
Supplementary information

Genomic and genetic insights into Mendel's pea genes

In the format provided by the
authors and unedited

Supplementary notes and figures for:

Genomic and genetic insights into Mendel's pea genes

Cong Feng^{1,*}, Baizhi Chen^{1,*}, Julie Hofer^{2,*}, Yan Shi^{1,*}, Mei Jiang^{1,*}, Bo Song¹, Hong Cheng¹, Lu Lu¹, Luyao Wang¹, Alex Howard², Abdel Bendahmane³, Anissa Fouchal³, Carol Moreau^{2,4}, Chie Sawada², Christine LeSignor⁵, Cuijun Zhang¹, Eleni Vikeli², Georgios Tsanakas², Hang Zhao¹, Jitender Cheema^{2,6}, J. Elaine Barclay², Junliang Hou¹, Liz Sayers², Luzie Wingen², Marielle Vigouroux², Martin Vickers², Mike Ambrose², Marion Dalmais³, Paola Higuera-Poveda², Pengfeng Li¹, Quan Yuan¹, Rebecca Spanner^{2,7}, Richard Horler², Roland Wouters², Smitha Chundakkad², Tian Wu¹, Xiaoxiao Zhao¹, Xiuli Li¹, Yuchen Sun¹, Zejian Huang¹, Zhen Wu¹, Xing Wang Deng⁸, Burkhard Steuernagel², Claire Domoney², Noel Ellis^{2,#}, Noam Chayut^{2,#}, Shifeng Cheng^{1,#,\$}

¹ Shenzhen Branch, Guangdong Laboratory of Lingnan Modern Agriculture, Key Laboratory of Synthetic Biology, Ministry of Agriculture and Rural Affairs, Agricultural Genomics Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen, China

² John Innes Centre, Norwich Research Park, Norwich NR4 7UH, UK

³ INRAE UMR 1403, Institute of Plant Sciences Paris-Saclay, Bâtiment 630, Rue Noetzlin, 91190 Gif-sur-Yvette, France

⁴ Paleogenomics Laboratory, INRAE Clermont-Auvergne-Rhône-Alpes, CS 60032, 63001 Clermont-Ferrand, France

⁵ INRAE UMR1347-Agroecologie pôle GEAPSI, 17 rue SULLY BP 86510, 21065 DIJON Cedex, France

⁶ EMBL-EBI, European Molecular Biology Laboratory, Wellcome Genome Campus, Hinxton, Cambridgeshire, CB10 1SD, UK

⁷ Department of Plant Pathology, University of Minnesota, St. Paul, MN 55108, USA

⁸ State Key Laboratory of Wheat Improvement, School of Advanced Agricultural Sciences and School of Life Sciences, Peking-Tsinghua Center for Life Sciences, