Theor Appl Genet (2010) 120:1597–1610
DOI 10.1007/s00122-010-1279-8

ORIGINAL PAPER

Brassica GLABRA2 genes: analysis of function related to seed oil content and development of functional markers

Guohua Chai · Zetao Bai · Fang Wei · Graham J. King · Chenggang Wang · Lei Shi · Caihua Dong · Hong Chen · Shengyi Liu

Received: 5 November 2009 / Accepted: 19 January 2010 / Published online: 17 February 2010

Abstract Regulation of seed oil accumulation in oilseed rape (*Brassica napus*) has important economic significance. However, few genes have been characterized that affect final seed oil content. Through a mutant identification, the class IV homeodomain-ZIP transcription factor GLABRA2 (GL2) has been found to regulate seed oil accumulation in *Arabidopsis*, in addition to its role in trichome development. In this study, we isolated four distinct orthologues of *GL2* from *B. napus* (AC-genome), *B. rapa* (A) and *B. oleracea* (C), using an overlapping-PCR strategy. The four *GL2* orthologues were very similar, with 96.10–99.69% identity in exon regions, 75.45–93.84% in intron regions, 97.34–99.87% in amino acid sequences. Alignments of the four genes revealed that the A-genome sequences of *BnaA.GL2.a* from *B. napus* and *BraA.GL2.a* from *B. rapa* are more similar than the others, and likewise the C-genome sequences of *BnaC.GL2.b* from *B. napus* and *BolC.GL2.a* from *B. oleracea* are more similar. *BnaA.GL2.a* and *BraA.GL2.a* from the A-genome are highly expressed in roots, whilst *BnaC.GL2.b* and *BolC.GL2.a* from the C-genome are preferentially expressed in seeds. Transgenic ectopic overexpression and suppression of *BnaC.GL2.b* in *Arabidopsis* allowed further investigation of the effect on seed oil content. Overexpression generated two phenotypes: the wild-type-like and the *gl2*-mutant-like (an *Arabidopsis* glabrous mutant of *gl2-2*), with increases in seed oil content of 3.5–5.0% in the *gl2*-mutant-like transgenic plants. Suppression resulted in increases of 2.5–6.1% in seed oil content, and reduced trichome number at the leaf margins. These results suggest that *BnaC.GL2.b* can negatively regulate oil accumulation in *Arabidopsis* seeds. As a result of comparing the four *GL2* genes, three A/C-genome-specific primer sets were developed and a C-genome-specific *Eco*RV cleavage site was identified, which can be used as functional markers to distinguish these orthologues within *Brassica* species. The genes identified and their molecular markers developed in this study will be valuable both for oilseed rape breeding focusing on improvement of seed oil content, and for detecting gene flow between populations.

Communicated by C. Quiros.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1279-8) contains supplementary material, which is available to authorized users.

G. Chai · Z. Bai · F. Wei · C. Wang · L. Shi · C. Dong · H. Chen · S. Liu (✉)
The Key Lab of Oil Crops Biology,
The Ministry of Agriculture of the People's Republic of China,
Oil Crops Research Institute, Chinese Academy of Agricultural Sciences, No. 2 Xudong 2nd Road, Wuhan 430062,
Hubei, People's Republic of China
e-mail: liusy@oilcrops.cn

G. J. King
Rothamsted Research, Harpenden, Hertfordshire AL5 2JQ, UK

Introduction

Oilseed rape (*Brassica napus* L., *Brassicaceae*) is the third largest oilseed crop in the world, providing approximately 13% of the world's supply of vegetable oil (Hajduch et al. 2006). In spite of its economic importance, there has been little work on functional analysis of the key genes controlling seed oil content in oilseed rape. The reasons could be that oilseed rape (genome AACC, $2n = 38$) originated from a spontaneous hybridization between *B. rapa* L. (syn. *campestris*; AA, $2n = 20$) and *B. oleracea* L. (CC, $2n = 18$) and comprises two sets of homoeologous

chromosomes from the two species. Thus the *B. napus* genome contains many sequentially homologous genes (Parkin et al. 2003; Snowdon et al. 2002). Most genes in the amphidiploid genomes are predicted to exist as multiple copies (≥2) with highly similar structures (Lysak et al. 2005). It is therefore difficult to distinguish such paralogues within *B. napus*. This, together with the complex oil biosynthetic pathway and unclear control mechanisms, has so far constrained efforts to understand the regulation, and to modify seed oil quantity and quality in oilseed rape.

In contrast, considerable advances have been made in understanding the genetic and biochemical basis of seed oil accumulation in the closely related *Arabidopsis* species. Screens of *Arabidopsis* mutants have identified a number of transcription factors (TFs), including WRINKLED1 (WRI1), GLABRA2 (GL2), LEAFY COTYLEDON1 (LEC1), LEC2, ABSCISIC ACID INSENSITIVE3 (ABI3), FUSCA3 (FUS3), PICKLE (PKL) and 6b interacting protein 1-like 1 (ASIL1), which have been shown to affect oil accumulation in seed (Focks and Benning 1998; Shen et al. 2006; Mu et al. 2008; Baud et al. 2007; Finkelstein and Somerville 1990; Tiedemann et al. 2008; Ogas et al. 1999; Gao et al. 2009). Manipulation of such regulators is believed to be a more effective strategy for increasing oil content than the manipulation of single biochemical steps in plant seed oil biosynthesis (Shen et al. 2006). Apart from GL2, the TFs identified to date control seed oil accumulation via involvement in the process of plastidial glycolysis, pyruvate dehydrogenase action and lipid synthesis during *Arabidopsis* seed maturation. To our knowledge, the mechanism by which the *GL2* gene regulates seed oil content remains unclear. The *GL2* gene belongs to the class IV homeodomain (HD)-ZIP subfamily and features three conserved domains, a homeodomain DNA-binding domain, two-successive-ZIP domains, and a START motif (Ariel et al. 2007). Further functional analysis has suggested that the *GL2* gene also plays a crucial role in root hairless cell specification, leading to the local outgrowth of the trichome in shoot epidermal cells and seed mucilage production (Szymanski et al. 2000; Nakamura et al. 2006). In addition to *Arabidopsis*, full-length sequences of *GL2* genes have also been cloned from other plant species, where they mostly exist as multiple homologous genes. These include five homologous *GL2* paralogues in maize (Ingram et al. 1999, 2000), two in rice (Ito et al. 2002; Yang et al. 2002), two in Norway spruce (Ingouff et al. 2001, 2003), two in cotton (Guan et al. 2008), and one in *Phalaenopsis* (Nadeau et al. 1996). However, to date, there has been no further study of function affecting seed oil content, and no *GL2* genes have been isolated from *Brassica*. We therefore sought to clone and characterize the *Brassica GL2* orthologues, particularly in the context of investigating their roles in regulating seed oil content.

Molecular markers have been used to assist rapeseed breeding and selection procedures (Snowdon and Friedt 2004). However, use of traditional markers such as RFLP, AFLP and SSR are limited, largely due to their incomplete correlation with gene function. In contrast, functional (or 'in-gene') markers (FMs) are usually designed from polymorphisms within transcribed regions of functional genes and can dramatically facilitate accurate selection of target characters in breeding programs (Wei et al. 2009). It has now been demonstrated that functional markers can be applied with high efficiency in marker-assisted selection (MAS) for rapid transfer of a character into an otherwise desirable genotype. For instance, Blake et al. (2004) designed genome-specific PCR primer pairs for a series of starch biosynthesis genes, such as *Agp-L*, *SUT*, *Wx*, *Agp-S* and *SsI* in common wheat.

In this study, we have isolated four homologous *GL2* genes, *BnaA.GL2.a*, *BnaC.GL2.b* from *B. napus*, *BraA.GL2.a* from *B. rapa*, and *BolC.GL2.a* from *B. oleracea*. Their expression patterns were divided into two groups with *BnaC.GL2.b* predominantly expressed in seeds. We have demonstrated that when *BnaC.GL2.b* is introduced into *Arabidopsis* plants it functionally substitutes for *Arabidopsis GL2*. We suggest that *BnaC.GL2.b* may participate in regulation of seed oil content in oilseed rape. Based on sequence divergence of four *GL2* orthologues, three specific primer sets (exon-anchored, A- and C-genome-specific) were developed and an *Eco*RV cleavage site unique to the C-genome was identified. These functional markers can be used to distinguish the four *GL2* orthologues and different *Brassica* species.

Materials and methods

Plant materials and growth conditions

Nine *Brassica* cultivars were used, representing *B. napus* var. *oleifera* (cv. Zhongshuang 9, M083 and '583'), *B. rapa* var. *parachinensis* (cv. Sijiucaixin), var. *rosularis* (cv. Wutacai), var. *campestris* (cv. Yellow Sarson), and *B. oleracea* var. *acephala* (cv. Yuyiganlan), var. *alboglabra* (cv. Xianggangbaihuajielan), var. *capitata* (cv. Lvbao). The plants were grown in a greenhouse at 28°C (day) and 14°C (night) with a natural photoperiod. *Arabidopsis thaliana* ecotype Col-0 was used as transformation recipient. Col-0 seeds were sown in planting soil (1:1 mixture of peat moss enriched soil: vermiculite) and cold treated (at 4°C) for 3–4 days prior to transfer to growth chambers with a day/night cycle of 25°C (16 h light) and 20°C (8 h dark).

