii*.

The identified wheat gene belongs to a gene family of at least four genes, designated as *TaRKD1-4*, since they exhibit sequence similarity with a family of strictly plant specific, putative novel transcription factors which share the highly conserved RWP-RK domain. The gene family can be further subdivided into the *NLP*- and the *RKD*-subfamily. Up to now little is known about the function of these genes. Therefore, the studies described here aimed at a first functional characterisation of *RKD* genes in wheat and *Arabidopsis*.

The *RKD* gene family of wheat consists of at least four different genes, with two of them being expressed. Single cell RT-PCR detects transcripts in the egg apparatus, especially in the egg cell and the synergids, as well as in the zygote. Preliminary results of RNAi experiments suggest that *TaRKD* genes might be involved in embryo sac development.

The genome of the model plant *Arabidopsis thaliana* contains at least five *RKD* genes. Using RT-PCR, transcripts of *AtRKD1,2,3,4* can only be detected in flower buds, whereas *AtRKD5* exhibits a more ubiquitous expression pattern. Based on *in situ* hybridisation, transcripts of *AtRKD1,2,4* have been specifically localised in the egg apparatus of mature embryo sacs. Promoter-GUS fusions revealed that the *AtRKD1*- and *AtRKD2*-gene promoters are specifically active in the egg apparatus of the mature gametophyte. A search for common promoter elements reveals, that all four *AtRKD* genes with expression in the egg apparatus share an at least nonameric sequence motif also found in the promoters of two other egg cell-specifically expressed genes. Together these data suggest that *AtRKD* genes are specifically functional during gametophyte development, most likely in the egg apparatus of the mature embryo sac.

Transient expression of *AtRKD*::*GFP* gene fusions reveals the localisation of AtRKD1,2,3,4 proteins in the cell nucleus. The nuclear localisation supports the assumption that RKD proteins function as transcription factors.

To analyse a "loss of function" situation, T-DNA insertion mutants of *AtRKD1* and *AtRKD2* genes have been investigated. Homozygous single gene T-DNA insertion mutants for *AtRKD1,2,3,4* do not show an obvious phenotype, suggesting functional redundancy. In two different double mutants (*rkd1-1/rkd2-2* and *rkd1-3/rkd2-2)* a 40% reduction of seed set was found. The observed seed abortion was most likely caused by the distortion of the female gametophyte development at the stage of the functional megaspore. Since, this phenotype was not tightly linked to the T-DNA insertion, it needs further investigations.

Transgenic lines with two different RNAi constructs directed against *AtRKD2* have been analysed. A reduction of seed set was observed in several plants. Ovule clearing revealed a distortion of the gametophyte development at the level of the functional megaspore in some plants, but needs to be further analysed in the next generation.

"Gain of function" experiments were performed using the coding sequence of the genes *AtRKD1,2,3,4* driven by the double *CaMV35S* gene promoter. Severe growth distortions and callus like proliferation were observed for *AtRKD1* and *AtRKD2*. In case of *AtRKD3* overexpression the seed set is slightly reduced, whereas no effect could be found in plants with ectopic expression of *AtRKD4*. The AtRKD1- and AtRKD2-induced proliferating

tissue can be divided into a white and a green part. Similar proliferating tissue can be induced by ectopic expression of AtRKD1-GFP and AtRKD2-GFP fusion proteins, which allow to follow their expression at the protein level. Both fusion proteins can only be detected in the non-differentiating, white proliferating tissue and are absent in the differentiating green part of the callus.

cDNA arrays, RT-PCR and *in situ* hybridisation revealed the expression of several gametophytically expressed marker genes in the AtRKD1- and AtRKD2-induced white proliferating tissue, suggesting that AtRKD1 and AtRKD2 might trigger a gametophytic developmental pathway.

Although further work is required to reveal the function of RKD proteins, our data suggest a crucial role for the development of the female gametophyte.

6. *ZUSAMMENFASSUNG*

Der typische Lebenszyklus einer Pflanze umfasst zwei aufeinander folgende Generationen, den Sporophyten und den Gametophyten. Der Gametophyt generiert Gameten und der Sporophyt produziert Sporen. Die weibliche Gametophytentwicklung beginnt mit der Auskeimung der meiotisch reduzierten funktionellen Megaspore. Die Sporophytentwicklung beginnt nach der Befruchtung mit der ersten Teilung der Zygote.

Einige Pflanzenarten weichen von dieser sexuellen Reproduktion ab und folgen einem apomiktischen Entwicklungsweg. Apomiktische Reproduktion umfasst drei wesentliche Komponenten: a) die Vermeidung der meiotischen Reduktion (Apomeiose), b) die befruchtungsunabhängige Initiation der Embryogenese (Parthenogenese) und c) die autonome oder pseudogame Entwicklung des Endosperm.

Zwei cDNA-Bibliotheken wurden von sexuellen und parthenogentischen Eizellen isogener Weizenlinien des Salmon-Systems erzeugt. Durch eine Strategie, die auf experimentellen und virtuellen Subtraktionstechniken beruht, wurden Kandidaten für eizellspezifisch exprimierte Gene identifiziert. Eines dieser Gene besitzt Sequenzähnlichkeit mit *MINUS DOMINANCE (MID),* einem Gen, das essentiell für die Entwicklung von Minus-Gameten der Grünalge *Chlamydomonas reinhardtii* ist.

Das identifizierte Gen gehört zu einer Familie von mindestens vier Genen, die als *TaRKD1-4* bezeichnet wurden, da sie Sequenzähnlichkeit mit einer Familie pflanzenspezifischer, neuer Transkriptionsfaktoren besitzen, die durch eine gemeinsame, hoch konservierte RWP_RK-Domäne charakterisiert sind. Die Genfamilie kann in die NLP- und die RKD-Subfamilie unterteilt werden. Bisher ist die Funktion dieser Gene weitgehend unbekannt. Ziel der vorliegenden Arbeiten war die funktionelle Charakterisierung von *RKD*-Genen in Weizen und *Arabidopsis*.

Die *RKD*-Genfamilie des Weizens besteht aus mindestens vier verschiedenen Genen, von denen zwei exprimiert werden. Durch Einzelzell-RT-PCR wurden Gentranskripte nur im Eiapparat, speziell in der Eizelle und den Synergiden, sowie in der Zygote nachgewiesen.

Vorläufige Ergebnisse von RNAi-Experimenten weisen darauf hin, dass *TaRKD*-Gene an der Entwicklung des Embryosacks beteiligt sind.

Das Genom der Modellpflanze *Arabidopsis thaliana* enthält mindestens fünf *RKD*-Gene. Durch RT-PCR wurden Transkripte der Gene *AtRKD1,2,3,4* ausschließlich in Blütenknospen gefunden, während *AtRKD5*-Transkripte ein ubiquitäres Expressionsmuster zeigen. Mit Hilfe von *in situ*-Hybridisierungstechniken wurden die Transkripte von *AtRKD1,2,4* ausschließlich im Eiapparat des reifen Embryosacks nachgewiesen. Promotor-GUS Fusionen zeigten, dass die Promotoren der Gene *AtRKD1* und *AtRKD2* spezifisch im Eiapparat des reifen Embryosacks aktiv sind. Die Suche nach gemeinsamen Promotorelementen zeigt, dass in den Promotoren aller vier im Eiapparat exprimierten Gene ein neun Basenpaare langes Sequenzmotiv vorkommt. Dieses Motiv wird auch in den Promotoren zweier weiterer eizellspezifisch exprimierter Gene gefunden. Alle diese Beobachtungen weisen darauf hin, dass *AtRKD*-Gene spezifische Funktionen für die Entwicklung des Gametophyten, insbesondere des Eiapparat im reifen Embryosack, besitzen.

Durch transiente Expression von *AtRKD*::*GFP* Genfusionen wird gezeigt, dass die Proteine AtRKD1,2,3,4 im Zellkern lokalisiert sind. Dies unterstützt die Vermutung, dass RKD Proteine als Transkriptionsfaktoren wirken können.

Um die Auswirkungen des Verlustes der RKD-Funktion zu ermitteln, wurden T-DNA-Insertionsmutanten der Gene *AtRKD1* und *AtRKD2* untersucht. Homozygote Einzelgen-Mutanten zeigten keinen offensichtlichen Phänotyp, was auf eine funktionelle Redundanz der Gene hinweist. Daher wurden zwei Doppelmutanten *(rkd1-1/rkd2-2 und rkd1-3/rkd2-2)* erzeugt und analysiert. Sie weisen eine Reduktion des Samenansatzes von 40% auf. Dies wird wahrscheinlich durch eine Unterbrechung der Gametophytentwicklung auf dem Stadium der funktionellen Megaspore verursacht. Genotypisierung segregierender Nachkommen zeigen aber, dass die Mutationen und der Phänotyp nicht gekoppelt sind. Weiterführende Arbeiten dazu sind erforderlich.

Die kodierenden Regionen der Gene *AtRKD1,2,3,4* wurden unter Kontrolle des *CaMV35S* Promotors ektopisch exprimiert. Die Überexpression der Gene *AtRKD1* und *AtRKD2* führt zu schwerwiegenden Wachstumsstörungen und kallusartigen Proliferation. In *AtRKD3*-Überexpressionlinien ist der Samenansatz geringfügig reduziert, während *AtRKD4*-

Überexpressionslinien sich vom Wildtyp nicht unterscheiden. Das AtRKD1- and AtRKD2-induzierte proliferierende Gewebe kann in farblose, undifferenzierte und differenzierte, grüne Abschnitte unterteilt werden. Ähnlich proliferierendes Gewebe entsteht bei der Überexpession von *AtRKD1*::*GFP* und *AtRKD2*::*GFP* Fusionskonstrukten, wobei deren Expression auf Proteinebene verfolgt werden kann. Beide Fusionsproteine werden nur im nichtdifferenzierten, farblosen, proliferierenden Gewebe nachgewiesen, sie fehlen im differenzierten grünen Teil des Kallus.

cDNA Arrays, RT-PCR und *in situ*-Hybridisierung belegen die Expression einer Reihe von gametophytischen Markergenen in dem AtRKD1- und AtRKD2-induzierten farblosen proliferierenden Gewebe. Dies führt zu der Hypothese, dass AtRKD1 und AtRKD2 einen gametophytischen Entwicklungsweg kontrollieren.

Wenngleich weitere Arbeiten erforderlich sind, belegen die verfügbaren Ergebnisse eine wesentliche Rolle von RKD-Proteinen für die Entwicklung des weiblichen Gametophyten.

6. *REFERENCES*

Acosta-Garcia, G. and Vielle-Calzada, J.P. (2004). A classical ARABINOGALACTAN protein is essential for the initiation of female gametogenesis in *Arabidopsis*. Plant Cell **16,** 2614-2628.

Ahn, J.H., Kim, J., Yoo, S.J., Yoo, S.Y., Roh, H., Choi, J.H., Choi, M.S., Chung, K.S., Han, E.J., Hong, S.M., Jung, S.H., Kang, H.J., Kim, B.K., Kim, M.D., Kim, Y.K., Kim, Y.H., Lee, H., Park, S.H., Yang, J.H., Yang, J.W., Yoo, D.H., Yoo, S.K. and Lee, S.J. (2007). Isolation of 151 mutants that have developmental defects from T-DNA tagging. Plant Cell Physiol. **48,** 169–178.

Alexander, M.P. (1969). Differential staining of aborted and nonaborted pollen. Stain Technol. **44,** 117–122.

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

Altpeter, F., Vasil, V., Srivastava, V., Stöger, E. and Vasil, I. K. (1996). Accelerated production of transgenic wheat (*Triticum aestivum*, L.) plants. Plant Cell Rep. **16**, 12–17.

Bevan, M.W. (1984). Binary *Agrobacterium* vectors for plant transformation. Nucleic Acids Res. **12,** 8711-8721.

Bicknell, R.A. and Koltunow, A.M. (2004). Understanding apomixis: recent advances and remaining conundrums. Plant Cell **16 Suppl,** S228-S245.

Bicknell, R.A., Borst, N.K. and Koltunow, A.M. (2000). Monogenic inheritance of apomixis in two *Hieracium* species with distinct developmental mechanisms. Heredity **84,** 228–237.

Birky, C.W. (1995). Uniparental inheritance of mitochondrial and chloroplast genes: mechanisms and evolution. Proc. Natl. Acad. Sci. **92,** 11331-11338.

Boutilier, K., Offringa, R., Sharma, V.K., Kieft, H., Ouellet, T., Zhang, L., Hattori, J., Liu, C.M., van Lammeren, A.A.M., Miki, B.L.A., Custers, J.B.M. and Campagne M.M.L. (2002). Ectopic expression of *BABY BOOM* triggers a conversion from vegetative to embryonic growth. Plant Cell **14,** 1737–1749.

Brunaud, V., Balzergue, S., Dubreucq, B., Aubourg, S., Samson, F., Chauvin, S., Bechtold, N., Cruaud, C., DeRose, R., Pelletier, G., Lepiniec, L., Caboche, M. and Lecharny A. (2002). T-DNA integration into the Arabidopsis genome depends on sequences of pre-insertion sites. EMBO Rep. **3,** 1152-1157.

Carman, J.G. (1997). Asynchronous expression of duplicate genes in angiosperms may cause apomixis, bispory, tetraspory, and polyembryony. Biol. J. Linn. Soc. **61,** 51-94.

Chaudhury, A.M., Craig, S., Dennis, E.S. and Peacock, W.J. (1998). Ovule and embryo development, apomixis and fertilization. Curr. Opin. Plant Biol. **1,** 26–31.

Chen, H., Nelson, R.S. and Sherwood J.L. (1994). Enhanced recovery of transformants of *Agrobacterium tumefaciens* after freeze-thaw transformation and drug selection. BioTechniques **16,** 664-668.

Choi, Y., Gehring, M., Johnson, L., Hannon, M., Harada, J.J., Goldberg, R.B., Jacobsen, S.E. and Fischer, R.L. (2002). DEMETER, a DNA glycosylase domain protein, is required for endosperm gene imprinting and seed viability in *Arabidopsis*. Cell **110,** 33–42.

Christensen, C.A., Subramanian, S., and Drews, G.N. (1998). Identification of gametophytic mutations affecting female gametophyte development in *Arabidopsis*. Developmental Biology **202,** 136-151.

Christensen, C.A., Gorsich, S.W., Brown, R.H., Jones, L.G., Brown, J., Shaw, J.M. and Drews, G.N. (2002). Mitochondrial *GFA2* is required for synergid cell death in *Arabidopsis*. Plant Cell **14,** 2215–2232.

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

Cordts, S., Bantin, J., Wittich, P.E., Kranz, E., Lorz, H. and Dresselhaus, T. (2001). *ZmES* genes encode peptides with structural homology to defensins and are specifically expressed in the female gametophyte of maize. Plant J. **25,** 103–114.

Crane, C.F. (2001). Classification of apomictic mechanisms. In Flowering of Apomixis: From Mechanisms to Genetic Engineering, Y. Savidan, J.G. Carman, and T. Dresselhaus, eds (Mexico: CIMMYT, IRD, European Commission DG VI), pp. 24–34.

Curtis, M. and Grossniklaus, U. (2003). A Gateway™ cloning vector set for high-throughput functional analysis of genes in plants. Plant Phys. **133,** 462-469.

Drews, G.N., Lee, D. and Christensen, G.A. (1998). Genetic analysis of female gametophyte development and function. Plant Cell **10,** 5–17.

Ebel, C., Mariconti, L. and Gruissem, W. (2004) Plant *RETINOBLASTOMA* homologues control nuclear proliferation in the female gametophyte. Nature **429,** 776-780.

Escobar-Restrepo, J.M., Huck, N., Kessler, S., Gagliardini, V., Gheyselinck, V., Yang, W.C. and Grossniklaus, U. (2007). The FERONIA receptor-like kinase mediates male-female interactions during pollen tube reception. Science **317,** 656 – 660.

Evans, M.M. (2007). The *indeterminate gametophyte 1* gene of maize encodes a LOB domain protein required for embryo sac and leaf development. Plant Cell **19,** 46-62.

Ferris, P.J. and Goodenough, U.W. (1997). Mating type in *Chlamydomonas* is specified by mid, the *minus-dominance* gene. Genetics **146,** 859-869.

Friedman, W.E. (2001). Developmental and evolutionary hypotheses for the origin of double fertilization and endosperm. Comptes rendus de l'Academie des sciences **324,** 559-567.

Gasser, C.S., Broadhvest, J. and Hauser, B.A. (1998). Genetic analysis of ovule development. Annu. Rev. Plant Physiol. Plant Mol. Biol. **49,** 1–24.

Gehring, M., Huh, J.H., Hsieh, T.F., Penterman, J., Choi, Y., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (2006). DEMETER DNA glycosylase establishes *MEDEA* polycomb gene self-imprinting by allele-specific demethylation. Cell **124,** 495–506.

Grimanelli, D., Leblanc, O., Espinosa, E., Perotti, E., De Leon, D.G. and Savidan, Y. (1998). Mapping diplosporous apomixis in tetraploid *Tripsacum*: One gene or several genes? Heredity **80,** 33–39.

Groß-Hardt, R., Kägi, C., Baumann, N., Moore, J.M., Baskar, R., Gagliano, W.B., Jürgens, G. and Grossniklaus, U. (2007). *LACHESIS* restricts gametic cell fate in the female gametophyte of *Arabidopsis*. PLoS Biology **5,** 494-500.

Grossniklaus, U. and Schneitz, K. (1998). The molecular and genetic basis of ovule and megagametophyte development. Semin. Cell Dev. Biol. **9,** 227–238.

Grossniklaus, U., Vielle-Calzada, J.P., Hoeppner, M.A. and Gagliano, W.B. (1998). Maternal control of embryogenesis by *MEDEA*, a Polycomb group gene in *Arabidopsis*. Science **280,** 446–450.

Gutierrez-Marcos J.F., Costa L.M. and Evans M.M.S. (2006). Maternal gametophytic baseless1 is required for development of the central cell and early endosperm patterning in maize (*Zea mays*). Genetics **174,** 317 - 329.

Haecker, A., Groß-Hardt, R., Geiges, B., Sarkar, A., Breuninger, H., Herrmann M. and Laux, T. (2004). Expression dynamics of *WOX* genes mark cell fate decisions during early embryonic patterning in *Arabidopsis thaliana*. Development **131,** 657–668.

Haig, D. and Westoby, M. (1989). Parent specific gene expression and the triploid endosperm. Nature **134,** 147-155.

Harding, E.W., Tang, W.N., Nichols, K.W., Fernandez, D.E. and Perry, S.E. (2003). Expression and maintenance of embryogenic potential is enhanced through constitutive expression of *AGAMOUS-LIKE 15*. Plant Phys. **133,** 653–663.

Hejatko, J., Pernisova, M., Eneva, T., Palme, K. and Brzobohaty, B., (2003). The putative sensor histidine kinase CKI1 is involved in female gametophyte development in *Arabidopsis*. Mol. Genet. Genomics **269,** 443-453.

Higashiyama, T., Yabe, S., Sasaki, N., Nishimura, Y., Miyagishima, S., Kuroiwa, H. and Kuroiwa, T. (2001). Pollen tube attraction by the synergid cell. Science **293,** 1480–1483.

Hobe, M., Müller, R., Grünewald, M., Brand, U. and Simonet, R. (2003). Loss of CLE40, a protein functionally equivalent to the stem cell restricting signal CLV3, enhances root waving in *Arabidopsis*. Dev. Genes Evol. **213,** 371–381.

Hofmeister, W., (1851). Vergleichende Untersuchungen der Keimung. Entfaltung und Fruchtbildung höherer Kryptogamen (Moose, Farrn, Equisetaceen, Rhizocarpeen und Lycopodiaceen) und die Samenbildung der Coniferen. Leipzig, Germany.

Honys, D. and Twell, D. (2003). Comparative analysis of the *Arabidopsis* pollen transcriptome. Plant Phys. **132,** 640-652.

Huanca-Mamani, W., Garcia-Aguilar, M., Leon-Martinez, G., Grossniklaus, U. and Vielle-Calzada, J.P. (2005). CHR11, a chromatin-remodeling factor essential for nuclear proliferation during female gametogenesis in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **102,** 17231–17236.

Huck, N., Moore, J.M., Federer, M. and Grossniklaus, U. (2003). The *Arabidopsis* mutant *feronia* disrupts the female gametophytic control of pollen tube reception. Development **130,** 2149-2159.

Jefferson, A.J. (1987). Assaying chimaeric genes in plants: The GUS gene fusion system. Plant Mol. Biol. Rep. **5**, 387-405.

Jessup, R.W., Burson, B.L., Burow, G.B., Wang, Y.W., Chang, C., Li, Z., Paterson, A.H. and Hussey, M.A. (2002). Disomic inheritance, suppressed recombination, and allelic interactions govern apospory in buffelgrass as revealed by genome mapping. Crop Sci. **42,** 1688–1694.

Johnston, A.J., Meier, P., Gheyselinck, J., Wuest. S.E., Federer, M., Schlagenhauf. E., Becker, J.D. and Grossniklaus, U., (2007). Genetic substraction profiling identifies genes essential for *Arabidopsis* reproduction and reveals interaction between the female gametophyte and the maternal sporophyte. Genome Biol., in press

Jones, H.D., Doherty, A. and Wu, H. (2005). Review of methodologies and a protocol for the *Agrobacterium*-mediated transformation of wheat. Plant Methods **1,** doi: 10.1186/1746-4811-1-5.

Kasahara, R.D., Portereiko, M.F., Sandaklie-Nikolova, L., Rabiger, D.S. and Drews, G.N. (2005). *MYB98* is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. Plant Cell **17,** 2981-2992.

Kerk, N.M., Ceserani, T., Tausta, S.L., Sussex, I.M. and Nelson, T.M. (2003). Laser capture microdissection of cells from plant tissues. Plant Phys. **132,** 27-35.

Kim, H.U., Li, Y. and Huang, A.H. (2005). Ubiquitous and endoplasmic reticulum-located lysophosphatidyl acyltransferase, LPAT2, is essential for female but not male gametophyte development in *Arabidopsis*. Plant Cell **17,** 1073-1089.

Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. Science **303,** 521-523.

Koltunow, A.M. (1993). Apomixis: Embryo sacs and embryos formed without meiosis or fertilization in ovules. Plant Cell **5,** 1425–1437.

Koltunow, A.M. and Grossniklaus, U. (2003). Apomixis: A developmental perspective. Annu. Rev. Plant Biol. **54,** 547–574.

Koltunow, A.M., Truettner, J., Cox, K.H., Wallroth, M. and Goldberg, R.B. (1990). Different temporal and spatial gene expression patterns occur during anther development. Plant Cell **2**, 1201–1224.

Köhler, C. and Grossniklaus, U. (2005). Seed development and genomic imprinting in plants. In Progress in Molecular and Subcellular Biology: Epigenetics and Chromatin (P. Jeanteur, ed), Springer Verlag, Berlin, Germany

Köhler, C., Hennig, L., Spillane, C., Pien, S., Gruissem, W. and Grossniklaus, U. (2003). The Polycomb-group protein MEDEA regulates seed development by controlling expression of the MADS-box gene *PHERES1*. Genes Dev. **17,** 1540–1553.

Kumlehn, J., Lörz, H. and Kranz, E. (1998). Differentiation of isolated wheat zygotes into embryos and normal plants. Planta **205,** 327–333.

Kumlehn, J., Kirik, V., Czihal, A., Altschmied, L., Matzk, F., Lörz, H. and Bäumlein, H. (2001). Parthenogenetic egg cells of wheat: cellular and molecular studies. Sex. Plant Reprod. **14,** 239–243.

Kumlehn, J., Serazetdinova, L., Hensel, G., Becker, D. and Lörz, H. (2006). Genetic transformation of barley (*Hordeum vulgare* L.) via infection of androgenetic pollen cultures with *Agrobacterium tumefaciens*. Plant Biotechnology Journal **4**, 251-261.

Kwee, H.S. and Sundaresan, V. (2003). The *NOMEGA* gene required for female gametophyte development encodes the putative APC6/CDC16 component of the Anaphase Promoting Complex in *Arabidopsis*. Plant J. **36,** 853-866.

Laux, T. (2003). The stem cell concept in plants: a matter of debate. Cell **113,** 281–283.

Lê, Q., Gutierrez-Marcos, J.F., Costa, L.M., Meyer, S., Dickinson, H.G., Lorz, H., Kranz, E. and Scholten, S. (2005). Construction and screening of subtracted cDNA libraries from limited populations of plant cells: a comparative analysis of gene expression between maize egg cells and central cells. Plant J. **44,** 167-178.

Leblanc, O., Grimanelli, D., Gonzalez de Leon, D. and Savidan, Y. (1995). Detection of the apomictic mode of reproduction in maize-*Tripsacum* hybrids using maize RFLP markers. Theor. Appl. Genet. **90,** 1198–1203.

Lenhard, M. and Laux, T. (2003). Stem cell homeostasis in the *Arabidopsis* shoot meristem is regulated by intercellular movement of CLAVATA3 and its sequestration by CLAVATA1. Development **130,** 3163-3173.

León, G., Holuigue, L. and Jordana, X. (2007). Mitochondrial complex II is essential for gametophyte development in *Arabidopsis*. Plant Phys. **143,** 1534–1546.

Li, J.J. and Herskowitz, I. (1993). Isolation of ORC6, a component of the yeast origin recognition complex by a one-hybrid system. Science **262,** 1870-1874.

Lin, H. and Goodenough, U.W. (2007). Gametogenesis in the *Chlamydomonas reinhardtii minus* mating type is controlled by two genes, *MID* and *MTD1*. Genetics **176,** 913–925.