Isolation of full-length *BnaC.GL2.b* cDNA and *B. napus GL2* 5′ non-coding region

A clone (210c011) was identified from a sequenced oilseed rape cDNA library, which had similarity (87.32%) to the *Arabidopsis GL2* (L32873) sequence and was designated as *BnaC.GL2.b* according to the standard nomenclature of Ostergaard and King (2008). The library had been constructed from mixed mRNA of *B. napus* leaves treated with fungus *Sclerotinia sclerotiorum* or chemicals (benzothiadiazole, oxalic acid and methyl jasmonate). Following sequencing, the clone was shown to contain an intact open reading frame (ORF) and encoded a GL2 protein similar to that of *Arabidopsis*.

The *GL2* 5′ non-coding region of *B. napus* was cloned using the Genome Walker Universal Kit (Takara Clontech, Kyoto, Japan). Manufacturer's instructions were followed with *B. napus* (Zhongshuang 9) genomic DNA being digested with *Eco*RV, *Dra*I, *Pvu*II and *Stu*I. The digested products were ligated to the Genome Walker Adaptor to form four libraries. For the primary PCR, DNA of each library, was used as the template and was annealed with the adaptor primer 1 (AP1) and the *GL2*-specific primer 1 (GSP1) (5′-CTGCCGGAGGATGCATTCCGAAATATC-′3). Subsequently, nested PCR was performed with the primer set of the adaptor primer 2 (AP2) and the *GL2*-specific primer 2 (GSP2) (primer 16 in Fig. 1: 5′-TGGGTTGTTT CGATGACATCTCGACGG-′3).

PCR strategy for cloning other three *GL2* homologous genes

By comparing *B. napus BnaC.GL2.b* cDNA sequence with *Arabidopsis GL2* genomic DNA sequence, we designed conservative primers on exon regions and non-coding regions to amplify multiple *B. napus GL2* homologous genes. PCR was performed with genomic DNA and cDNA from Zhongshuang 9, respectively. To obtain more sequence variability, ten short amplification fragments were clustered and assembled and every intron was amplified with three primer sets (Fig. 1; Table 1). By comparing the sequence identity of two overlapping fragments, *B. napus GL2* homologous sequences were obtained.

Based on the *B. napus GL2* sequences, three pairs of primers (P1/P4, P6/P7 and P8/P9 in Table 1; Fig. 1) were designed to amplify *BraA.GL2.a* and *BolC.GL2.a* separately from genomic DNA or cDNA of *B. rapa* (Yellow Sarson) and *B. oleracea* (Yuyiganlan). The expected PCR fragments (F3, F5 and F6 in Fig. 1) were assembled to encompass a complete coding region and a partial untranslated region. The most upstream primer (P1) and the most downstream primer (P9) were uniquely located in the corresponding 5′ and 3′ untranslated regions. The other four primers (P4, 6, 7 and 8) were designed based on the regions conserved between *BnaA.GL2.a* and *BnaC.GL2.b*.

Sequencing PCR product and sequence analysis

All PCR reactions were performed in a reaction mixture (50 μl) containing 100 ng of genomic DNA or cDNA, 0.2 mM dNTP, 0.5 mM each primer, 1.25 unit LA *Taq* and 50× LA-PCR Buffer II (Mg^{2+} plus) (Takara, Kyoto, Japan). Amplification products were separated by electrophoresis on 1% agarose gels, purified using a Mini-DNA Rapid Purification Kit (BioDev, Beijing, China), and cloned into PMD18-T (Takara) to sequence both strands using the dideoxy chain termination method with the ABI 3730 sequencer.

The integrated nucleotide and deduced amino acid sequences were analyzed using the software DNAMAN (Lynnon Biosoft Co., Canada). Sequences were initially aligned using the local NCBI standalone Blast software, and then submitted to DNAsp 4.0 (http://www.ub.es/dnasp/) to calculate the nucleotide diversity. A phylogenetic tree was obtained by analyzing the amino acid sequence divergence of the GL2 proteins, and a dendrogram constructed using the neighbor-joining method with arithmetic mean

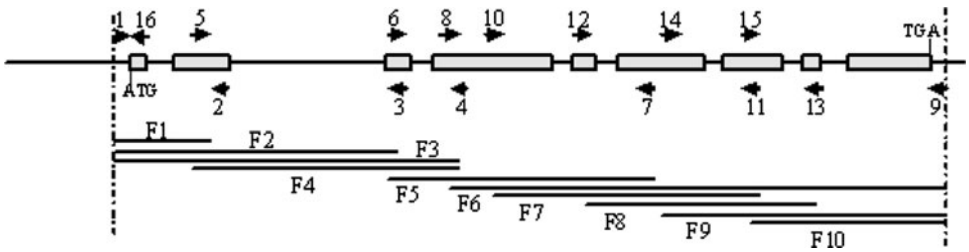

Fig. 1 Primer positions (*arrowheads*) in the *GL2* genes of *Brassica napus*. 15 different primers were used for PCR to determine the *GL2* genes structures. Primer 16 was annealed with the sequence of AP2 primer in nested PCR to amplify the 5′ non-coding regions by the Genome Walking method (Clontech, GenomeWalker Universal Kit). ATG and TGA indicate start and stop codons, respectively. *Rectangles* indicate exon regions. *Solid lines* stand for introns. *Bars* (F1–F10, correspond to primer pairs in Table 1) indicate cloned DNA fragments by PCR

Table 1 Oligonucleotide primers used for PCR amplification

Fragment	Upstream primer (F)	Downstream primer (R)	Annealing temperature[a]	Fragment length
F1[b]	1[c]: 5′-AGATTTCGGAGCGTATGTCAATGG-′3	2: 5′-GCTACTCTCCTTCCCAGAAAG-′3	58	262[d]
F2	1: 5′-AGATTTCGGAGCGTATGTCAATGG-′3	3: 5′-CGGATGAGGTGTTTCTTTGA-′3	60	1,130
F3	1: 5′-AGATTTCGGAGCGTATGTCAATGG-′3	4: 5′-TGACTCTCTCATCGCTTTGT-′3	58	1,390
F4	5: 5′-GAGATGAGCAGCGAGAACTCT-′3	4: 5′-TGACTCTCTCATCGCTTTGT-′3	56	1,103
F5	6: 5′-CGGACGAGAAGCAAAGACA-′3	7: 5′-CTCAGCAGTGAACCCTCCTT-′3	60	1,161
F6	8: 5′-TCAAGGCTATTCAAGAACGG-′3	9: 5′-GTGAGTTATTGTATCATCAAC-′3	50	2,217
F7	10: 5′-GAAGCCTTGAGACTGGCCGTGAG-′3	11: 5′-GTGTCATCCTCTGCGCCATCTTG-′3	65	986
F8	12: 5′-GATCGACGGTGCCATTCAATTG-′3	13: 5′-ACAATAGACTGAACATGAGC-′3	55	987
F9	14: 5′-GTGGAGCACCTTGACTTGT-′3	9: 5′-GTGAGTTATTGTATCATCAAC-′3	53	1,162
F10	15: 5′-CCTGTTTCTCCTACTCTTCTC-′3	9: 5′-GTGAGTTATTGTATCATCAAC-′3	50	698
Bnactin	5′-TCTGGCATCACACTTTCTACAACGAGC-′3	5′-CAGGGAACATGGTCGAACCACC-′3	57	669
Bractin	5′-GGAACTGGAATGGTGAAGGCT-′3	5′-ATCCCAACCATGACACCAGTG-′3	53	104
Boactin	5′-CCATCATGAAGTGTGACGTGGA-′3	5′-TGCTCATACGGTCTGCAATCC-′3	53	102
PBnGL2A	5′-ATGGAGGAGGACAACAACGC-′3	5′-GGGCGACCCAATGTTTAGCA-′3	52	208
PBnGL2C	5′-GTGGAGGAGGACAACAACAGTA-′3	5′-AGGGTGGCAACCCAATGACG-′3	54	215
PBGL2	5′-ATGTCAATGGCCGTCGAGAT-′3	5′-CGGATGAGGTGTTTCTTTGA-′3	50	360
PGL2Tr	5′-AGTCTAGAAGATTTCGGAGCGTATGT-′3	5′-GCTCTAGAAGGTAAACAACATCTCC-′3	54	2,321
PK	5′-TCCACAGTTCAGCCTCTCTT-′3	5′-TCAACAATCTTCGATTTGTAGGC-′3	50	880
BnaC.GL2.b	5′-GCACAAGAGGACAGAAACACA-′3	5′-GCTTCTCTTAATGTTGTGGAGC-′3	58	151
GL2	5′-ATATGGGTGCTGCAAGACAG-′3	5′-GGAGGATTTGGATGTTGCTT-′3	50	124
Atactin	5′-CCCGCTATGTATGTCGCCA-′3	5′-AACCCTCGTAGATTGGCACAG-′3	52	196
BnGL2A	5′-TATCAGTGCCGGTCTAACATCTTCAAG-′3	5′-GATATATGAGTGGAATCAAAACGGT-′3	54	984
BnGL2C	5′-ACTGAGCTAATACATGCATGGTC-′3	5′-TATGAGTGGAATTAATCAAAACG-′3	50	894

[a] PCR annealing temperature (°C) which was used in the second step of 35 cycles

[b] Ten amplification fragments (F1–F10) corresponding to those presented by ten bars in Fig. 1, respectively

[c] Primers 1–15 stand for primer positions and their combinations for PCR amplification indicated by arrows in Fig. 1

[d] The expected length (bp) of the amplified regions (F1–F10) predicted from the *Arabidopsis GL2* gene

incorporated into the MEGA version 2.1 (Kumar et al. 2001).