Lotan, T., Ohto, M., Matsudaira Yee, K., West, M.A.L., Lo, R., Kwong, R.W., Yamagishi, K., Fischer, R.L., Goldberg, R.B. and Harada, J.J. (1998). *Arabidopsis* LEAFY COTYLEDON1 is sufficient to induce embryo development in vegetative cells. Cell **93,** 1195–1205.

Luo, M., Bilodeau, P., Dennis, E.S., Peacock, W.J. and Chaudhury, A. (2000). Expression and parent-of-origin effects for *FIS2*, *MEA* and *FIE* in the endosperm and embryo of developing *Arabidopsis* seeds. Proc. Natl. Acad. Sci. **97,** 10637–10642.

Márton, M., Cordts, S., Broadhvest, J. and Dresselhaus, T. (2005). Micropylar pollen tube guidance by *EGG APPARATUS 1* of maize. Science **307,** 573-576.

Matzk, F. (1996). The `Salmon system`of wheat - a suitable model for apomixis research. Heriditas **125,** 299-301.

Matzk, F., Prodanovic, S., Bäumlein, H. and Schubert, I. (2005). The Inheritance of apomixis in *Poa pratensis* confirms a five locus model with differences in gene expressivity and penetrance. Plant Cell **17,** 13-24.

Mayer, K.F., Schoof, H., Haecker, A., Lenhard, M., Jürgens, G. and Laux, T. (1998). Role of WUSCHEL in regulating stem cell fate in the *Arabidopsis* shoot meristem. Cell **95,** 805–815.

Mogie, M. (1992). The Evolution of Asexual Reproduction in Plants. (London: Chapman and Hall).

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

Niewiadomski, P., Knappe, S., Geimer, S., Fischer, K., Schulz, B., Unte, U.S., Rosso, M.G., Ache, P., Flugge, U.I. and Schneider, A. (2005). The *Arabidopsis* plastidic glucose 6-phosphate/phosphate translocator GPT1 is essential for pollen maturation and embryo sac development. Plant Cell **17,** 760–775.

Ning, J., Peng, X.B., Qu, L.H., Xin, H.P., Yan, T.T. and Sun, M. (2006). Differential gene expression in egg cells and zygotes suggests that the transcriptome is restructed before the first zygotic division in tobacco. FEBS Lett. **580,** 1747-1752.

Nogler, G.A. (1984). Genetics of apospory in apomictic Ranunculus auricomus. V. Conclusion. Bot. Helv. **94,** 411–422.

Noyes, R.D. (2000). Diplospory and parthenogenesis in sexual x agamospermous (apomictic) *Erigeron* (*Asteraceae*) hybrids. Int. J. Plant Sci. **161,** 1–12.

Noyes, R.D. and Rieseberg, L.H. (2000). Two independent loci control agamospermy (apomixis) in the triploid flowering plant *Erigeron annuus*. Genetics **155,** 379–390.

Ohad, N., Yadegari, R., Margossian, L., Hannon, M., Michaeli, D., Harada, J.J., Goldberg, R.B. and Fischer, R.L. (1999). Mutations in *FIE*, a WD Polycomb group gene, allow endosperm development without fertilization. Plant Cell **11,** 407–415.

Page, D.R. and Grossniklaus, U. (2002). The art and design of genetic screens: *Arabidopsis thaliana.* Nat. Rev. Genet. **3,** 124–136.

Pagnussat, G.C., Yu, H.J., Ng, Q.A., Rajani, S., Mayalagu, S., Johnson, C.S., Capron, A., Xie, L.F., Ye, D. and Sundaresan, V. (2005). Genetic and molecular identification of genes required for female gametophyte development and function in *Arabidopsis*. Development **132,** 603-614.

Paz-Ares, J. and consortium TR (2002). REGIA, an EU project on functional genomics of transcription factors from *Arabidopsis thaliana*. Comp. Funct. Genom **3,** 102–108.

Peleman, J., Boerjan, W., Engler, G., Seurinck, J., Botterman, J., Alliotte, T., Van Montagu, M. and Inzé, D. (1989). Strong cellular preference in the expression of a housekeeping gene of *Arabidopsis thaliana* encoding *S*-adenosylmethionine synthetase. Plant Cell **1,** 81–93.

Pischke, M.S., Jones, L.G., Otsuga, D., Fernandez, D.E., Drews, G.N. and Sussman, M.R. (2002). An *Arabidopsis* histidine kinase is essential for megagametogenesis. Proc. Natl. Acad. Sci. USA **99,** 15800–15805.

Portereiko, M.F., Lloyd, A., Steffen, J.G., Punwani, J.A., Otsuga, D. and Drews, G.N. (2006a). AGL80 is required for central cell and endosperm development in *Arabidopsis*. Plant Cell **18,** 1862–1872.

Portereiko, M.F., Sandaklie-Nikolova, L., Lloyd, A., Dever, C.A., Otsuga, D. and Drews, G.N. (2006b). *NUCLEAR FUSION DEFECTIVE 1* encodes the *Arabidopsis* RPL21 M protein and is required for karyogamy during female gametophyte development and fertilization. Plant Phys. **141,** 957–965.

Reidt, W., Wohlfarth1, T., Ellerström, M., Czihal, A., Tewes, A., Ezcurra, I., Rask, L. and Bäumlein H. (2000). Gene regulation during late embryogenesis: the RY motif of maturation-specific gene promoters is a direct target of the FUS3 gene product. Plant J. **21,** 401-408.

Richert, J., Kranz, E., Lörz, H. and Dresselhaus, T. (1996). A reverse transcriptase-polymerase chain reaction assay for gene expression studies at the single cell level. Plant Sci. **114,** 93-99.

Roche, D., Chen, Z.B., Hanna, W.W. and Ozias-Akins, P. (2001). Non-Mendelian transmission of an apospory-specific genomic region in a reciprocal cross between sexual pearl millet (*Pennisetum glaucum*) and an apomictic F1 (*P. glaucum* x *P. squamulatum*). Sex. Plant Reprod. **13,** 217–223.

Rosso, M.G., Li, Y., Strizhov, N., Reiss, B., Dekker, K. and Weisshaar, B. (2003). An *Arabidopsis thaliana* T-DNA mutagenized population (GABI-Kat) for flanking sequence tag-based reverse genetics. Plant Mol. Biol. **53,** 247-259.

Rotman, N., Rozier, F., Boavida, L., Dumas, C., Berger, F. and Faure, J.E. (2003). Female control of male gamete delivery during fertilization in *Arabidopsis thaliana*. Curr. Biol. **13,** 432–436.

Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989). Molecular Cloning: a Laboratory Manual, 2nd edn. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.

Sager, R. and Granick, S. (1954). Nutritional control of sexuality in *Chlamydomonas reinhardtii*. J. Gen. Physiol. **37,** 729-742.

Savidan, Y. (1983). Genetics and utilization of apomixis for the improvement of guineagrass (*Panicum maximum Jacq*.). In Proceedings of the 14th International Grasslands Congress (1981), Lexington, KY, J.A. Smith and V.W. Hayes, eds (Boulder, CO: Westview Press), pp. 182–184.

Schauser, L., Roussis, A., Stiller, J. and Stougaard, J. (1999). A plant regulator controlling development of symbiotic root nodules. Nature **402,** 191-195.

Schauser, L., Wieloch, W. and Stougaard, J. (2005). Evolution of NIN-like proteins in *Arabidopsis*, rice, and *Lotus japonicus*. J. Mol. Evol. **60,** 229-237.

Schiefthaler, U., Balasubramanian, S., Sieber, P., Chevalier, D., Wisman, E. and Schneitz, K. (1999). Molecular analysis of *NOZZLE*, a gene involved in pattern formation and early sporogenesis during sex organ development in *Arabidopsis thaliana*. Proc. Natl. Acad. Sci. USA **96,** 11664–11669.

Schmidt, E.D., Guzzo, F., Toonen, M.A. and de Vries, S.C. (1997). A leucine-rich repeat containing receptor-like kinase marks somatic plant cells competent to form embryos. Development **124,** 2049-2062.

Schneitz, K. (1999). The molecular and genetic control of ovule development. Curr. Opin. Plant Biol. **2,** 13–17.

Schneitz, K., Balasubramanian, S. and Schiefthaler, U. (1998). Organogenesis in plants: The molecular and genetic control of ovule development. Trends Plant Sci. **3,** 468–472.

Schroeter, E. H., Kisslinger, J. A. and Kopan, R. (1998). Notch-1 signalling requires ligand-induced proteolytic release of intracellular domain. Nature **393,** 382-386.

Schoof, H., Lenhard, M., Haecker, A., Mayer, K.F., Jürgens, G. and Laux, T. (2000). The stem cell population of *Arabidopsis* shoot meristems in maintained by a regulatory loop between the *CLAVATA* and *WUSCHEL* genes. Cell **100,** 635–644.

Sherwood, R.T., Berg, C.C. and Young, B.A. (1994). Inheritance of apospory in buffelgrass. Crop Sci. **34,** 1490–1494.

Shi, D.Q., Liu, J., Xiang, Y.H., Ye, D., Sundaresan, V. and Yang, W.C. (2005). *SLOW WALKER 1*, essential for gametogenesis in *Arabidopsis*, encodes a WD40 protein involved in 18S ribosomal RNA biogenesis. Plant Cell **17,** 2340-2354.

Siddiqi, I., Ganesh, G., Grossniklaus, U. and Subbiah, V. (2000). The *DYAD* gene is required for progression through female meiosis in *Arabidopsis*. Development **127,** 197-207.

Smith, J. (1841). Notice of a plant which produces seeds without any apparent action of pollen. Transactions of the Linnaean Society of London (meeting of June 18 1839).

Spillane, C., Curtis, M.D. and Grossniklaus, U. (2004). Apomixis technology development-virgin births in farmers' fields? Nature Biotech. **22,** 687-691.

Spillane, C., Steimer, A. and Grossniklaus, U. (2001). Apomixis in agriculture: The quest for clonal seeds. Sex. Plant Reprod. **14,** 179–187.

Sprunck, S., Baumann, U., Edwards, K., Langridge, P. and Dresselhaus, T. (2005). The transcript composition of egg cells changes significantly following fertilization in wheat (*Triticum aestivum* L.). Plant J. **41,** 660-672.

Steffen, J.G., Kang, I.H., Macfarlane, J. and Drews G.N. (2007). Identification of genes expressed in the *Arabidopsis* female gametophyte. Plant J. **51,** 281-292.

Stone, S.L., Kwong, L.W., Yee, K.M., Pelletier, J., Lepiniec, L., Fischer, R.L., Goldberg, R.B. and Harada J.J. (2001). *LEAFY COTYLEDON2* encodes a B3 domain transcription factor that induces embryo development. Proc. Natl. Acad. Sci. **98,** 11806 - 11811.

Valle, C.B., Glienke, C. and Leguizamon, G.O.C. (1994). Inheritance of apomixis in *Brachiaria*, a tropical forage grass. Apomixis Newsl. **7,** 42–43.

Válóczi A., Várallyay, É., Burgyán J. and Havelda Z. (2006). Spatiotemporal accumulation of microRNAs is highly coordinated in developing plant tissues. Plant J. **47,** 140-151.

van Dijk, P.J., Tas, I.C.Q., Falque, M. and Bakx-Schotman, T. (1999). Crosses between sexual and apomictic dandelions (*Taraxacum*). II. The breakdown of apomixis. Heredity **83,** 715–721.

Verdeil, J.L., Alemanno, L., Niemenak, N. and Tranbarger, T.J. (2007). Pluripotent versus totipotent plant stem cells: dependence versus autonomy? Trends Plant Sci. **12,** 245-252.

Vielle-Calzada, J.P., Nuccio, M.L., Budiman, M.A., Thomas, T.L., Burson, B.L., Hussey, M.A. and Wing, R.A. (1996). Comparative gene expression in sexual and apomictic ovaries of *Pennisetum ciliare* (L.) Link. Plant Mol. Biol. **32,** 1085–1092.

Vielle-Calzada, J.P., Thomas, J., Spillane, C., Coluccio, A., Hoeppner, M.A. and Grossniklaus, U. (1999). Maintenance of genomic imprinting at the *Arabidopsis MEDEA* locus requires zygotic *DDM1* activity. Genes Dev. **13,** 2971-2982.

Voigt, P.W. and Burson, B.L. (1983). Breeding of apomictic Eragrostis curvula. In Proceedings of the 14th International Grasslands Congress (1981), Lexington, KY, J.A. Smith and V.W. Hayes, eds (Boulder, CO: Westview Press), pp. 160–163.

Wang, M.B. and Waterhouse P.M. (1997). A rapid and simple method of assaying plants transformed with hygromycin or PPT resistance genes. Plant Molecular Biology Reporter **15,** 209–215.

Wu, M.F., Tian, Q. and Reed, J.W. (2007). *Arabidopsis microRNA167* controls patterns of ARF6 and ARF8 expression, and regulates both female and male reproduction. Development **133,** 4211-4218.

Yadegari, R. and Drews, G.N. (2004). Functional analysis of the female gametophyte. Plant Cell **16 Suppl,** 133-141.

Yang, H., Kaur, N.., Kiriakopolos, S. and McCormick, S. (2006). EST generation and analyses towards identifying female gametophyte-specific genes in *Zea mays* L. Planta, **224,** 1004-1014.

Yang, W., Jefferson, R.A., Huttner, E., Moore, J.J., Gagliano, W.B. and Grossniklaus, U. (2005). An egg apparatus-specific enhancer of *Arabidopsis*, identified by enhancer detection. Plant Phys. **139,** 1421–1432.

Yang, W.C., Ye, D., Xu, J. and Sundaresan, V. (1999). The *SPOROCYTELESS* gene form *Arabidopsis* is required for initiation of sporogenesis and encodes a novel nuclear protein. Genes Dev. **13,** 2108–2117.

Yoo, S.Y., Bomblies, K., Yoo, S.K., Yang, J.W., Choi, M.S., Lee, J.S., Weigel, D. and Ahn, J.H. (2005). The 35S promoter used in a selectable marker gene of a plant transformation vector affects the expression of the transgene. Planta **221,** 523–530.

Yu, H.J., Hogan, P. and Sundaresan, V. (2005). Analysis of the female gametophyte transcriptome of *Arabidopsis* by comparative expression profiling. Plant Phys. **139,** 1853-1869.

Zhao, H., Lu, M., Singh, R. and Snell, W.J. (2001). Ectopic expression of a *Chlamydomonas* mt+-specific homeodomain protein in mt− gametes initiates zygote development without gamete fusion. Genes Dev. **15,** 2767-2777.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R. and Siebert, P.D. (2001) Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. Biotechniques **30,** 892-897.

Zuo, J., Niu, Q.W., Frugis, G. and Chau, N.H. (2002). The *WUSCHEL* gene promotes vegetative-to embryonic transition in Arabidopsis. Plant J. **30,** 349–359.

Curriculum vitae

Personal data

Name: Kőszegi, Dávid

Place of birth: Budapest, Hungary

Date of birth: 20 February 1979

Nationality: Hungarian

Languages : Hungarian (native), English, German, Italian

Contact Addresses

Leibniz Institute of Plant Genetics and Crop Plant Research (IPK)

Corrensstrasse 3

D-06466 Gatersleben

E-mail: koszegi@ipk-gatersleben.de

Phone: ++49-39-482-5751

Fax: ++49-39-482-5139

Education and employment

2005-2007 Post graduate student at the Institute für Pflanzengenetik und Kulturpflanzenforschung, Gatersleben, Germany
Ph.D. thesis: *RKD* genes: a novel transcription factor family involved in the female gametophyte development of *Arabidopsis* and wheat

2002-2004 Agricultural Research Institute of the Hungarian Academy of Sciences, Martonvásár, Hungary

1997-2002 M.Sc. in Biology and Chemistry, Eötvös Loránd University, Budapest
Thesis title: Isolation of HMW storage protein genes from the 'Bánkúti 1201' wheat cultivar

Publication of results

Talks:

Koszegi, D.,: Towards the initiation of parthenogenesis *via* egg cell-specific expression of RepA or orf13 genes. – 1st ISC IPK student conference Gatersleben, 22.–25.06.2005.

Kumlehn, J., Hähnel, U., Koszegi, D., Czihal, A., Tiedemann, J., Altschmied, L. and Bäumlein, H.: Studies on autonomous embryo formation from Salmon wheat egg cells. – 19th International Congress of the Sexual Plant Reproduction (ICSPR), Budapest/Hungary, 11.–15.07.2006.

Koszegi, D., Czihal, A., Hähnel, U., Varshney, A., Tewes, A., Johnston, A., Kumlehn, J., Altschmied, L., Matzk, F., Grossniklaus, U. and Bäumlein, H.: Novel transcription factors involved in the gametophyte-sporophyte transition in wheat and *Arabidopsis*. 3rd International Apomixis Conference, Wernigerode, Germany, 27.06.-01.07.2007.

Posters:

Matzk, F., Prodanovic, S., Czihal, A., Tiedemann, J., Koszegi, D., Arzenton, F., Blattner, F., Kumlehn, J., Altschmied, L., Johnston, A., Grossniklaus, U. and Bäumlein, H.: Apomixis-preliminary data from wheat egg cells, *Poa pratensis, Hypericum perforatum* and *Arabidopsis*. – 12th International Conference on Plant Embryology, Cracow/Poland, 05.–07.09.2005.

Matzk, F., Prodanovic, S., Czihal, A., Tiedemann, J., Koszegi, D., Arzenton, F., Blattner, F., Kumlehn, J., Altschmied, L., Johnston, A., Grossniklaus, U. and Bäumlein, H.: Apomixis – preliminary data from *Poa pratensis, Hypericum*, wheat egg cells and *Arabidopsis*. Institutstag IPK, Gatersleben, 13.10.2005.

Hähnel, U., Johnston, A., Koszegi, D., Tiedemann, J., Czihal, A., Grossniklaus, U., Kumlehn, J., Bäumlein, H. and Altschmied, L.: Potential transcription factor genes from barley and *Arabidopsis* specifically expressed in the egg apparatus. – 19. Tagung 'Molekularbiologie der Pflanzen', Dabringhausen, 07.–10.03.2006.

Koszegi, D., Czihal, A., Tiedemann, J., Hähnel, U., Prodanovic, S., Arzenton, F., Johnston, A., Grossniklaus, U., Matzk, F., Blattner, F., Kumlehn, J., Schubert, I., Altschmied, L., and Bäumlein, H.: Apomixis – preliminary lessons from wheat egg cells, *Hypericum perforatum* and *Arabidopsis thaliana*. – 2nd ISC IPK student conference Gatersleben, 30.05.–01.06.2006.

Koszegi, D., Czihal, A., Tiedemann, J., Hähnel, U., Prodanovic, S., Arzenton, F., Johnston, A., Grossniklaus, U. Matzk, F., Blattner, F., Kumlehn, J., Altschmied, L., and Bäumlein, H.: Apomixis – preliminary lessons from wheat egg cells, *Hypericum perforatum* and *Arabidopsis thaliana*. – 19th International Congress on Sexual Plant Reproduction (ICSPR), Budapest/Hungary, 11.–15.07.2006.

Koszegi, D., Czihal, A., Hähnel, U., Varshney, A., Tewes, A., Johnston, A., Kumlehn, J., Altschmied, L., Matzk, F., Grossniklaus, U. and Bäumlein, H.: Novel transcription factors involved in the gametophyte-sporophyte transition in wheat and *Arabidopsis*. – 3rd Plant Science Student Conference, Halle 5-8.06.2007.

Acknowledgements

This work was done in the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) in Gatersleben. I would like to thank all the co-workers of the Institute who helped in this work, specially from the groups of Gene Regulation, Expression Mapping and Plant Reproductive Biology.

I am very grateful to my supervisor, Helmut Bäumlein and to my mentor Jochen Kumlehn for the support and for valuable discussions during the time of my Ph.D. studies.

I acknowledge the following people, who helped in different ways: Annegret Tewes and Sabine Skiebe for helping the transient expression assays; Alok Varshney for helping in wheat microdissection; Susanne König for DNA sequencing; Brigitte Schäfer for the high quality graphic and photographic work. I am especially thankful to Anett Busching, Elke Liemann, Monika Gottowik, Alexandra Rech, Erika Grützemann, Petra Hoffmeister, Susanne Knüpffer, Conny Marthe and Ingrid Otto for excellent technical assistance. I was very happy to work together with Isolde Saalbach and Götz Hensel on wheat transformation and analysis of the plants.

I would like to express my gratitude towards Amal Johnston and Ueli Grossniklaus from University of Zürich for the fruitful cooperation and for the *in situ* hybridisations.

My best thanks to Andreas Czihal for helping in hybridisation experiments and general advice about molecular work.

I am very much obliged to Timothy F. Sharbel and Jose Maria Corral Garcia for giving me the chance to get familiar with the GeXP system.

I am indebted to Twan Rutten who put a lot of effort in taking high quality microscopical images and hystological analysis.

I am thankful to Ulrich Wobus for giving me the oportunity to work in this appreciated institute.

Grateful acknowledgements are due to Lothar Altschmied, Helmut Bäumlein, Jochen Kumlehn and Gudrun Mönke for critical reading of the manuscript and for useful comments.

And the *RKD* genes happily lived ever after in the gametophyte…

ERKLÄRUNGEN

GEMÄß § 5 ABS. 2 DER PROMOTIONSORDNUNG VOM 17.09.1998

Ich erkläre hiermit, daß ich mich mit der vorliegenden wissenschaftlichen Arbeit erstmals um die Erlangung des Doktorgrades bewerbe. Ich habe die Arbeit selbständig und ohne fremde Hilfe verfaßt, und nur die von mir angegebenen Quellen und Hilfsmittel benutzt. Wörtlich oder inhaltlich entnommene aus anderen Werken habe ich Stellen als solche kenntlich gemacht.

Gatersleben, November 2007

Kőszegi, Dávid

8. *APPENDIX*

I. Results of cDNA array experiments

macroarray experiment from the AtRKD1 overexpression in *Arabidopsis*.
*the P80 plant was an escape plant (unsuccessfully transformed plant), served as a control
** WT1 and WT2 wild type plants, used as a negative controls; P85, P87 and P88 AtRKD1 overexpressing plants; P80 see above

Probe ID	GENE ID	Description	WT signal*			Mutant signal				Mean WT signal	Mean mutant signal	Fold change	Regulation
			WT1	WT2	P80	P78	P85	P87	P88				
D - 04 : 07 (a)	At5g13080	WRKY-like protein	0,61	1,87	0,83	4,24	28,62	15,17	32,56	1,10	20,15	18,26	up
D - 03 : 04 (a)	At1g21970	LEAFY COTYLEDON 1	0,56	0,48	0,17	5,03	5,96	3,47	7,72	0,40	5,55	13,75	up
A - 05 : 16 (a)	At1g02220	NAC domain containing protein 3	0,47	0,74	0,21	2,06	9,77	1,95	12,12	0,47	6,48	13,68	up
B - 05 : 13 (a)	At1g80730	zinc finger protein 54	0,44	0,43	0,33	4,45	6,48	3,19	5,70	0,40	4,96	12,39	up
B - 10 : 02 (a)	At3g66656	MADS-box family protein (MADS118)	0,53	0,44	1,32	7,49	11,55	7,57	9,87	0,76	9,12	11,95	up
E - 08 : 16 (a)	At1g29860	WRKY DNA-binding protein 71	0,91	0,70	0,52	3,68	7,38	12,65	10,06	0,71	8,44	11,89	up
C - 08 : 13 (a)	At2g01500	WUSCHEL related homeobox 6	0,66	0,55	0,31	2,91	4,66	4,12	2,96	0,51	3,66	7,23	up
F - 11 : 11 (a)	At1g77950	MADS-box family protein (MADS67)	0,47	0,78	0,22	3,57	3,88	3,40	2,54	0,49	3,35	6,83	up
D - 03 : 05 (a)	At4g18770	MYB domain protein 98	0,63	0,48	0,17	2,16	3,75	3,21	2,24	0,43	2,84	6,66	up

C - 04 : 10 (a)	At1g54330	NAC domain containing protein 20	0,50	0,52	1,89	7,52	7,76	5,10	5,12	0,97	6,38	6,57	up
A - 10 : 09 (a)	At5g15150	HOMEOBOX PROTEIN 3	0,50	0,43	0,38	2,31	4,21	1,64	2,48	0,44	2,66	6,09	up
A - 12 : 08 (a)	At4g01720	WRKY DNA-binding protein 47	0,48	0,46	0,91	5,29	6,01	0,89	2,30	0,62	3,62	5,87	up
A - 04 : 13 (a)	At3g57920	squamosa promoter-binding protein	0,55	0,45	0,66	3,16	4,51	1,60	3,67	0,55	3,24	5,85	up
B - 06 : 13 (a)	At2g26150	Heat stress transcription factor A-2	1,77	0,59	0,28	4,20	6,10	3,69	6,52	0,88	5,13	5,83	up
D - 08 : 11 (a)	At2g40340	DREB subfamily A-2 of ERF/AP2 transcription factor family	0,52	0,60	0,56	2,42	4,42	1,77	3,87	0,56	3,12	5,57	up
E - 08 : 13 (a)	At5g45980	WUSCHEL related homeobox 8	0,91	0,68	0,52	3,03	4,45	2,81	4,18	0,70	3,62	5,14	up
F - 06 : 03 (a)	At4g14550	Aux/IAA protein 14 (solitary root)	1,13	0,49	0,79	2,95	6,14	3,91	3,49	0,80	4,12	5,13	up
E - 01 : 16 (a)	At4g00050	basic helix-loop-helix (bHLH) family protein 16, UNFERTILISED EMBRYO SAC 10	0,64	0,86	0,33	2,12	4,42	2,62	2,97	0,61	3,03	4,97	up
C - 09 : 12 (a)	At1g28460	MADS-box family protein (MADS59)	0,96	0,54	1,44	5,19	5,61	4,23	4,25	0,98	4,82	4,92	up
D - 04 : 12 (a)	At3g04100	MADS-box family protein (MADS57)	0,61	0,40	0,19	1,83	2,51	1,50	2,00	0,40	1,96	4,90	up
E - 03 : 09 (a)	At4g18170	WRKY DNA-binding protein 28	1,21	3,88	2	17,96	10,73	7,01	9,59	2,36	11,32	4,79	up
B - 09 : 13 (a)	At5g26930	zinc finger (GATA type) family protein 22	0,58	1,30	0,79	4,21	5,08	3,04	4,48	0,89	4,20	4,72	up