Determination of the number of *Brassica GL2* orthologues

Southern blot was performed to determine the number of *GL2* orthologues in the three *Brassica* species. According to the description of Wang et al. (2009), genomic DNAs (25 μg each) of young *Brassica* plant leaves (Zhongshuang 9, Yuyiganlan, Yellow Sarson) were firstly digested with the restriction endonucleases *Eco*RI and *Hin*dIII (Takara), respectively. The products were separated by size in 0.8% agarose gel in 1× TAE buffer (0.04 M Tris acetate and 0.001 M EDTA), and then transferred to Nylon membrane (Roche, Indianapolis, IN, USA). The DNA blot was hybridized with 340 bp of the 3′ end *BnaC.GL2.b* cDNA sequence labeled with digoxigenin (Roche) by PCR using the primers SC1882F: (5′CGAGAGAAAAGAGCACATG′3) and SC2222R: (5′ATGTTGTGGAGCGTGACA′3). The Dig-labeled probe sequence is highly conservative among the four homologous *GL2* genes in the *Brassica* species.

Additionally, the four *GL2* genomic DNA sequences were used as query in BLAST comparison ($E < 10^{-5}$) against the comprehensive *Brassica* genomic DNA database in our group, which contain whole gDNA sequences of *B. rapa* (Chiifu) and *B. oleracea* ('267') completely sequenced by a whole genome sequencing (WGS) approach (data unpublished). The sequences were aligned using a local implementation of the NCBI standalone Blast software (packageblast-2.2.18-ia32-win32.exe, http://www.ncbi.nlm.nih.gov/BLAST/download.shtml).

RT-PCR detection of transcription levels of the four *GL2* genes in different organs

To study expression patterns of the four *GL2* genes, semi-quantitative RT-PCR was adopted for representatives of three *Brassica* species (Zhongshuang 9, Yuyiganlan, Yellow Sarson). Transcription levels were detected in samples from 4-week-old roots, stems, rosette leaves, flowers, 4 days after pollination (4-DAP) siliques, and mature seeds 30-DAP. Total RNA extraction, redigestion with DNase I and first-strand cDNA synthesis were carried out as described by Wang et al. (2009). Quantitative uniformity of the cDNA samples was assessed by PCR using the control primers of *Bnactin* (*B. napus actin*, AF1118122), *Bractin* (*B. rapa actin*, EX087730), and *Boactin* (*B. oleracea actin*, AF044573) (Table 1). The expression patterns of *BnaA.GL2.a* and *BnaC.GL2.b* were detected by RT-PCR with primers PBnGL2A and PBnGL2C, respectively (Table 1). The transcription levels of *BraA.GL2.a* and *BolC.GL2.a* were detected with primers PBGL2 (Table 1). All RT-PCRs were replicated twice.

Generation of *BnaC.GL2.b* overexpression and antisense transgenic *Arabidopsis* plants

Based on the observed preferential expression of the C-genome *GL2* genes in seeds, *B. napus BnaC.GL2.b* was selected to study further the effect on seed oil content in *Arabidopsis* by overexpression and antisense knockdown strategies. A full-length *BnaC.GL2.b* cDNA fragment (2,253 bp) was PCR-amplified using primers PGL2Tr (Table 1) with Ex *Taq* DNA polymerase (Takara), and subcloned into vector pGΩ4A containing two enhancers, a double CaMV 35S promoter, the lead sequence of tobacco mosaic virus (*germin*), and the CaMV Nos terminator (Wang et al. 2009). The resulting constructs harbored the *BnaC.GL2.b*-coding region either in a sense or antisense orientation, which was confirmed by digesting with *Xba*I/*Hin*dIII. The construct was transformed into *Agrobacterium tumefaciens* (EHA105), together with the pSoup helper plasmid, for *Arabidopsis* transformation (Vain et al. 2003). To confirm positive transgenic lines, Basta resistant T_2 plants were analyzed using two strategies: (1) genomic DNAs were used as templates for PCR analysis by primers PK (Table 1) designed based on the *BnaC.GL2.b*-coding sequence, in order to confirm the integration of the T-DNA into the *Arabidopsis* genome; and (2) RNA was isolated from young leaves of putative transgenic plants and analyzed by Semi-quantitative RT-PCR to test for expression of *B. napus BnaC.GL2.b* (primers BnaC.GL2.b) (Table 1). Real-time RT-PCR was used to detection of *Arabidopsis* endogenous *GL2* (At1g79840, primers GL2) expression by a Chromo4 real-time PCR detection system (Bio-Rad) and SYBR Premix Ex *Taq* (Takara). *Atactin* (AT3G18780, primers Atactin) was used as the control gene of internal reference. Relative mRNA levels were calculated by iQ5 software (Bio-Rad) and normalized to the concentration of *Atactin* mRNA. Quantification of each cDNA sample was examined three times and the averaged data were analyzed statistically.

Arabidopsis seed oil measurement

Arabidopsis wild-type (Col-0) and T_2 transgenic plants were cultivated and harvested under similar growth conditions. Seed oil content was measured according to the method described by Wei et al. (2008). Briefly, a small quantity of dried sample (<20 mg) was quantified, crushed and extracted by petroleum ether using an ultrasound cleaning bath KQ-500DB (Ultrasound instrument Co., Ltd). After addition of internal standard (methylheptadecanoate, C17: 0), KOH–methanol solution was added to the extracted sample for fatty acid methyl ester (FAME) preparation. Following addition of distilled water, the sample was shaken and centrifuged. The supernatant was used for

GC analysis (Agilent 7890N gas chromatograph). The relative peak areas (analyte area/IS area) were used for quantification of fatty acid content, based on response factors (Christie WW 1991). The total fatty acid content was converted to the total oil content by converting the forms of the FAMEs to their relative forms of the fatty acid glycerides (Li et al. 2006). All samples were replicated at least three times.

Genome-specific primer design and identification of informative restriction sites

The differences between the four *GL2* nucleotide sequences allowed us to design three pairs of primers to identify and distinguish between copies associated with each of the two *Brassica* genomes. In addition to a pair of exon-anchored primers (Primer10 and 11), genome-specific primer sets were designed for the A-genome (primer: BnGL2A) and C-genome (BnGL2C), respectively (Table 1). PCR was performed on genomic DNA from the nine cultivars of three *Brassica* species mentioned above.

A genome-specific transition substitution present in the four *GL2* genes was used to generate another functional marker. *GL2* partial genome fragments including the substitution were amplified from Zhongshuang 9 genomic DNA using primers 14 and 9, and from Yellow Sarson or Yuyiganlan using primers 8 and 9 (Table 1). PCR products were subsequently subcloned into PMD18-T that possessed an *Eco*RI site, but not an *Eco*RV site. Six independent positive clones (*B. napus*) and three others (*B. rapa* or *B. oleracea*) were selected and sequenced. After excluding clones having identical sequences, four representative clones named as 'ZS9-1' (for *BnaC.GL2.b*), 'ZS9-2' (*BnaA.GL2.a*), 'Yuyi' (*BolC.GL2.a*) and 'YS' (*BnaA.GL2.a*) were digested by *Eco*RI/ *Eco*RV.

Results

Cloning of four homologous *GL2* genes

Based on similarity to the *Arabidopsis GL2* sequence, a clone was identified from a *B. napus* cDNA library that contained an intact ORF and encoded a GL2 protein, BnaC.GL2.b, similar to that of *Arabidopsis*.

To determine the complete translation regions of the *B. napus GL2* paralogues, 5′ non-coding regions were identified by the Genome Walking method, based on the coding sequence of *BnaC.GL2.b*. The presence of two different amplification fragments in the 5′ non-coding regions suggested that *B. napus* may possess at least two *GL2* paralogues. Although the two fragments had high sequence similarity, they were different in length (731 and 500 bp). Promoter prediction using NNPP (Neural Network Promoter Prediction: http://www.fruitfly.org/seq_tools/promoter.html) indicated that both fragments include transcription start sites (TSS) and some basal elements of the promoter including TATA-box, CAAT-box and G-box. Within the 3′ non-coding regions, *BnaC.GL2.b* from the cDNA library possessed a 162-bp fragment containing a poly (A) tail following the stop codon TGA.

We then developed a novel PCR-based strategy to clone three homologous *GL2* genes (*BnaA.GL2.a*, *BraA.GL2.a*, and *BolC.GL2.a*) from three *Brassica* species. This was based on the sequence obtained from the clone (*BnaC.GL2.b*) and the two non-coding regions obtained from genome walking. To clone *GL2* homologous genes from *B. napus*, we designed 15 primers and amplified 10 short fragments, which covered the whole *GL2* genomic DNA region, at least three times for each intron (Fig. 1). Sequences of the PCR fragments were oriented and assembled to form the two *GL2* homologous genes. Using the same PCR strategy, we also acquired the *GL2* genes from *B. rapa* and *B. oleracea*, with each containing a 14-bp fragment of the 5′ UTR and a 16-bp fragment of the 3′ UTR.

Characterization of *Brassica GL2* orthologues

Alignment of the nucleotide sequences of *BnaA.GL2.a*, *BnaC.GL2.b*, *BraA.GL2.a* and *BolC.GL2.a* indicated that all four genes have identical structures which contain nine exons and eight introns similar to the known HD-ZIP IV genes of *Arabidopsis* (SFig. 1 in Electronic Supplementary Material, Fig. 2). Although the four genes have a high similarity the predicted exons are more conserved than the introns (Table 2). Dinucleotide sequences at the 5′- and 3′-ends of all introns strictly follow the universal GT-AG rule (Breathnach and Chambon 1981). Sequence of the C-genome orthologue of *B. napus* (*BnaC.GL2.b*) is most similar to *Arabidopsis GL2* (Table 2).