E - 11 : 05 (a)	At1g30490	PHAVOLUTA, HD-Zip protein	0,52	1,60	1,36	4,69	7,37	5,41	4,39	1,16	5,47	4,71	up
A - 09 : 14 (a)	At1g08010	zinc finger (GATA type) family protein 11	1,98	2,36	1,92	7,11	12,57	8,80	10,07	2,09	9,64	4,62	up
D - 08 : 13 (a)	At5g59340	WUSCHEL related homeobox 2	0,52	0,60	0,39	2,01	2,28	2,62	2,38	0,50	2,32	4,61	up
A - 10 : 16 (a)	At2g28340	zinc finger (GATA type) family protein 7	0,50	3,63	3,28	9,62	16,67	7,98	10,23	2,47	11,13	4,50	up
E - 09 : 11 (a)	At2g40950	Basic leucine zipper 17	0,87	0,52	1,19	3,07	6,28	2,14	3,79	0,86	3,82	4,44	up
C - 08 : 05 (a)	At2g36080	DNA-binding protein, putative	1,13	1,92	1,03	4,33	7,64	4,37	7,75	1,36	6,02	4,43	up
A - 01 : 09 (a)	At1g62300	WRKY DNA-binding protein 6	7,68	3,91	7,29	34,26	21,53	33,51	21,44	6,29	27,69	4,40	up
D - 05 : 11 (a)	At5g53950	CUP-SHAPED COTYLEDON 2	0,59	0,45	0,23	0,38	2,79	1,76	2,20	0,42	1,78	4,21	up
C - 02 : 13 (a)	At5g01380	similar to DNA-binding protein-related	0,59	1,39	1,39	4,71	6,59	3,55	3,97	1,12	4,71	4,19	up
A - 08 : 14 (a)	At1g62990	homeodomain transcription factor (KNAT7)	0,35	0,47	0,70	1,59	2,62	1,32	2,72	0,51	2,06	4,07	up
F - 09 : 12 (a)	At3g54340	Floral homeotic protein APETALA3	0,49	0,91	0,66	1,65	2,77	1,95	4,71	0,69	2,77	4,03	up
B - 01 : 15 (a)	At1g68810	basic helix-loop-helix (bHLH) family protein 27	0,44	0,45	2,02	0,85	6,31	2,23	5,89	0,97	3,82	3,94	up
F - 07 : 04 (a)	At3g24650	Abscisic acid-insensitive protein 3	0,89	0,51	0,99	1,81	3,57	3,90	2,69	0,80	2,99	3,76	up
A - 07 : 13 (a)	At1g02030	zinc finger (C2H2 type) family protein	0,57	0,49	0,45	1,16	2,21	1,51	2,00	0,50	1,72	3,42	up

F - 10 : 03 (a)	At3g28910	MYB domain protein 30	1,85	2,82	14,46	20,50	19,30	21,28	20,84	6,38	20,48	3,21	up
A - 02 : 04 (a)	At5g61860	Floricaula/leafy protein	0,69	0,64	1,24	2,55	2,43	3,28	2,23	0,86	2,62	3,06	up
D - 12 : 13 (a)	At3g23250	MYB domain protein 15	2,23	0,72	4,25	3,87	13,36	4,04	7,23	2,40	7,13	2,97	up
D - 11 : 11 (a)	At1g34190	NAC domain containing protein 17	0,51	1,18	2,75	5,19	5,46	1,52	4,51	1,48	4,17	2,82	up
E - 04 : 15 (a)	At2g43000	NAC domain containing protein 42	0,54	1,09	1,31	1,74	4,86	0,84	3,55	0,98	2,75	2,80	up
A - 07 : 14 (a)	At2g45050	zinc finger (GATA type) family protein 2	0,57	0,49	2,46	3,11	2,63	1,86	4,74	1,17	3,09	2,63	up
E - 07 : 13 (a)	At2g37430	zinc finger (C2H2 type) family protein 64	1,53	0,91	6,16	2,37	12,60	5,09	9,94	2,87	7,50	2,62	up
B - 01 : 09 (a)	At4g24240	WRKY DNA-binding protein 7	0,92	1,66	3,14	2,64	4,97	4,79	6,78	1,91	4,80	2,51	up
D - 03 : 08 (a)	At5g13330	member of the ERF (ethylene response factor) subfamily B-4 of ERF/AP2 transcription factor family	3,67	7,15	0,17	3,85	15,88	2,27	12,95	3,66	8,74	2,39	up
B - 02 : 13 (a)	At1g76890	plant trihelix DNA-binding protein (GT2)	0,97	0,98	6,40	3,34	12,61	4,65	4,33	2,78	6,23	2,24	up
B - 08 : 09 (a)	At2g22430	homeodomain leucine zipper class (HD-Zip I) protein, HOMEOBOX PROTEIN 6	274,30	174,60	87,00	76,16	38,25	163,10	44,52	178,63	80,51	0,45	down
F - 10 : 13 (a)	At1g22640	MYB domain protein 3	65,30	40,82	12,82	12,56	8,24	36,99	12,08	39,65	17,47	0,44	down
C - 07 : 03 (a)	At4g22950	AGAMOUS-LIKE 19	24,49	18,85	5,02	9,21	1,92	13,35	1,96	16,12	6,61	0,41	down

F - 11 : 15 (a)	At2g04240	Encodes a small protein with an N-terminal trans-membrane domain and a RING-H2 zinc finger motif located at the C-terminus	30,71	41,85	4,55	12,62	4,53	18,31	6,09	25,70	10,39	0,40	down
A - 05 : 14 (a)	At1g24625	ZINC FINGER PROTEIN 7	39,50	29,14	6,47	13,21	6,48	13,33	6,74	25,04	9,94	0,40	down
B - 05 : 03 (a)	At4g22950	AGAMOUS-LIKE 19	21,68	15,34	4,64	7,48	1,53	11,19	1,48	13,89	5,42	0,39	down
F - 09 : 01 (a)	At2g46870	similar to NGA2 (NGATHA2)	27,47	19,35	6,04	4,67	5,33	13,06	4,02	17,62	6,77	0,38	down
B - 07 : 11 (a)	At1g52880	NAC domain containing protein 18	68,64	47,34	19,14	30,45	12,06	13,19	13,19	45,04	17,22	0,38	down
A - 01 : 04 (a)	At2g34720	CCAAT-binding transcription factor (CBF-B/NF-YA) family protein	47,15	41,12	24,72	27,85	0,70	26,60	1,94	37,66	14,27	0,38	down
A - 06 : 11 (a)	At2g35940	BEL1-like homeodomain 1	27,09	23,31	9,75	7,46	5,17	12,88	3,41	20,05	7,23	0,36	down
F - 08 : 05 (a)	At3g15210	member of the ERF (ethylene response factor) subfamily B-1 of ERF/AP2 transcription factor family (ATERF-4)	760,60	164,00	196,00	207,80	47,24	156,90	88,23	373,53	125,04	0,33	down
A - 02 : 16 (a)	At2g26580	Axial regulator YABBY5	16,65	10,05	12,57	8,48	1,51	3,62	3,45	13,09	4,27	0,33	down
E - 07 : 08 (a)	At2g45660	AGAMOUS-LIKE 20, AGL20, SOC1	161,10	99,85	64,99	16,91	17,95	88,21	16,84	108,65	34,98	0,32	down
E - 02 : 16 (a)	At4g18390	TCP transcription factor 2	9,71	16,53	2,12	2,37	1,74	5,49	2,47	9,45	3,02	0,32	down
E - 09 : 08 (a)	At2g47890	zinc finger (B-box type) family protein, CONSTANS-LIKE 13 (COL13)	12,60	15,65	6,27	4,06	4,42	2,32	2,9	11,51	3,43	0,30	down

F - 12 : 05 (a)	At2g31230	encodes a member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family (ERF 15)	20,86	11,28	9,59	4,54	1,37	9,55	0,96	13,91	4,11	0,30	down
B - 03 : 14 (a)	At5g53420	unknown protein	50,86	83,56	21,22	14,36	18,16	14,18	12,23	51,88	14,73	0,28	down
D - 07 : 13 (a)	At1g27730	Related to Cys2/His2-type zinc-finger proteins, SALT TOLERANCE ZINC FINGER, STZ	289,10	48,64	119,60	9,55	48,31	48,17	61,75	152,45	41,95	0,28	down
D - 11 : 14 (a)	At3g46620	zinc finger (C3HC4-type RING finger) family protein	22,94	3,16	11,18	1,88	1,97	5,40	4,26	12,43	3,38	0,27	down
F - 01 : 06 (a)	At4g36900	member of the DREB subfamily A-5 of ERF/AP2 transcription factor family	61,79	35,65	14,98	15,07	2,26	17,26	5,31	37,47	9,98	0,27	down
E - 08 : 05 (a)	At5g47220	ETHYLENE RESPONSIVE ELEMENT BINDING FACTOR 2	28,94	19,82	29,43	5,72	7,04	6,38	7,91	26,06	6,76	0,26	down
E - 12 : 13 (a)	At5g07580	member of the ERF (ethylene response factor) subfamily B-3 of ERF/AP2 transcription factor family	42,17	35,24	106,30	19,37	8,61	26,40	9,08	61,24	15,87	0,26	down
E - 10 : 12 (a)	At5g53420	unknown protein	40,64	77,24	24,58	10,92	15,73	10,90	9,46	47,49	11,75	0,25	down
F - 01 : 16 (a)	At4g00480	MYC-related protein with a basic helix-loop-helix motif at the C-terminus and a region similar to the maize B/R family at the N-terminus	58,07	6,52	19,01	2,79	1,37	11,65	9,82	27,87	6,41	0,23	down
B - 07 : 16 (a)	At2g23340	member of the DREB subfamily A-5 of ERF/AP2 transcription factor family	18,11	8,94	11,06	2,07	0,97	6,32	1,08	12,70	2,61	0,21	down
A - 03 : 02 (a)	At3g61970	NGA2 (NGATHA2)	10,73	6,94	1,80	1,56	0,16	1,63	1,69	6,49	1,26	0,19	down

A - 12 : 10 (a)	At3g61890	homeodomain leucine zipper class I (HD-Zip I) protein, HOMEOBOX PROTEIN 1	33,58	34,29	2,83	4,18	7,54	1,21	4,50	23,57	4,36	0,18	down
C - 10 : 02 (a)	At2g47190	MYB domain protein 2	25,79	19,61	0,73	0,37	1,88	4,56	3,05	15,38	2,47	0,16	down
C - 11 : 04 (a)	At5g04340	zinc finger (C2H2 type) family protein 6	93,62	25,49	13,11	3,92	3,81	11,30	8,88	44,07	6,98	0,16	down
C - 02 : 05 (a)	At5g25890	IAA-ALANINE RESISTANT 2	13,99	9,74	3,24	0,83	1,89	1,52	1,05	8,99	1,32	0,15	down
D - 10 : 10 (a)	At1g74930	member of the DREB subfamily A-5 of ERF/AP2 transcription factor family	61,98	3,96	29,96	3,52	1,68	9,73	2,63	31,97	4,39	0,14	down
C - 03 : 08 (a)	At1g46768	member of the DREB subfamily A-5 of ERF/AP2 transcription factor family	25,86	11,06	2,93	3,13	1,89	1,64	0,55	13,28	1,80	0,14	down
A - 01 : 08 (a)	At2g33720	unknown protein	205,80	40,35	37,58	25,71	1,13	13,75	2,63	94,58	10,81	0,11	down
C - 12 : 10 (a)	At1g18710	MYB domain protein 47	39,85	43,48	3,12	4,26	2,45	1,70	3,16	28,82	2,89	0,10	down
F - 01 : 10 (a)	At4g16750	member of the DREB subfamily A-4 of ERF/AP2 transcription factor family	17,83	10,49	1,94	1,83	0,11	1,03	0,39	10,09	0,84	0,08	down
A - 02 : 02 (a)	At5g60910	FRUITFULL AGAMOUS-LIKE 8	102,30	53,40	11,14	5,20	0,22	5,31	5,31	55,61	4,01	0,07	down
C - 12 : 02 (a)	At1g56650	MYB domain protein 75, production of ANTOCYANIN PIGMENT 1	118,30	114,00	0,67	8,17	2,29	6,05	1,87	77,66	4,60	0,06	down
A - 07 : 05 (a)	At1g48000	MYB domain protein 112	60,44	27,19	0,23	0,59	1,17	2,49	2,13	29,29	1,60	0,05	down
E - 07 : 04 (a)	At1g66390	MYB domain protein 90, production of ANTOCYANIN PIGMENT 2	31,43	38,43	0,97	0,23	0,30	1,00	0,27	23,61	0,45	0,02	down

II. Gene sequences

Bold letters are the coding regions, normal letters are introns, italic letters are putative untranslated regions. The amino acid sequence can be seen on top of the nucleic acid sequence in a one letter code. Stars indicate the STOP codons.

AtRKD1 (At1g18790) genomic sequence

```
          M  K  S  F  C  K  L  E  Y  D  Q  V  F  G  K  E  N  N  S  F
     1    ATGAAATCGTTTTGCAAGTTGGAGTATGATCAAGTGTTTGGCAAAGAAAATAATTCATTC

          S  F  L  N  H  S  S  L  Y  S  H  Q  S·
    61    TCATTTCTAAACCACTCATCACTTTACTCTCATCAAAGGTAAATTAATTGATCATGCATG

   121    ACTTTTTTTTAGTTTTACATGTTTATATAATAGTTCTTGATCCATGTTTTGTTGTTGTGT

                     ·S  E  L  A  N  P  F  F  E  L  E  D  E  M  L  P  S·
   181    GTTCTTAATGCAGCGAGTTAGCAAATCCTTTCTTCGAGTTGGAAGACGAGATGCTTCCTT

          ·S  A  T  S  S  N  C  F  T  S  A  S  S  F  L  A  L  P  D  L  E·
   241    CTGCTACCTCTAGTAATTGTTTTACTTCTGCCTCAAGCTTTCTGGCTTTACCTGATCTTG

          ·E  P  I  S  I  V  S  H  E  D·
   301    AACCCATCTCCATTGTGTCTCATGAAGGTATGTATATTGTAGTTGTTTGATCTCATGGAT

                                                ·D  I  L  S  V  Y  G·
   361    TTTATGTGTGCACTTCTCATTCTTGTCGTTCTGTTTAAGCAGATATACTTAGTGTGTATG

          ·G  S  A  S  W  T  A  E  E  T  M  F  V  S  D  F  A  K  K  S  E·
   421    GTTCTGCTTCATGGACCGCAGAAGAGACGATGTTCGTTTCTGATTTTGCGAAAAAGAGTG

          ·E  T  T  T  T  K  K  R  R  C  R  E  E  C  F  S  S  C  S  V  S·
   481    AAACCACAACTACCAAGAAGAGGAGATGCAGAGAAGAATGTTTTTCTAGTTGTTCTGTTT

          ·S  K  T  L  S  K  E  T  I  S  L  Y  F  Y  M  P  I  T  Q  A  A·
   541    CAAAGACATTGTCGAAGGAAACCATCTCATTGTACTTTTACATGCCGATAACTCAAGCGG

          ·A  R  E  L  N  I  G  L  T  L  L  K  K  R  C  R  E  L  G  I  K·
   601    CTAGAGAGCTTAACATTGGTTTAACTCTTTTGAAGAAGAGATGCCGCGAATTGGGTATTA

          ·K  R  W  P  H  R  K  L  M  S  L  Q  K  L  I  S  N  V  K
   661    AACGTTGGCCTCATCGTAAGCTCATGAGCCTACAAAAACTCATCAGCAATGTCAAGGTAT

   721    AAACAACTATAAACATAAGACAACTAAGTTATAACAAGTTAAATTCCTCTATTGTGTATC

   781    ATTAGGAATTATTTTTTGCATTTTTTAGTTCGCCAAATTTTCTTGAAAATTTGATTTTTG

   841    TTGTTTTGGAAATGAAAACGCTATATAAAGCCTTAATGCTATGGTCTACCGATTTAGTGA

                                             E  L  E  K  M  E  G  E  E  N·
   901    ATTTGATGTGACTATTTTTATAATTGTAACAGGAGCTAGAGAAGATGGAAGGGGAAGAAA

          ·N  E  D  K  L  R  N  A  L  E  K  L  E  K  E  K  K  T  I  E  K·
   961    ATGAAGATAAGCTAAGAAACGCTTTGGAAAAGCTCGAGAAGGAGAAGAAAACGATTGAGA

          ·K  L  P  D  L  K  F  E  D  K  T  K  R  L  R  Q  A  C  F  K  A·
```

```
1021   AGTTACCAGATTTGAAGTTTGAGGATAAGACAAAGAGATTGAGACAAGCTTGTTTCAAGG

      ·A  N  H  K  R  K  R  R  S  G  M  S  T  P  I  T  S  S  S  S S·
1081   CTAACCATAAGAGGAAGAGAAGAAGTGGCATGTCCACGCCCATCACATCATCATCTTCTT

      ·S  A  S  A  S  S  S  S  Y  S  S  V  S  G  F  E  R  *
1141   CTGCTTCTGCTTCTTCTTCTTCTTACTCTTCTGTTTCGGGTTTTGAGAGATAA
```

AtRKD2 (At1g74480) genomic sequence

```
        M  A  D  H  T  T  K  E  Q  K  S  F  S  F  L  A  H  S  P  S
1      ATGGCTGATCACACAACCAAAGAACAGAAGTCATTCTCATTCCTAGCTCATTCTCCATCC

        F  D  H  S·
61     TTTGATCACAGGTTAATTTCAGAAGCTTTTGTATAGTTTCATTTTGTTATTAGTTTTTGT

                                                    S  S  L  S  Y  P·
121    TTTCTTATAATTTATTTTCTCTGTATTTTTCCAAACGCAATGCAGCTCCTTAAGTTATCC

      ·P  L  F  D  W  E  E  D  L  L  A  L  Q  E  N  S  G  S  Q  A  F·
181    TTTATTCGACTGGGAAGAAGATCTTCTTGCTCTCCAAGAAAACTCTGGCTCTCAAGCATT

      ·F  P  F  T  T  T  S  L  P  L  P  D  L  E  P  L  S  E  D·
241    TCCTTTTACTACAACTTCTCTGCCTTTACCTGATCTTGAACCCTTGTCTGAAGGTATAAC

301    ATATTATTACATAGTTTTTATATTCATATAGAGCTATATACATAGAGGTCTCATGGTCTT

                  ·D  V  L  N  S  Y  S  S  A  S  W  N  E  T  E  Q
361    GTTTGTTCTGTAGATGTACTCAATTCATACAGCTCTGCGTCATGGAACGAAACAGAGCAA

        N  R  G  D  G  A  S  S  E  K  K  R  E  N  G  T  V  K  E  T
421    AACAGAGGAGATGGCGCTTCATCGGAGAAGAAGAGGGAAAATGGAACAGTGAAAGAGACA

        T  K  K  R  K  I  N  E  R  H  R  E  H  S  V  R  I  I  S  D
481    ACTAAGAAGAGGAAAATCAATGAGAGACACAGAGAACATAGCGTGAGAATCATCAGCGAT

        I  T  T  Y  T  T  S  S  A  P  T  T  L  S  K  E  T  V  S  R
541    ATTACTACCTACACAACTAGTTCAGCTCCAACGACATTGTCAAAGGAAACTGTCTCTCGC

        Y  F  Y  M  P  I  T  Q  A  A  I  A  L  N  V  G  L  T  L  L
601    TACTTCTACATGCCCATAACTCAGGCTGCAATAGCACTTAACGTTGGTTTAACTCTACTA

        K  R  R  C  R  E  L  G  I  R  R  W  P  H  R  K  L  M  S  L
661    AAAAGGAGATGTCGCGAATTGGGTATTCGCCGATGGCCTCATCGTAAACTTATGAGCTTA

        N  T  L  I  S  N  V  K
721    AACACTTTGATCAGTAACGTCAAGGTATGTAGAACAAGGCACTACACTTAATATTATTGT

                                                    E  L  Q  K  M  E  G·
781    TTTCAGATAATTGAACTCAAATGTTTTATCATATGAATAGGAGCTGCAGAAGATGGAAGG

      ·G  E  E  N  A  E  K  L  Q  D  A  L  E  M  L  E  K  E  K  R  T·
841    CGAAGAGAATGCAGAAAAACTGCAGGACGCGTTGGAGATGCTTGAGAAGGAGAAGAGGAC

      ·T  I  E  D  L  P  D  L  E  F  K  D  K  T  K  R  L  R  Q  A  C·
901    AATTGAGGATTTGCCGGATTTGGAGTTTAAGGACAAGACAAAGAGGCTAAGACAAGCTTG

      ·C  F  K  A  N  H  K  R  K  K  K  R  S  L  K  S  D  Q  S  Q  V·
961    TTTCAAGGCTAACCACAAGAGGAAGAAGAAGAGAAGTCTCAAGTCCGATCAGTCTCAAGT
```

```
      ·V P  S  C  S  S  S  G  S  V  P  S  D  E  S  V  D  E  A  G  M·
1021  ACCCTCGTGTTCAAGCAGCGGATCAGTTCCTAGTGATGAGTCGGTTGATGAAGCAGGAAT

      ·M E  S  D  E  E  M  K  Y  L  L  C  G  F  S  S  E  F  T  S  G·
1081  GGAGAGTGATGAAGAAATGAAGTATCTCTTGTGTGGTTTCTCAAGTGAATTTACTAGTGG

      ·G L  *
1141  TTTGTGA
```

AtRKD3 (At5g66990) genomic sequence

```
       M  A  D  Q  R  P  L  M  T  W  L  E  A  N  N  Y  E  S  F  L
   1   ATGGCTGATCAAAGACCTCTAATGACCTGGTTAGAGGCCAACAACTATGAATCATTCCTT

       Q  E  D  I  F  S  F  L  D  Q  S  L  F  V  D  P  H  S·
  61   CAAGAAGACATATTCTCGTTTCTCGATCAATCACTTTTCGTCGATCCTCACAGGTATAAG

 121   TCACTATGTTTCTACTGTTTGGTCAAACAAAAAAATGATTTCCCTGCGTTGAATCTCAAA

                                          ·S  S  F  I  D  P  F  K  D  F  Q·
 181   TGCTTTTTTTGGTTGTCTTGAGTTTTTGCAGCTCTTTCATTGACCCTTTTAAGGATTTTC

      ·Q  T  Q  N  W  F  S  L  Q  D  S  I  V  N  H  I  S  T  T  F  A·
 241  AAACCCAAAATTGGTTTTCTCTCCAAGACAGCATTGTTAATCATATATCTACTACCTTTG

      ·A  A  D  H  T  F  L  A  S  L  D  L  E  A  I  S  S  T  F  S  L·
 301  CGGCTGATCATACGTTTCTGGCTTCACTTGATCTTGAAGCTATCTCTAGTACTTTCTCTC

      ·L  D·
 361  TAGGTATATATATTTACATAAATGTATGTAGATATACACACACATATATGCAAGCTTGGA

                                          ·D  I  S  S  G  W  W  N  E  N  N·
 421  TTGTAAATCTAATCTTATATTATGTAGCAGATATATCGAGTGGATGGTGGAACGAGAATA

      ·N  G  N  Y  N  N  Q  V  E  P  N  L  D  E  I  S  R  T  N  T  M·
 481  ATGGTAACTACAATAACCAGGTCGAACCAAACCTTGATGAAATTTCAAGAACTAATACCA

      ·M  G  D  P  N  M  E  Q  I  L  H  E  D  V  N  T  M  K  E  K  T·
 541  TGGGAGATCCAAATATGGAGCAAATATTGCATGAAGATGTTAACACAATGAAAGAGAAAA

      ·T  S  Q  K  R  I  I  M  K  R  R  Y  R  E  D  G  V  I  N  N  M·
 601  CAAGCCAGAAGAGGATAATTATGAAGAGGCGATATAGAGAAGATGGAGTCATCAATAATA

      ·M  S  R  E  M  M  K  Q  Y  F  Y  M  P  I  T  K  A  A  K  E  L·
 661  TGTCAAGGGAAATGATGAAGCAGTACTTCTACATGCCGATAACTAAAGCAGCCAAGGAGC

      ·L  N  I  G  V  T  L  L  K  K  R  C  R  E  L  G  I  P  R  W  P·
 721  TTAACATTGGTGTAACCCTCTTGAAGAAAAGATGTCGTGAGTTAGGTATTCCTCGTTGGC

      ·P  H  R  K  L  T  S  L  N  A  L  I  A  N  L  K
 781  CTCACCGTAAGCTCACGAGCCTAAACGCTCTAATTGCTAATCTCAAGGTACGTTCGTCAA

 841  CAGACATTTACAATCAAATTGCCTATACATTTTGGGGTTACAACACTTATATATGTGATG

                       D  L  L  G  N  T  K  G  R  T  P  K  S  K  L  R·
 901  GTCTTCTTTGAAAGGACTTGTTAGGGAACACGAAGGGGAGAACGCCCAAGAGTAAGCTGA
```

```
      ·R  N  A  L  E  L  L  E  M  E  K  K  M  I  E  E  V  P  D  L  E·
  961 GGAACGCTTTGGAGCTTTTGGAGATGGAGAAGAAGATGATTGAGGAAGTTCCCGATTTGG

      ·E  F  G  D  K  T  K  R  L  R  Q  A  C  F  K  A  K  Y  K  R  R·
 1021 AATTTGGGGATAAGACTAAGAGGTTAAGACAGGCTTGCTTCAAGGCTAAATACAAACGGA

      ·R  R  L  F  S  S  S  S  *
 1081 GAAGGCTCTTCTCATCTTCTTCATGA
```