The most striking differences between the four orthologues are the high variability of intron 5 (nucleotide position: 2,100–2,400 bp) and intron 7 (3,100–3,300 bp), which

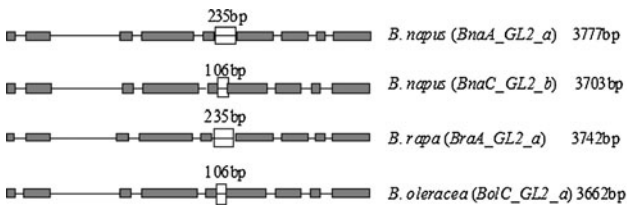

Fig. 2 Summary of the structures of the four *Brassica GL2* genes. *Box* represents the fifth intron in each gene, and the corresponding sequence length is indicated *above the box*. These *GL2* sequences have been deposited in GenBank with accession numbers EU826520 (*BnaA.GL2.a*), EU826521 (*BnaC.GL2.b*), EU826522 (*BraA.GL2.a*) and EU826523 (*BolC.GL2.a*)

Table 2 Homologies of genomic DNA and amino acid sequences within four *Brassica GL2* and *Arabidopsis GL2* genes

Gene pair	Genomic DNA sequence similarity (%)			Amino acid sequence similarity (%)
	Start codon to stop codon	Exon region	Intron region	
BnaA.GL2.a/BnaC.GL2.b	91.64	96.99	83.65	98.40
BnaA.GL2.a/BraA.GL2.a	96.52	98.76	93.84	98.67
BnaA.GL2.a/BolC.GL2.a	88.01	96.76	75.45	98.27
BnaC.GL2.b/BraA.GL2.a	89.09	96.10	78.18	97.34
BnaC.GL2.b/BolC.GL2.a	95.84	99.69	89.95	99.87
BnaA.GL2.a/GL2	70.53	87.58	37.05	90.87
BnaC.GL2.b/GL2	72.20	87.67	45.08	91.40
BraA.GL2.a/GL2	70.89	87.57	40.90	90.46
BolC.GL2.a/GL2	72.12	87.65	46.26	91.27

were reflected by their nucleotide diversity (Pi) values (Fig. 3). The GC contents of intron 5 and the corresponding *GL2* genomic DNA are 27.7 and 41.9% in *BnaA.GL2.a*, 28.3 and 42.3% in *BnaC.GL2.b*, 27.7 and 42.5% in *BraA.GL2.a*, and 28.3 and 42.5% in *BolC.GL2.a*. As expected, the GC content of intron 5 is much less than that of the corresponding genomic DNA in each gene. Interestingly, the 235-bp intron 5 of *B. napus BnaA.GL2.a* is identical with the fifth intron of *B. rapa BraA.GL2.a*, but extremely diverged (only 24% identity) from the 106-bp intron 5 of the two *GL2* genes from *B. oleracea* origin (Fig. 2).

Sequence alignment of the complete genes revealed greater similarity within two orthologues from either A or C origin than between homoeologues from the different genomes. Sixty-three genome-specific single nucleotide polymorphisms (SNPs) (7 transversions, 50 transitions, 6 insertions /deletions) were detected in the *GL2* coding regions, which resulted in eleven amino acid substitutions (Table 3). 82% of the genome-specific substitutions are synonymous mutations (SFig. 1).

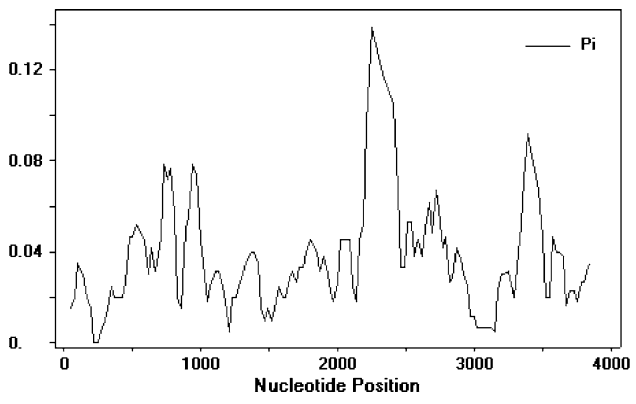

Fig. 3 Genetic diversity of the four *Brassica GL2* genes. The most variable region is present at the position of 2,100–2,400 bp (calculated from the start codon), and the corresponding nucleotide diversity (Pi) value is nearly 0.14

Analysis of putative *Brassica* GL2 proteins

The proteins deduced from the *GL2* exon sequences all have 750 amino acids, with predicted molecular weights and isoelectric points values of 83.76 kDa and 7.00 for *BnaA.GL2.a*, 83.70 kDa and 7.02 for *BnaC.GL2.b*, 83.77 kDa and 6.57 for *BraA.GL2.a*, and 83.65 kDa and 7.28 for *BolC.GL2.a*.

Comparison with other representative HD-Zip IV proteins (*Arabidopsis* GL2, *Phalaenopsis* O39, rice ROC1 and maize ZmOCL1) suggested that the four proteins were typical of HD-ZIP IV proteins described by Nakamura et al. (2006) (Fig. 4a, b). In particular, the HD is N-terminally located, in contrast to most HD proteins, where the HD is located in the middle or in the C-terminal region (Chan et al. 1998). In the HD, the sequence of the third helix region (QVKFWFQNRRTQIK) is strictly conserved, which provides sequence-specific interaction with DNA (Gehring et al. 1994) (Fig. 4b). A basic amino acid cluster (RKRKK) acts as a nuclear localization signal (PredictNLS server: http://cubic.bioc.columbia.edu/predictNLS/) in the N-terminal end of HD (SFig. 1), similar to that reported by Hicks et al. (1995). Following the conserved HD, two successive truncated leucine zipper motifs (LZ), including a loop and a STAR-related lipid transfer (START) domain, were found. Additionally, SignalP 3.0 (http://www.cbs.dtu.dk/services/SignalP/) predicted no signal peptides in the four GL2 proteins.

A phylogenetic tree based on the sequence similarities of nine taxonomically diverged HD-ZIP IV proteins generated a clear division into monocot and dicot origin. As expected, the *Arabidopsis GL2* gene is the most closely related to four *Brassica GL2* genes (Fig. 4c).

Table 3 A/C-genome-specific amino acids variation loci in four *Brassica* GL2 proteins

Gene	Amino acid variation loci[a]										
	238	269	404	412	418	419	477	660	697	709	745
BnaA.GL2.a	Gln	Leu	Ala	Met	–	Ala	Lys	Met	Ala	Met	Cys
BnaC.GL2.b	His	Val	Val	Val	Thr	Ser	Arg	Leu	–	Leu	Ser
BraA.GL2.a	Gln	Leu	Ala	Met	–	Ala	Lys	Met	Ala	Met	Cys
BolC.GL2.a	His	Val	Val	Val	Thr	Ser	Arg	Leu	–	Leu	Ser