AtRKD4 (At5g53040) genomic sequence

```
       M  S  S  S  K  H  S  S  V  F  N  Y  S  A  L  F  L  S  L  F
    1  ATGAGTTCGTCAAAACATTCCTCTGTTTTTAACTATTCTGCTCTGTTTCTATCACTGTTT

       L  Q  Q  M  D  Q  N  S  L  H  H  L  D  S  P  K  I  E  N  E
   61  CTTCAACAAATGGATCAGAACTCTCTTCATCATCTCGATTCTCCAAAAATCGAAAACGAG

       Y  E  P  D  S  L  Y  D  M  L  D  K·
  121  TATGAACCAGATTCGTTATACGACATGTTAGATAAGTAAGTTGGAAACCCTTTGATACTT

  181  TTTCTTGCTAAACAAGTTATCTCTCTTTTCTTTACTGTTTTTTCTCATTCGTTTTTGGTC

          ·K  L  P  P  L  D  S  L  L  D  M  E  D  L  K  P  N  A  G
  241  TACAGGTTGCCTCCGCTTGATTCTCTCCTAGATATGGAAGATTTGAAACCAAATGCAGGG

       L  H  F  Q  F  H  Y  N  S  F  E  D  F  F  E  N  I  E  V  D
  301  TTGCACTTTCAGTTCCATTACAATAGCTTTGAAGATTTCTTCGAAAACATTGAAGTGGAT

       N  T  I  P  S  D  I  H  L  L  T  Q  E  P  Y  F  S  S  D  S
  361  AACACAATTCCATCTGATATTCACTTGTTGACACAAGAGCCCTACTTCTCAAGTGACTCC

       S  S  S  S  P  L  A  I  Q  N  D  G  L  I  S  N  V  K  V  E
  421  TCTTCCTCTTCACCATTGGCTATCCAAAACGACGGTCTCATTTCCAACGTGAAAGTTGAA

       K  V  T  V  K  K  K  R  N  L  K  K  K  R  Q  D  K  L  E  M
  481  AAGGTAACAGTTAAGAAGAAGAGGAACCTTAAGAAAAAGAGGCAAGACAAATTGGAGATG

       S  E  I  K  Q  F  F  D  R  P  I  M  K  A  A  K  E  L  N  V
  541  TCTGAGATCAAACAATTTTTCGATAGGCCGATCATGAAAGCGGCTAAAGAACTGAACGTG

       G  L  T  V  L  K  K  R  C  R  E  L  G  I  Y  R  W  P  H  R
  601  GGACTCACTGTGTTGAAGAAGCGATGCAGGGAATTAGGAATTTACCGGTGGCCTCACCGG

       K  L  K  S  L  N  S  L  I  K  N  L  K
  661  AAGCTCAAGAGTCTAAACTCTCTTATAAAGAATCTCAAGGTGCAACAACATTAATTTATT

  721  AATAGCATTTGCTTGTGATTCACGAGTAGCGTTTTAGGTTTTATTATATTACGTTCTTGA

                                                 N  V  G  M  E  E  E·
  781  TTTCGAAAGGTTAATAGTTTGTTCATGTTTTTGGTGATAGAATGTTGGAATGGAAGAGGA

       ·E  V  K  N  L  E  E  H  R  F  L  I  E  Q  E  P  D  A  E  L  S·
  841  AGTGAAGAACTTGGAGGAACATAGGTTTCTTATTGAACAAGAACCTGATGCAGAACTCAG

       ·S  D  G  T  K  K  L  R  Q  A  C  F  K  A  N  Y  K  R  R  K  S·
  901  TGATGGAACCAAGAAGCTAAGGCAAGCTTGTTTCAAAGCCAATTATAAGAGAAGAAAATC

       ·S  L  G  D  D  Y  Y  *
```

961 **ACTTGGTGATGATTATTATTGA**

TaRKD1 genomic sequence

1 *GACCCATCTCTCTCGTCCAGACAGCGACAAGTACGTGGTGATCTCTGTACATATTGTGTT*

M E M Q·
61 *GCTCTTTTCATCGATCGTCAGCCGCTCTGCCTTCCGATCGAGTTGCTCC***ATGGAGATGCA**

·Q Q Y F G G C G D G D A D W F H Q L A L L·
121 **GCAGTACTTCGGCGGCTGCGGCGATGGCGATGCTGACTGGTTCCATCAGCTCGCCTTGCT**

·L P P L P V S S S L P P L P M S·
181 **CCCGCCTTTGCCGGTCTCTTCGTCTCTGCCGCCTCTCCCCATGAG**GTCAGTTCAGCGCGT

241 CCTGCAAGTTTCGCCCCATGAGCTCTATTTAGCGCGACATAGAGTCTGATTTAGCATTCT

·S E G S C L P M A A A A P T L P L G·
301 CGTCACCAG**CGAGGGCAGCTGCTTACCTATGGCCGCCGCCGCCCCAACGCTTCCTCTTGG**

·G D C S S A L M I R·
361 **GGATTGCTCATCAGCTCTCATGATTAG**GTGATGATCGATCGGTTCCCTCGATGTAATTAT

421 GCTAAAAAGCCTATGGTTGCTTGAAATATGATTGACACGGCTGATCGATTGGCATGGTTT

·R P E E Q M G C L Q M I P P Q A V A·
481 GATTGATCAG**GCCGGAAGAACAGATGGGCTGCCTGCAGATGATACCTCCACAGGCTGTTG**

·A D D E Y S S Y A T N N V D V L P P F P A·
541 **CCGATGATGAGTACAGCAGCTACGCCACCAACAATGTCGACGTCCTCCCGCCGTTTCCTG**

·A G L D D P T A G L D D A L L M E S F R D·
601 **CAGGTCTCGATGATCCCACGGCAGGCCTCGACGACGCGCTGCTCATGGAGTCCTTCAGAG**

·D I D L E E F A D A V G P K I K T E P L D·
661 **ACATCGACCTGGAGGAGTTCGCCGACGCCGTCGGCCCCAAGATTAAGACCGAGCCTCTCG**

·D D A M V P A D H D F A A Q V Q Q A R P V·
721 **ACGACGCCATGGTGCCGGCGGATCACGATTTCGCGGCGCAAGTGCAACAGGCGCGCCCCG**

·V V I M N Q Q Q L N A P H G V R L L N D P·
781 **TGGTGATCATGAACCAGCAGCAGCTGAATGCGCCACACGGCGTGCGCCTGCTCAATGATC**

·P D D D D S A V V A G G Y E A A A V G C A·
841 **CCGACGACGATGACTCAGCTGTCGTCGCCGGGGGCTATGAGGCGGCGGCCGTTGGGTGCG**

·A E Q K R V R P A P R R V R K S S G G S R·
901 **CTGAGCAGAAGCGGGTGAGGCCGGCGCCACGTCGTGTGCGGAAGAGCAGCGGTGGGTCAC**

·R P A A G G K S L D H I G F E E L R T Y F·
961 **GCCCTGCCGCCGGTGGGAAAAGCCTCGATCACATAGGGTTTGAGGAGCTGCGTACGTATT**

·F Y M P I T K A A R E M N V G L T V L K K·
1021 **TCTACATGCCTATCACCAAGGCGGCGCGGGAGATGAACGTCGGTCTCACCGTGCTCAAGA**

·K R C R E L G V A R W P H R K M K S L R S·
1081 **AGCGCTGCCGTGAGCTCGGTGTCGCCCGTTGGCCTCACCGGAAGATGAAGAGCCTCAGGT**

```
      ·S  L  I  L  N  I  Q
1141  CTCTCATCCTTAACATCCAGGTACGTAGGTGCACACATATCGGTACAACACCACATTTGA

1201  AATGTCGCACACCGCATGTTAGCTATTCCCTCCATTCCAATTTGAACTAGAATCACCACA

                                                  D  M  G  K  G  A  T
1261  AGAATTATGGAACAACGGGTAATAGAAAATTGCGCGCAGGACATGGGGAAGGGCGCCACG

       S  P  A  A  V  Q  G  E  L  E  A  L  E  R  Y  C  A  I  M  E
1321  TCGCCGGCGGCGGTGCAAGGGGAGCTGGAGGCGCTTGAGAGGTATTGTGCCATAATGGAG

       E  N  P  A  I  E  L  T  E  Q  T  K  K  L  R  Q  A  C  F  K
1381  GAGAACCCGGCGATCGAGCTGACGGAGCAGACCAAGAAGCTGAGGCAGGCCTGCTTTAAG

       E  N  Y  K  R  R  R  A  A  A  S  V  N  L  L  E  H  C  Y  N
1441  GAGAACTACAAGAGGAGGAGAGCGGCGGCCTCCGTCAACTTGCTCGAGCATTGCTACAAC

       D  L  G  S  H  E  Q  Q  M  P  L  P  Q  M  G  F  F  G  F  *
1501  GACTTGGGCAGTCATGAGCAGCAGATGCCATTGCCACAGATGGGTTTCTTTGGGTTCTAA

1561  AGGGACAGGAATTCACATTAACAGTATATATACTTGGGGCAACAACTGAAGAAGTAAATA

1681  AATGACGTCGACGCATGCATGGGAGGTGCCGCTTATTTTTGCGTCATTTACGAGATGCAA

1741  ACATTTCTGTATACAAGATGCAAACATTTCTCTATGAATTGAGTGGCTCGTGC
```