[a] Variation loci based on amino acid sequence of *BnaA.GL2.a*

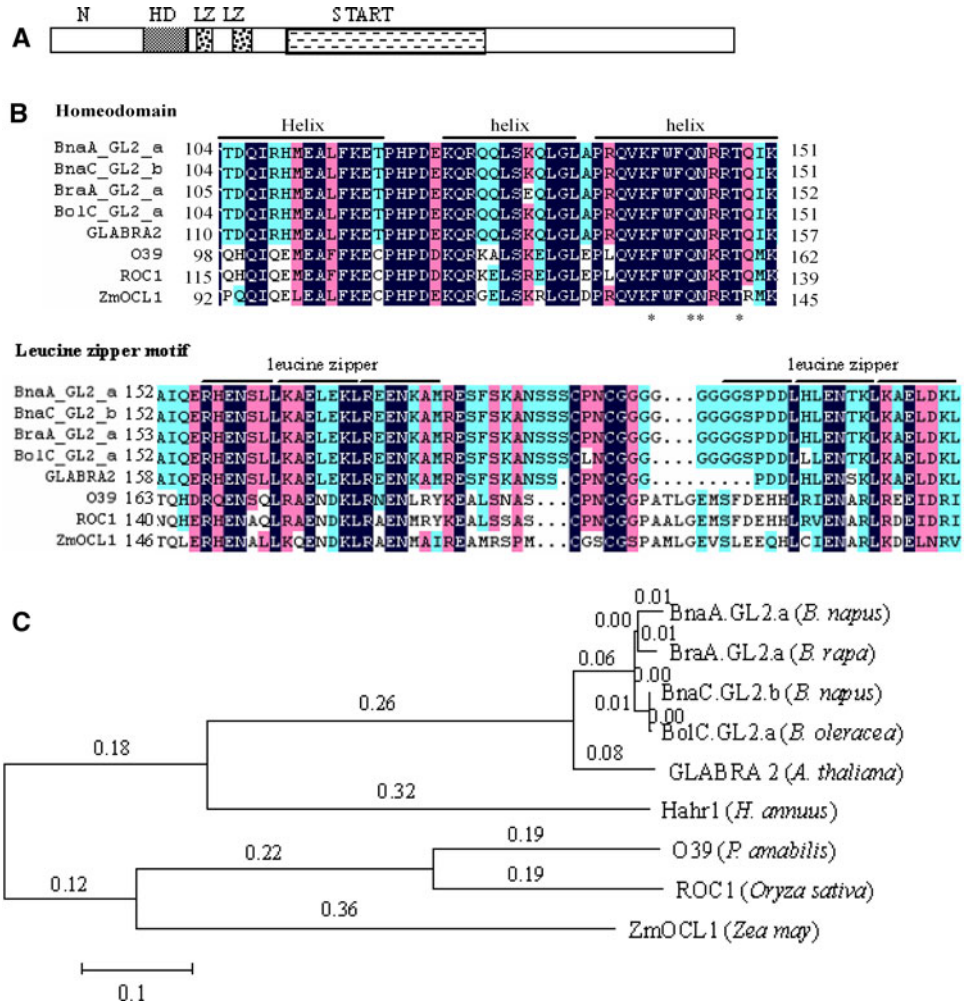

Fig. 4 The structure of the four GL2 proteins and sequence comparison within the HD-GL2 family. **a** Schematic diagram illustrating the domains conserved among the HD-GL2 members. Conserved domains, including acidic amino acid region (*N*), homeodomain (*HD*), two successive truncated leucine zipper motifs (*LZ*) and START domain (*START*), are marked. **b** Alignment of the sequences in the HD, leucine zipper motif of the four GL2 proteins with the corresponding sequences of *Arabidopsis* GLABRA2, *Phalaenopsis* O39, Rice ROC1 and Maize ZmOCL1. Identical and similar amino acids are marked. *Asterisks* denote potential DNA sequence-specific contact residues. **c** Phylogenetic tree showing the predicted relationships among the representative HD-GL2 proteins. The four *Brassica* GL2, GLABRA2, O39, ROC1, ZmOCL1 and Sunflower Hahr1 were included. The tree was constructed by using the neighbor-joining method. The branch lengths are proportional to the sequence divergence. *Numbers* indicate genetic distance. GenBank accession number of above-mentioned proteins: GLABRA2 (L32873), O39 (U34743), ROC1 (AB077993), ZmOCL1 (Y17898), and Hahr1 (L76588)

Copy numbers of *Brassica GL2* gene family

To determine copy numbers of *GL2* gene family in the three *Brassica* species, this study carried out Southern blot using a conservative 340-bp C-terminal region fragment of *Brassica GL2* cDNA sequence as the hybridization probe, and exhaustive blast analysis against the whole genomic DNA sequences of *B. rapa* and *B. oleracea* (unpublished

sequence data), respectively. The result of Southern blot showed that both *Eco*RI and *Hin*dIII digestions gave rise to two hybridized bands in genomic DNA of *B. napus*, one in either *B. rapa* or *B. oleracea*, indicating two copies of *GL2* genes in *B. napus* and one in either *B. rapa* or *B. oleracea* (Fig. 5), which was the same to the result of blast analysis against the whole genomic DNA sequences of *B. rapa* and *B. oleracea* (SFig. 2). Furthermore, corresponding orthologues from *B. napus*, *B. oleracea* or *B. rapa* showed the same or very similar size in sequence length (indicated with arrows in Fig. 5), which verified the origin of *BnaA.GL2.a* and *BnaC.GL2.b* in *B. napus*.

Organ-specific transcription of *Brassica GL2* homoeologues

Previous analysis of *Arabidopsis GL2* transcription level suggested that the *GL2* gene was universally expressed in different tissues during the whole life cycle (Nakamura et al. 2006). Here we used semi-quantitative RT-PCR to detect expression levels of the four *Brassica GL2* genes in six organs (root, stem, leaf, silique, flower and seed). As expected, the orthologues were transcribed in a wide range of organs. However, expression patterns were different between orthologues of the A-genomes and the C-genomes, and between the two orthologous *GL2* genes of *B. napus*.

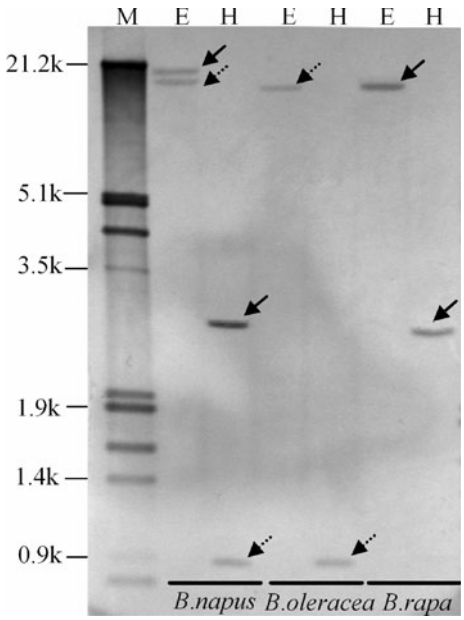

Fig. 5 Southern blot detection for members of *GL2* gene family in *B. napus*, *B. rapa*, and *B. oleracea*. Twenty-five micrograms of genomic DNAs were digested with *Eco*RI (*E* in figure) and *Hin*dIII (*H*), respectively, and hybridized by a cDNA fragment of 340 nucleotides conservative among *Brassica GL2* orthologues. The *arrows* point to the bands of *B. napus* at the similar marker positions where their orthologues in *B. oleracea* and *B. rapa* were located. *M* lambda DNA/*Eco*RI + *Hin*dIII marker (Fermentas Biotech)

The transcription of the two A-genome *GL2* orthologues was universally enriched in roots and relatively weaker in 30-DAP seeds. In contrast, the two C-genome *GL2* orthologues had higher levels of transcription in 30-DAP seeds and 4-DAP siliques, and lower levels in roots and rosette leaves (Fig. 6).

Modulation of seed oil content by overexpression and suppression of *BnaC.GL2.b*

A T-DNA insertion *gl2* mutant (a glabrous mutant of *gl2-2*) had previously been shown to increase seed oil content in *Arabidopsis*, in addition to positively controlling trichome development (Shen et al. 2006). Our phylogenetic analysis indicated that *B. napus BnaC.GL2.b* is the most similar orthologue to *Arabidopsis GL2*. To investigate whether *BnaC.GL2.b* also affects seed oil content, we generated *Arabidopsis BnaC.GL2.b* overexpression and suppression transgenic plants. Multiple independent Basta resistant T_2 transgenic plants were obtained and preliminarily verified positive by PCR amplification (data not shown). Ectopic overexpression of *BnaC.GL2.b* driven by 35S promoter in the wild-type *Arabidopsis* generated two phenotypes: the wild-type-like and the *gl2*-mutant-like (an *Arabidopsis* glabrous mutant of *gl2-2*). Suppression of *BnaC.GL2.b* in antisense transgenic plant reduced trichome number at the leaf margins (Fig. 7a). Based on their phenotype, nine transgenic plants were selected for further detection of *B. napus BnaC.GL2.b* transcription levels by semi-quantitative RT-PCR. The results indicated that *BnaC.GL2.b* was drastically detected in T2 transgenic plants (Fig. 7b). Real-time quantitative RT-PCR analysis of *Arabidopsis* endogenous *GL2* expression showed that their decreased transcription levels were found both in the

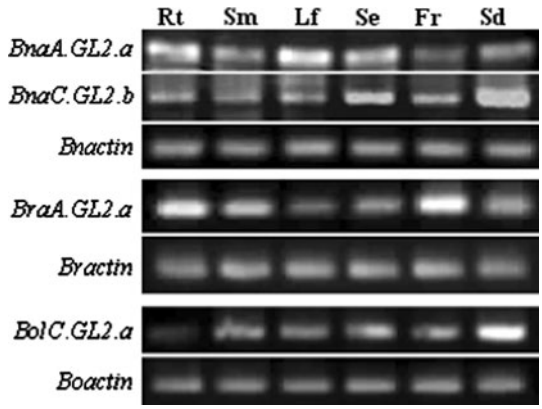

Fig. 6 Semi-quantitative RT-PCR detection of transcription levels of the four *GL2* genes in six various organs. *Bnactin* (*B. napus actin*, AF1118122), *Bractin* (*B. rapa actin*, EX087730) and *Boactin* (*B. oleracea actin*, AF044573) were used as the internal controls, respectively. *Rt* Root, *Sm* stem, *Lf* rosette leaf, *Se* 4-DAP (days after pollination) silique, *Fr* flower, *Sd* 30-DAP seeds

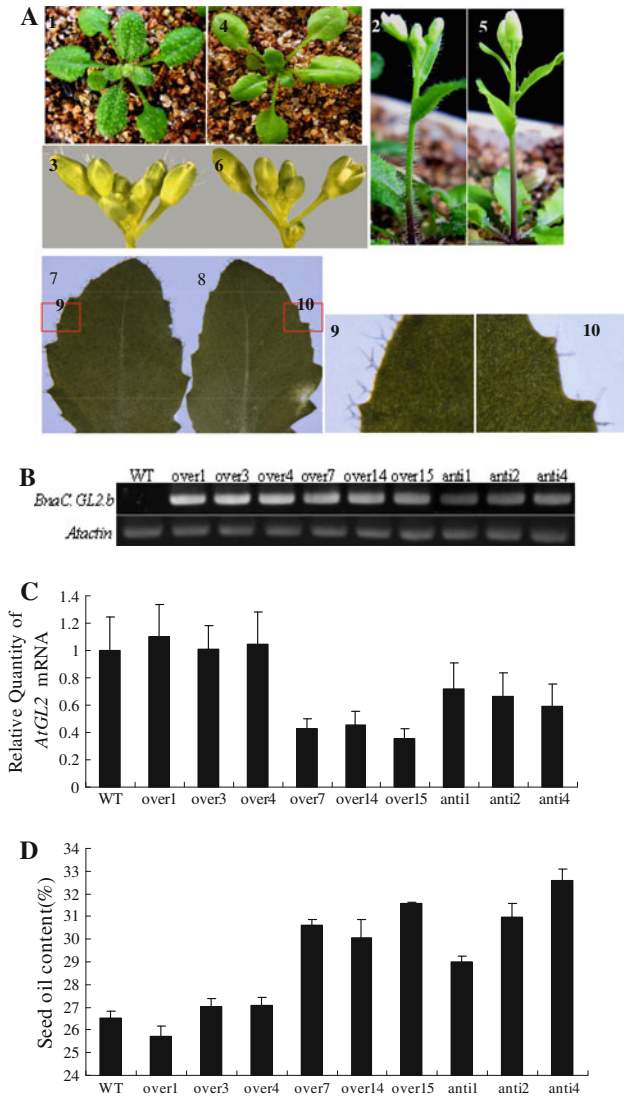

Fig. 7 Phenotypes identification and molecular detection of *BnaC.GL2.b* T_2 transgenic plants. *Arabidopsis* was transformed with *B. napus BnaC.GL2.b* driven by 35S promoter in wild-type genetic background. *WT* Wild-type (Col-0); wild-type-like overexpression transgenic plants: over1, over3, over4; *gl2*-mutant-like overexpression plants: over7, over14, over15; antisense plants: anti1, anti2, anti4. **a**. Trichome of wild-type and *BnaC.GL2.b* transgenic plants. Overexpression transgenic plants with wild-type-like (*1*, *2*, *3*) or *gl2*-mutant-like (*4*, *5*, *6*) phenotypes. Seedlings (*1*, *4*), stems (*2*, *5*) and buds (*3*, *6*) are shown. The rosette leaves of wild-type (*7*) and antisense transgenic plant (*8*). *Red boxes* (9 and 10) indicate magnified leaf trichomes. **b** Semi-quantitative RT-PCR detection of *B. napus BnaC.GL2.b* expression in wild-type and *BnaC.GL2.b* transgenic plants. *Atactin* (AT3G18780) was used as the internal control. **c** Real-time quantitative RT-PCR analysis of *Arabidopsis GL2* expression in wild-type and *BnaC.GL2.b* transgenic plants. Expression levels are given in relative units, with the wild-type level serving as 1 unit. The mean ± SD (three replicates) is indicated for each sample. **d** Seed oil content of wild-type and *BnaC.GL2.b* transgenic plants. *Bars* are standard deviation

gl2-mutant-like overexpression plants and the antisense transgenic plants, but not in WT and the wild-type-like overexpression plants (Fig. 7c).

In agreement with the transcription levels, seed oil content had a consistent increase of more than 3% in the *gl2*-mutant-like overexpression and antisense lines (Fig. 7d). ANOVA analysis indicated a significant difference ($P < 0.01$) in oil content between wild-type *Arabidopsis* (26.54 ± 0.301%) and either the *gl2*-mutant-like overexpression or antisense lines. There was no detectable difference in seed oil content between the wild-type-like overexpression lines and WT. The fatty acid composition as detected by FAME analysis showed no changes in transgenic and wild-type plants (data not shown). No obvious differences between T_2 transgenic plants and *Arabidopsis* wild-type plants were observed in other seed characteristics such as seed size and seed number. Our results indicated that suppression with *B. napus BnaC.GL2.b* can decrease trichome number at the leaf margin, and increase seed oil content in *Arabidopsis*.

Development of functional markers

Based on comparison of the four *GL2* sequences, we designed A- and C-genome-specific PCR primers and identified a genome-specific restriction enzyme site, which can both be used as functional markers to distinguish these *Brassica* homoelogues and A- and C-genomes. Analysis of the *GL2* genes indicated that intron 5 and intron 7 were more variable and genome-specific, and could be used to design three primer pairs (exon-anchored, A- and C-genome-specific) to amplify from genomic DNA of *B. napus*, *B. rapa* and *B. oleracea* (Fig. 8a, b). When a pair of exon-anchored primers spanning three introns (intron 4, 5, 6) was used, we observed two bands in each *B. napus* cultivar. However, only one was observed in either *B. oleracea* or *B. rapa* (Fig. 8a). This further confirmed that each of the two genes present in the amphidiploid *B. napus* is represented by one counterpart copy in each of the two progenitor diploid species. When using a pair of A-genome-specific primers, a unique fragment was amplified from both *B. napus* and *B. rapa*, but no amplified product was obtained from any *B. oleracea* cultivar. For the pair of C-genome-specific primers, only one fragment was amplified from *B. napus* and *B. oleracea*, and none from *B. rapa* (Fig. 8b). Thus these *GL2* genome-specific PCR primers, as functional molecular markers, can effectively identify and distinguish these *GL2* orthologoues and also the *Brassica* A- and C-genomes.

An informative transition substitution (A → C) was found in the four *GL2* genes: adenine (A) was present in *BnaA.GL2.a* (at position 3486) and *BraA.GL2.a* (at position 3451) of the A-genomes, whereas cytosine (C) was present in *BnaC.GL2.b* (at position 3415) and *BolC.GL2.a* (at position 3374) of the C-genomes (SFig. 1). This variation created an *Eco*RV (GATATC) cleavage site in *BnaC.GL2.b*

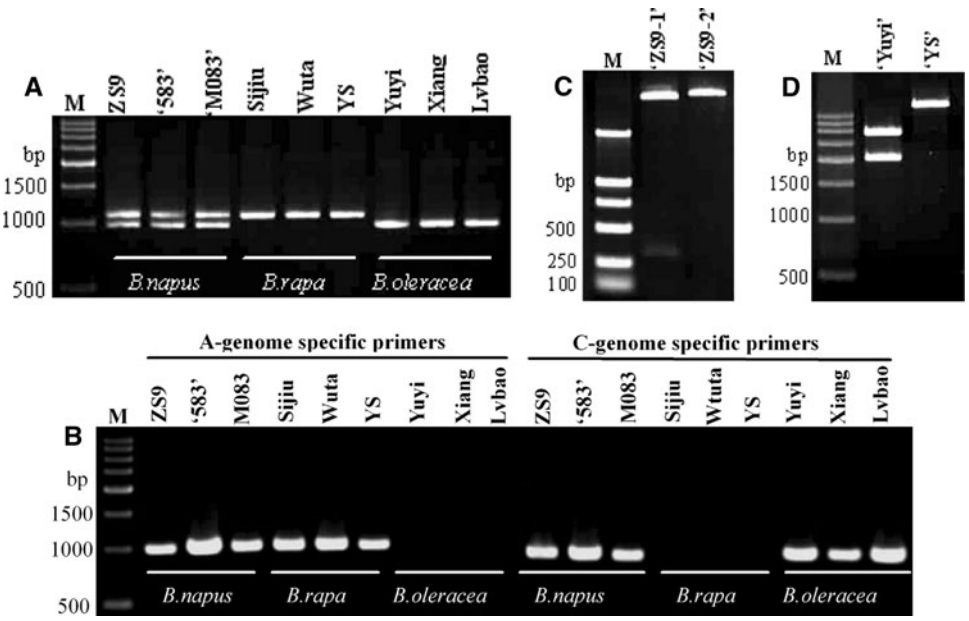

Fig. 8 Gel electrophoresis of PCR amplification and enzyme cutting results. In PCR (**a**, **b**), each primer pair was used to amplify different *Brassica* species. *B. napus*: ZS9 (ZhongShuang 9), '583' and M083; *B. rapa*: Sijiu (Sijiucaixin), Wuta (Wutacai) and YS (Yellow Sarson); *B. oleracea*: Yuyi (Yuyiganlan), Xiang (Xianggangbaihuajielan) and Lvbao (Lvbaojielan). *M* 500 bp ladder Marker. In enzyme cutting analysis (**c**, **d**), four representative clones amplified from the corresponding *GL2* genomic DNA were digested by *Eco*RI/ *Eco*RV, respectively. **a** An exon-anchored primer pair. **b** *Lanes 2*, *3*, *4*, *5*, *6*, *7*, *8*, *9*, *10* for the A-genome-specific primers, and *lanes 11*, *12*, *13*, *14*, *15*, *16*, *17*, *18*, *19* for the C-genome-specific primers. **c** Two clones amplified from *GL2* genomic DNA of *B. napus* (ZhongShuang 9). 'ZS9-1', a clone including the fragment of *BnaC.GL2.b*; 'ZS9-2', a clone including the fragment of *BnaA.GL2.a*. *M* DL2000 marker. **d** Two clones amplified from *GL2* genomic DNA of *B. oleracea* Yuyiganlan and *B. rapa* Yellow Sarson, respectively. 'Yuyi', a clone with the fragment of *BolC.GL2.a*; 'YS', a clone with the fragment of *BraA.GL2.a*. *M* 500 bp ladder marker

and *BolC.GL2.a*, whereas there is no such site in the *GL2* sequence of both A-genomes. Based on these sequence characteristics, *GL2* partial genome fragments incorporating this substitution were amplified in three *Brassica* species and subcloned into pMD18-T. Four representative clones ('ZS9-1' with a 1,212-bp fragment amplified from *BnaC.GL2.b*, 'ZS9-2' with a 1,164-bp fragment from *BnaA.GL2.a*, 'Yuyi' with a 1,162-bp fragment from *BolC.GL2.a*, and 'YS' with a 1,213-bp fragment from *BraA.GL2.a*) were digested with *Eco*RI/*Eco*RV, respectively. The results demonstrated that full digestion was achieved for 'ZS9-1' and 'Yuyi' from the C-genomes, whereas 'ZS9-2' and 'YS' from the A-genomes were resistant to digestion with *Eco*RV (Fig. 8c, d). These results indicated that the *Eco*RV cleavage site in the *GL2* gene sequences could also be used as a rapid practical marker for distinguishing these homoeologues as well as distinguishing between the *Brassica* A- and C-genomes.