TaRKD2 genomic sequence

```
   1  GACCCATCTCTCTCGTCCAGACAGCAAGTAGTACGTGGTGATCTCTGTACATATTGTGTT

                                                        M  E  M  Q·
  61  GCTCTTTTCATCGATCGTCAGTCGCTCTGCCTTCCGATCGAGTTCCCCCATGGAGATGCA

      ·Q  Q  Q  Y  F  G  G  D  G  D  A  D  W  F  H  Q  L  A  L  L  P·
 121  ACAGCAATACTTCGGCGGTGATGGCGATGCTGACTGGTTCCACCAGCTCGCCTTGCTCCC

      ·P  P  L  P  I  S  S  S  L  P  P  L  P  M  S·
 181  GCCTTTGCCGATCTCTTCGTCTCTGCCGCCTCTCCCCATGAGGTCAGTTCAGCGCGTCCT

 241  GCAAGTTTCTTAATTTGTACTGCAAGTTTCGCCCCATGAGCTCTATTCAGCACGACCTAG

                                  ·S  E  G  S  C  L  P  M  A  A  A  A·
 301  AGTCTGATTTAGCATTCTCGTCACCAGCGAGGGCAGCTGCTTACCCATGGCCGCCGCCGC

      ·A  A  A  A  L  P  L  G  D  C  S  S  A  L  M  I  R·
 361  CGCGGCGGCGCTTCCTCTTGGGGATTGCTCATCAGCTCTCATGATTAGGTGATGATCGAT

 421  CGGTTCCCTCGATGTAACTGTACTTAAAAGCGTAGTTGCTTAAAATATGATTAACACGGC

                                     ·R  P  E  E  Q  M  S  C  L  P  M
 481  TGATCGAATTCTATGGTTTGATCGATCAGGCCGGAAGAACAGATGAGCTGCCTGCCGATG

       N  P  S  P  A  V  V  D  D  V  Y  S  S  Y  A  P  N  N  V  D
 541  AACCCTTCGCCAGCTGTTGTCGATGATGTGTACAGCAGCTACGCCCCCAACAATGTCGAC
```

```
      V  L  P  P  F  P  A  G  L  D  D  A  L  L  M  E  S  F  S  D
601   GTCCTCCCGCCGTTTCCTGCAGGTCTCGACGACGCGCTGCTCATGGAGTCCTTCAGCGAT

      I  D  L  E  E  F  A  D  A  F  G  H  K  I  K  T  E  P  L  D
661   ATCGACCTGGAGGAGTTCGCCGACGCCTTCGGCCACAAGATTAAGACCGAGCCTCTCGAC

      D  A  M  V  P  A  D  H  D  F  A  A  Q  A  Q  Q  A  C  P  V
721   GACGCTATGGTGCCGGCGGATCACGATTTCGCGGCGCAAGCGCAACAGGCGTGCCCAGTG

      V  I  M  N  Q  Q  Q  L  N  A  P  R  D  V  R  L  L  I  D  P
781   GTGATCATGAACCAGCAGCAGCTGAATGCACCACGCGACGTGCGCCTGCTCATAGATCCC

      D  D  D  D  S  T  V  V  A  G  G  Y  E  A  A  A  V  G  C  A
841   GACGACGATGACTCAACTGTCGTCGCCGGGGGCTATGAGGCTGCGGCCGTTGGGTGCGCT

      E  Q  K  Q  V  R  P  A  P  R  R  V  R  K  S  S  G  G  A  R
901   GAGCAGAAGCAGGTGAGGCCGGCGCCACGTCGTGTGCGCAAGAGCAGCGGTGGGGCACGC

      P  A  A  G  G  K  S  L  D  H  I  G  F  E  E  L  R  T  Y  F
961   CCAGCCGCCGGTGGGAAAAGCCTCGATCACATAGGGTTTGAGGAGCTACGTACGTATTTC

      Y  M  P  I  T  K  A  A  R  E  M  N  V  G  L  T  V  L  K  K
1021  TACATGCCTATCACCAAGGCGGCGCGGGAGATGAACGTTGGTCTCACCGTGCTCAAGAAG

      R  C  R  E  L  G  V  A  R  W  P  H  R  K  M  K  S  L  R  S
1081  CGCTGCCGAGAGCTCGGGGTCGCCCGCTGGCCTCACCGGAAGATGAAGAGCCTCAGGTCA

      L  I  L  N  I  Q
1141  CTCATCCTCAACATCCAGGTACGTGCACACAATAGATCGCTACAACACCACATTCCACCG

1201  CGTGTTAGCTATTCTCTCTGTTCCATAATTCTTGTCGTGTTTTTAGTTCAGCAAAATATG

                                                        E  M  G  K  G·
1261  AAACGACAAGAATTATGGAACAAAGTGTAATGGAAAATTGTGTGCAGGAGATGGGGAAGG

     ·G  A  T  S  P  A  A  V  Q  G  E  L  E  A  L  E  R  Y  C  A  I·
1321  GCGCAACGTCGCCGGCGGCTGTGCAAGGGGAACTAGAGGCGCTTGAGAGGTATTGCGCCA

     ·I  M  E  E  N  P  A  I  E  L  T  E  Q  T  K  K  L  R  Q  A  C·
1381  TAATGGAGGAGAACCCGGCGATCGAGCTGACTGAGCAGACCAAGAAGCTGCGGCAGGCCT

     ·C  F  K  E  N  Y  K  R  R  R  A  A  A  S  V  N  L  L  D  H  C·
1441  GCTTTAAGGAGAACTACAAGAGGAGGAGAGCGGCGGCCTCCGTCAACTTGCTCGACCATT

     ·C  Y  N  D  L  A  S  H  E  Q  Q  M  P  L  P  Q  M  G  F  F  G·
1501  GCTACAACGACTTGGCCAGTCATGAGCAGCAGATGCCATTGCCACAGATGGGTTTCTTTG

     ·G  F  *
1561  GGTTCTAAAGGGACAGGAATTCACATTAACGGTATATATATACTGGGGCATATACTGAAA

1621  AAGTAAATAAGTAACCCACTGAAACGATCATAGATGATCGATGGGTATTCGGCACCTTTT

1681  TGCGTCATTTACATTATGCAATCATTTCTCTATGAATT
```

TaRKD3 genomic sequence

```
       1  GACCCATCTCTCTCGTCCAGACAGCAACAAGTACGTGGTGTTGCTCTGTTCCTCGATCGT

                                              M  E  M  Q  Q  Y  Y  G  G  G·
      61  CAGCCGCTGTCTCTTCCGATCGAAGTTCCTCCATGGAGATGCAGCAATACTACGGCGGCG

          ·G  G  G  C  D  G  D  G  D  W  F  H  Q  L  A  A  L  P  P  L  P·
     121  GCGGCGGGTGCGATGGCGATGGTGACTGGTTCCATCAGCTGGCCGCTCTCCCGCCTTTGC

          ·P  M  S  S  S  L  P  P  P  L  L  M  S·
     181  CGATGAGCTCATCGCTGCCGCCGCCTCTCCTCATGAGGTCAGTTCAGCACGTCCTGCAAC

     241  TTTCTTAATTTGTACCCGTCGATCTATCCATCACGACCCGGAGATTGACTTAGCATTTTT

                ·S  E  G  S  C  L  P  M  A  A  A  A  A  L  P  L  G  D
     301  GTCGCCAGCGAGGGCAGCTGCTTACCCATGGCCGCCGCCGCCGCGCTTCCTCTTGGGGAT

           C  S  S  A  L  M  I  R·
     361  TGCTCTTCAGCTCTCATGATTAGGTGATGATCGATCGGTTCCCTCGATGTAATTATGCTA

     421  AAAAGCCTATGGTTGCTTGAAATATGATTGACACGGCTGATCGATTGGCATGGTTTGATT

               ·R  P  E  E  Q  M  S  C  L  P  M  N  P  S  P  A  V  V  D·
     481  GATCAGGCCGGAAGAACAGATGAGCTGCCTGCCGATGAACCCTTCGCCAGCTGTTGTCGA

          ·D  D  V  Y  S  S  Y  A  P  N  N  V  D  V  L  P  P  F  P  A  G·
     541  TGATGTGTACAGCAGCTACGCCCCCAACAATGTCGACGTCCTCCCGCCGTTTCCTGCAGG

          ·G  L  D  D  A  L  L  M  E  S  F  S  D  I  D  L  E  E  F  A  D·
     601  TCTCGACGACGCGCTGCTCATGGAGTCCTTCAGCGATATCGACCTGGAGGAGTTCGCCGA

          ·D  A  F  G  H  K  V  K  T  E  P  L  D  D  A  M  V  P  A  D  H·
     661  CGCCTTCGGCCACAAGGTTAAGACCGAGCCTCTCGACGACGCTATGGTGCCGGCGGATCA

          ·H  D  F  A  A  Q  A  Q  Q  A  C  P  V  V  I  M  N  Q  Q  Q  L·
     721  CGATTTCGCGGCGCAAGCGCAGCAGGCGTGCCCAGTGGTGATCATGAACCAGCAGCAGCT

          ·L  N  A  P  R  D  V  R  L  L  I  D  P  D  D  D  D  S  T  V  V·
     781  GAATGCACCACGCGACGTGCGCCTGCTCATAGATCCCGACGACGATGACTCAACTGTCGT

          ·V  A  G  G  Y  E  A  A  A  V  G  C  A  E  Q  K  Q  V  R  P  A·
     841  CGCCGGGGGCTATGAGGCTGCGGCCGTTGGGTGCGCTGAGCAGAAGCAGGTGAGGCCGGC

          ·A  P  R  R  V  R  K  S  S  G  G  A  R  P  A  A  G  G  K  S  L·
     901  GCCACGTCGTGTGCGCAAGAGCAGCGGTGGGGCACGCCCAGCCGCCGGTGGGAAAAGCCT

          ·L  D  H  I  G  F  E  E  L  R  T  Y  F  Y  M  P  I  T  K  A  A·
     961  CGATCACATAGGGTTTGAGGAGCTACGTACGTATTTCTACATGCCTATCACCAAGGCGGC

          ·A  R  E  M  N  V  G  L  T  V  L  K  K  R  C  R  E  L  G  V  A·
    1021  GCGGGAGATGAACGTTGGTCTCACCGTGCTCAAGAAGCGCTGCCGAGAGCTCGGGGTCGC

          ·A  R  W  P  H  R  K  M  K  S  L  R  S  L  I  L  N  I  Q  E  M·
    1081  CCGCTGGCCTCACCGGAAGATGAAGAGCCTCAGGTCACTCATCCTCAACATCCAGGAGAT

          ·M  G  K  G  A  T  S  P  A  A  V  Q  G  E  L  E  A  L  E  R  Y·
    1141  GGGGAAGGGCGCAACGTCGCCGGCGGCTGTGCAAGGGGAACTAGAGGCGCTTGAGAGGTA
```

·Y R A I M E E N P A I E L T E R T K K L R·
1201 **TCGCGCCATAATGGAGGAGAACCCGGCGATCGAGCTGACTGAGCGGACCAAGAAGCTGCG**

·R Q A C F K E N Y K R R R A A A S V N L L·
1261 **GCAGGCCTGCTTTAAGGAGAACTACAAGAGGAGGAGAGCGGCGGCCTCCGTCAACTTGCT**

·L D H C Y N D L A S H E Q Q M P L P Q M G·
1321 **CGACCATTGCTACAACGACTTGGCCAGTCATGAGCAGCAGATGCCATTGCCACAGATGGG**

·G F F G F *
1381 **TTTCTTTGGGTTCTAA**

TaRKD4 genomic sequence

1 *GACCCATCTCTCTCGTCCAGACAGCAAGTAGTACGTGGTGATCTCTGTACATATTGTGTT*

M E M Q·
61 *GCTCTTTTCATCGATCGTCAGTCGCTCTGCCTTCCGATCGAGTTCCCCC***ATGGAGATGCA**

·Q Q Q Y F G G D G D A D W F H Q L A L L P·
121 **ACAGCAATACTTCGGCGGTGATGGCGATGCTGACTGGTTCCACCAGCTCGCCTTGCTCCC**

·P P L P I S S S L P P L P M S E G S C L P·
181 **GCCTTTGCCGATCTCTTCGTCTCTGCCGCCTCTCCCCATGAGCGAGGGCAGCTGCTTACC**

·P M A A A A A A A A L P L G D C S S S L M I·
241 **CATGGCCGCCGCCGCCGCGGCGGCGCTTCCTCTTGGGGATTGCTCATCATCTCTCATGAT**

·I R P E E Q M S C L P M N P S P A V V D D·
301 **TAGGCCGGAAGAACAGATGAGCTGCCTGCCGATGAACCCTTCGCCAGCTGTTGTCGATGA**

·D V Y S S Y A P N N V D V L P P F P A G L·
361 **TGTGTACAGCAGCTACGCCCCCAACAATGTCGACGTCCTCCCGCCGTTTCCTGCAGGTCT**

·L D D A L L M E S F S D I D L E E F A D A·
421 **CGACGACGCGCTGCTCATGGAGTCCTTCAGCGATATCGACCTGGAGGAGTTCGCCGACGC**

·A F G H K I K T E P L D D A M V P A D H D·
481 **CTTCGGCCACAAGATTAAGACCGAGCCTCTCGACGACGCTATGGTGCCGGCGGATCACGA**

·D F A A Q A Q Q A C P V V I M N Q Q Q L N·
541 **TTTCGCGGCGCAAGCGCAGCAGGCGTGCCCAGTGGTGATCATGAACCAGCAGCAGCTGAA**

·N A P R D V R L L I D P D D D D S T V V A·
601 **TGCACCACGCGACGTGCGCCTGCTCATAGATCCCGACGACGATGACTCAACTGTCGTCGC**

·A G G Y E A A A V G C A E Q K Q V R P A P·
661 **CGGGGGCTATGAGGCTGCGGCCGTTGGGTGCGCTGAGCAGAAGCAGGTGAGGCCGGCGCC**

·P R R V R K S S G G A R P A A G G K S L D·
721 **ACGTCGTGTGCGCAAGAGCAGCGGTGGGGCACGCCCAGCCGCCGGTGGGAAAAGCCTCGA**

·D H I G F E E L R T Y F Y M P I T K A A R·
781 **TCACATAGGGTTTGAGGAGCTACGTACGTATTTCTACATGCCTATCACCAAGGCGGCGCG**

·R E M N V G L T V L K K R C R E L G V A R·
841 **GGAGATGAACGTTGGTCTCACCGTGCTCAAGAAGCGCTGCCGAGAGCTCGGGGTCGCCCG**

·R W P H R K M K S L R S L I L N I Q E M G·
901 **CTGGCCTCACCGGAAGATGAAGAGCCTCAGGTCACTCATCCTCAACATCCAGGAGATGGG**

```
     ·G K  G  A  T  S  P  A  A  V  Q  G  E  L  E  A  L  E  R  Y  C·
 961 GAAGGGCGCAACGTCGCCGGCGGCTGTGCAAGGGGAACTAGAGGCGCTTGAGAGGTATTG

     ·C A  I  M  E  E  N  P  A  I  E  L  T  E  R  T  K  K  L  R  Q·
1021 CGCCATAATGGAGGAGAACCCGGCGATCGAGCTGACTGAGCGGACCAAGAAGCTGCGGCA

     ·Q A  C  F  K  E  N  Y  K  R  R  R  A  A  A  S  V  N  L  L  D·
1081 GGCCTGCTTTAAGGAGAACTACAAGAGGAGGAGAGCGGCGGCCTCCGTCAACTTGCTCGA

     ·D H  C  Y  N  D  L  A  S  H  E  Q  Q  M  P  L  P  Q  M  G  F·
1141 CCATTGCTACAACGACTTGGCCAGTCATGAGCAGCAGATGCCATTGCCACAGATGGGTTT

     ·F F  G  F  *
1201 CTTTGGGTTCTAA
```