Discussion

In this paper, we isolated four full-length *GL2* genes from *B. napus* and its progenitor diploid relatives by an overlapping-PCR strategy. Comparative analysis of exon sequences amplified from genomic DNA and the corresponding cDNA sequences showed only a few nucleotide differences between them. Blast analysis of four *GL2* genomic DNA sequences against our *Brassica* whole genome database and Southern blot both indicated that there were just one *GL2* orthologue in either *B. rapa* or *B. oleracea* and two in *B. napus*. The four *Brassica GL2* genes were sequentially divided into two groups and two orthologous genes from the same genomic origins (A or C) are more similar than those from two different genomic origins (A and C), as is the case with the *ALC* loci in three *Brassica* species, although the presence of three homologous *ALC* paralogues in *B. napus* predicted by southern blot (Hua et al. 2009). Sequence alignment indicated that the four GL2 proteins belong to the HD-ZIP IV subfamily and feature a set of distinctive features such as DNA-binding specificities, gene structures, and additional common motifs. The HD-Zip proteins are widely present in the plant kingdom, and conserved in the process of evolution (Ariel et al. 2007). Therefore, the phylogenetic tree of *GL2* genes further confirmed that *B. napus* originated from the natural combination of chromosomes in hybrids between its parental genomes. We believe the PCR strategy described here is generically applicable for cloning multiple homologous genes in polyploid plants and their progenitors.

Previous reports suggested seed oil content can be regulated by modifying key enzymes such as ACCase and DGAT in oilseed rape lipid biosynthesis (Roesler et al. 1997; Weselake et al. 2007). However, more recent studies have indicated that seed oil accumulation appears to be controlled by a coordinated regulatory mechanism, which is not only pertinent to major steps of lipid metabolism pathways, but also requires coordination of key components in carbohydrate metabolism, in particular the regulation of sucrose and hexose flux (Mu et al. 2008). Therefore, identification of the key genes involved in seed oil content at the level of carbohydrate metabolism represents another strategy for increasing seed oil bioassembly. Screens of *Arabidopsis* mutants have identified a series of important TFs, which can significantly affect seed oil content. Among these TFs, LEC1 and WRI1 act as master regulators because they appear to regulate the action of other TFs (Weselake et al. 2009). The *LEC1* gene encodes the CCAAT box-binding domain protein and controls seed oil biosynthesis by inducing expression of more than half of the plastidial fatty acid synthesis genes in *Arabidopsis* (Mu et al. 2008). The *WRI1* gene encodes an AP2/EREB (ethylene responsive element binding) domain protein and promotes oil accumulation by controlling the carbon flux and sugar metabolism in *Arabidopsis* (Cernac and Benning 2004). In addition to these two TFs, *Arabidopsis GL2* gene has been confirmed to play a key role in seed oil biosynthesis (Shen et al. 2006). In this paper, four *Brassica GL2* genes transcription levels detected by semi-quantitative RT-PCR showed that the two orthologous *GL2* genes from the same genomic origin (A or C) had a more similar expression pattern when compared to those arising from the different genomic origin. Hua et al. (2009) also reported two types of different expression patterns within the *ALC* homologous genes from three *Brassica* species. It is well known that cis-regulatory region of the gene is important for determination of its expression pattern. Compared to those of paralogues, cis-regulatory sequences of the orthologous genes exhibit dramatically slower rates of evolution, which makes their gene expression patterns more conservative (Castillo-Davis et al. 2004). Thus, sequence characteristics among the cis-regulatory regions of the four *GL2* orthologues could be the cause of two types of expression patterns. Generally speaking, orthologous genes are expected to retain similar (if not identical) functions across different species, whereas paralogues may more readily acquire novel functions within a genome (Sonnhammer and Koonin 2002; Chen et al. 2006). We therefore infer that the *GL2* orthologues from the A-genome may participate in root development, whilst the *GL2* orthologues from the C-genome may play an important role in regulating seed development.

Based on this hypothesis, we further studied the effect of *B. napus BnaC.GL2.b* on seed oil accumulation in overexpression and suppression transgenic *Arabidopsis* plants. In T_2 antisense transgenic plant, suppression of *BnaC.GL2.b* drastically increased oil accumulation, which suggested that BnaC.GL2.b protein could substitute *Arabidopsis* GL2 and regulate oil content in *Arabidopsis*. Surprisingly, some of *BnaC.GL2.b*-overexpressed plants also increased seed oil content, in addition to suppression of trichome cell differentiation, in transgenic *Arabidopsis*. The reason for the fact that the *35S:: BnaC.GL2.b* interrupted *Arabidopsis* endogenous *GL2* function is unclear. Previous reports indicated that the ectopically expressed *Arabidopsis GL2* or cotton *GaHOX1* both repressed the trichome development. One possible mechanism for such phenotype is 'squelching', in which overexpression of *GL2* and *GL2*-like genes might affect normal GL2 function by reducing the ratio of the heterodimers to the GL2 homodimer (Ohashi et al. 2002; Guan et al. 2008).

Conventional molecular markers used in oilseed rape have been based on genomic DNA sequence rather than the actual genes. In contrast, 'FMs' combine DNA polymorphisms and trait variation, which can rapidly transfer a character into an otherwise desirable genotype for genetics research and breeding (Rahman et al. 2008). Due to the high homology of *Brassica* A- and C-genomes, it is not easy to identify and distinguish *B. napus* functional paralogues and the genome types. Therefore, cloning and functional identification of important orthologous genes and development of FMs that allow mapping of the individual loci can be advocated. Wu et al. (2007) and Rahman et al. (2008) separately reported FMs specific for erucic acid at the *FAE1* loci, which confirmed that the usage of FMs is feasible in oilseed rape. In this paper we have evidence suggesting that the two *B. napus GL2* homologues might have different functions, with *BnaC.GL2.b* potentially being effective in regulating seed oil content in oilseed rape. Analysis of the four *GL2* sequences showed that both of intron 5 and 7 have A- and C-genome-specific features (Fig. 2), which allowed us to design three genome-specific primer sets, as the FMs, to distinguish these *Brassica* orthologues. The divergence of these genome-specific sequences also provides an opportunity for targeting-induced local lesions in genomes (TILLING) or EcoTILLING analysis of allelic DNA polymorphism at the *GL2* loci in a population using artificial mutagenesis or a natural germplasm (Wang et al. 2008). Besides PCR-based markers, 63 A/C-genome-specific SNPs were also identified in the *GL2* coding regions, of which, 79.37% belong to the transition type and are higher than transversion frequency (11.11%), similar to the result of Fang et al. (2006). Among these SNP loci, 82.15% of the genome-specific substitutions are synonymous mutations, which can be used as allele-specific PCR to perform genetic mapping, or to identify genome-specific restriction endonucleases sites to distinguish the

GL2 orthologues (see SFig. 1). The present study identified from these SNPs an *Eco*RV cleavage site unique to the C-genome, which effectively distinguished the four orthologues and *Brassica* A/C-genome. Thus the two types of functional markers exploited in this study offer another approach to accelerate breeding selection for seed oil content, as well as to identify interspecific hybrids and intergenomic gene flow of *Brassica*.

Acknowledgments This work was financially supported by National Basic Research Program of China (2006CB101600), the Hi-Tech Research and Development Program of China (2006AA10A112) and the National Natural Science Foundation of China (30671344). GJK was supported in the UK by the Biotechnology and Biological Sciences Research Council. Thanks are given to Juping Fang, Shunmo Huang, Xianhu Fu, Yueying Liu and Xuejiang Zhang in our lab for their technical assistance. The authors thank Dr Yongju Huang (Rothamsted Research, Harpenden, UK) and Dr Daoxin Xie (Tsinghua University, Beijing, China) for their constructive suggestion to the manuscript.

References

Ariel FD, Manavella PA, Dezar CA, Chan RL (2007) The true story of the HD-Zip family. Trends Plant Sci 12:419–426

Baud S, Mendoza MS, To A, Harscoët E, Lepiniec L, Dubreucq B (2007) WRINKLED1 specifies the regulatory action of LEAFY COTYLEDON2 towards fatty acid metabolism during seed maturation in *Arabidopsis*. Plant J 50:825–838

Blake NK, Sherman JD, Dvořák J, Talbert LE (2004) Genome-specific primer sets for starch biosynthesis genes in wheat. Theor Appl Genet 109:1295–1302

Breathnach R, Chambon P (1981) Organization and expression of eucaryotic split genes coding for proteins. Annu Rev Biochem 50:349–383

Castillo-Davis CI, Hartl D, Guillaume Achaz (2004) cis-Regulatory and protein evolution in orthologous and duplicate genes. Genome Res 14:1530–1536

Cernac A, Benning C (2004) *WRINKLED1* encodes an AP2/EREB domain protein involved in the control of storage compound biosynthesis in *Arabidopsis*. Plant J 40:575–585

Chan RL, Gago GM, Palena CM, Gonzalez DH (1998) Homeoboxes in plant development. Biochem Biophys Acta 1442:1–19

Chen F, Mackey AJ Jr, Stoeckert CJ, Roos DS (2006) OrthoMCL-DB: querying a comprehensive multi-species collection of ortholog groups. Nucleic Acids Res 34:363–368

Christie WW (1991) Gas chromatographic analysis of fatty acid methyl esters with high precision. Lipid Technol 3:97–98

Fang JG, Twito T, Zhang Z, Chao CCT (2006) Genetic relationships among fruiting-mei (*Prunus mume* Sieb. et Zucc.) cultivars evaluated with AFLP and SNP markers. Genome 49:1256–1264

Finkelstein RR, Somerville CR (1990) Three classes of abscisic acid (ABA)-insensitive mutations of Arabidopsis define genes that control overlapping subsets of ABA responses. Plant Physiol 94:1172–1179

Focks N, Benning C (1998) *wrinkled1*: a novel, low-seed-oil mutant of *Arabidopsis* with a deficiency in the seed-specific regulation of carbohydrate metabolism. Plant Physiol 118:91–101

Gao MJ, Lydiate DJ, Li X, Lui H, Gjetvaj B, Hegedus DD, Rozwadowski K (2009) Repression of seed maturation genes by a trihelix transcriptional repressor in Arabidopsis seedlings. Plant Cell 21:54–71

Gehring WJ, Qian YQ, Billeter M, Furukubo-Tokunaga K, Schier AF, Resendez-Perez D, Affolter M, Otting G, Wüthrich K (1994) Homeodomain-DNA recognition. Cell 78:211–223

Guan XY, Li QJ, Shan CM, Wang S, Mao YB, Wang LJ, Chen XY (2008) The HD-Zip IV gene *GaHOX1* from cotton is a functional homologue of the *Arabidopsis GLABRA2*. Physiol Plant 134:174–182

Hajduch M, Casteel JE, Hurrelmeyer KE, Song Z, Agrawal GK, Thelen JJ (2006) Proteomic analysis of seed filling in *Brassica napus*. Developmental characterization of metabolic isozymes using high-resolution two-dimensional gel electrophoresis. Plant Physiol 141:32–46

Hicks GR, Smith HM, Shieh M, Raikhel NV (1995) Three classes of nuclear import signals bind to plant nuclei. Plant Physiol 107:1055–1058

Hua SJ, Shamsi IH, Guo Y, Park H, Chen MX, Shi CG, Meng HB, Jiang LX (2009) Sequence, expression divergence and complementation of homologous *ALCATRAZ* loci in *Brassica napus*. Planta 230:493–503

Ingouff M, Farbos I, Lagercrantz U, von Arnold S (2001) *PaHB1* is an evolutionary conserved HD-GL2 homeobox gene expressed in the protoderm during Norway spruce embryo development. Genesis 30:220–230

Ingouff M, Farbos I, Wiweger M, von Arnold S (2003) The molecular characterization of *PaHB2*, a homeobox gene of the HD-GL2 family expressed during embryo development in Norway spruce. J Exp Bot 54:1343–1350

Ingram GC, Magnard JL, Verges P, Dumas C, Rogowsky PM (1999) *ZmOCL1*, an HDGL2 family homeobox gene, is expressed in the outer cell layer throughout maize development. Plant Mol Biol 40:343–354

Ingram GC, Boisnard-Lorig C, Dumas C, Rogowsky PM (2000) Expression patterns of genes encoding HD-ZipIV homeodomain proteins define specific domains in maize embryos and meristems. Plant J 22:401–414

Ito M, Sentoku N, Nishimura A, Hong SK, Sato Y, Matsuoka M (2002) Position dependent expression of GL2-type homeobox gene, *Roc1*: significance for protoderm differentiation and radial pattern formation in early rice embryogenesis. Plant J 29:497–507

Kumar S, Tamura K, Jakobsen IB, Nei M (2001) MEGA2: molecular evolutionary genetics analysis software. Bioinformatics 17:1244–1245

Li YH, Beisson F, Pollard M, Ohlrogge J (2006) Oil content of Arabidopsis seeds: the influence of seed anatomy, light and plant-to-plant variation. Phytochemistry 67:904–915

Lysak MA, Koch MA, Pecinka A, Schubert I (2005) Chromosome triplication found across the tribe Brassiceae. Genome Res 15:516–525

Mu JY, Tan HL, Zheng Q, Fu FY, Liang Y, Zhang JA, Yang XH, Wang T, Chong K, Wang XJ, Zuo JR (2008) LEAFY COTYLEDON1 is a key regulator of fatty acid biosynthesis in Arabidopsis. Plant Physiol 148:1042–1054

Nadeau JA, Zhang XS, Li J, O'Neill SD (1996) Ovule development: identification of stage-specific and tissue-specific cDNAs. Plant Cell 8:213–239

Nakamura M, Katsumata H, Abe M, Yabe N, Komeda Y, Yamamoto KT, Takahashi T (2006) Characterization of the class IV homeodomain-leucine zipper gene family in *Arabidopsis*. Plant Physiol 141:1363–1375

Ogas J, Kaufmann S, Henderson J, Somerville C (1999) PICKLE is a CHD3 chromatin-remodeling factor that regulates the transition from embryonic to vegetative development in *Arabidopsis*. Proc Natl Acad Sci USA 96:13839–13844

Ohashi Y, Oka A, Ruberti I, Morelli G, Aoyama T (2002) Entopically additive expression of *GLABRA2* alters the frequency and spacing of trichome initiation. Plant J 29:359–369

Ostergaard L, King GJ (2008) Standardized gene nomenclature for the *Brassica* genus. Plant Methods 4:10

Parkin IAP, Sharpe AG, Lydiate DJ (2003) Patterns of genome duplication within the *Brassica napus* genome. Genome 46:291–303

Rahman M, Sun ZD, McVetty PBE, Li GY (2008) High throughput genome-specific molecular markers for erucic acid genes in *Brassica napus* (L.) for marker-assisted selection in plant breeding. Theor Appl Genet 117:895–904

Roesler K, Shintani D, Savage L, Boddupalli S, Ohlrogge J (1997) Targeting of the Arabidopsis homomeric acetyl-coenzyme A carboxylase to plastids of rapeseeds. Plant Physiol 113:75–81

Shen B, Sinkevicius KW, Selinger DA, Tarczynski MC (2006) The homeobox gene *GLABRA2* affects seed oil content in *Arabidopsis*. Plant Mol Biol 60:377–387

Snowdon RJ, Friedt W (2004) Molecular markers in *Brassica* oilseed breeding: current status and future possibilities. Plant Breed 123:1–8

Snowdon RJ, Friedrich T, Friedt W, Köhler W (2002) Identifying the chromosomes of the A- and C-genome diploid *Brassica* species *B. rapa* (syn. campestris) and *B. oleracea* in their amphidiploid *B. napus*. Theor Appl Genet 104:533–538

Sonnhammer EL, Koonin EV (2002) Orthology, paralogy and proposed classification for paralog subtypes. Trends Genet 18:619–620

Szymanski DB, Lloyd AM, Marks MD (2000) Progress in the molecular genetic analysis of trichome initiation and morphogenesis in *Arabidopsis*. Trends Plant Sci 5:214–219

Tiedemann J, Rutten T, Mönke G, Vorwieger A, Rolletschek H, Meissner D, Milkowski C, Petereck S, Mock HP, Zank T, Bäumlein (2008) Dissection of a complex seed phenotype: novel insights of *FUSCA3* regulated developmental processes. Dev Biol 317:1–12

Vain P, Afolabi AS, Worland B, Snaps JW (2003) Transgene behaviour in populations of rice plants transformed using a new dual binary vector system: pGreen/pSoup. Theor Appl Genet 107:210–217

Wang N, Wang Y, Tian F, King GJ, Zhang C, Long Y, Shi L, Meng J (2008) A functional genomics resource for *Brassica napus*: development of an EMS mutagenized population and discovery of *FAE1* point mutations by TILLING. New Phytol 180:751–765

Wang Z, Mao H, Dong CH, Ji RQ, Cai L, Fu H, Liu SY (2009) Overexpression of *Brassica napus MPK4* enhances resistance to *Sclerotinia sclerotiorum* in oilseed rape. MPMI 22:235–244

Wei F, Gao GZ, Wang XF, Dong XY, Li PP, Hua W, Wang X, Wu XM, Chen H (2008) Quantitative determination of oil content in small quantity of oilseed rape by ultrasound-assisted extraction combined with gas chromatography. Ultrason Sonochem 15:938–942

Wei B, Jing RL, Wang CS, Chen JB, Mao XG, Chang XP, Jia JZ (2009) *Dreb1* genes in wheat (*Triticum aestivum* L.): development of functional markers and gene mapping based on SNPs. Mol Breed 23:13–22

Weselake RJ, Shah S, Taylor DC, Zou JT, Laroche A, Moloney MM, Rakow G, Raney JP, Harwood JL (2007) Transformation of *Brassica napus* with diacylglycerol acyltransferase1 results in increased seed oil content. In: Benning C, Ohlrogge J (eds) Current advances in the biochemistry and cell biology of plant lipids. Aardvark Global, Salt Lake City, pp 232–234

Weselake RJ, Taylor DC, Rahaman MH, Shah S, Laroche A, McVetty PB, Harwood JL (2009) Increasing the flow of carbon into seed oil. Biotechnol Adv 27:866–878

Wu YH, Xiao L, Wu G, Lu CM (2007) Cloning of fatty acid elongase1 gene and molecular identification of A and C genome in *Brassica* species. Sci China Ser C Life Sci 50:343–349

Yang JY, Chung MC, Tu CY, Leu WM (2002) *OSTF1*: a HD-GL2 family homeobox gene is developmentally regulated during early embryogenesis in rice. Plant Cell Physiol 43:628–638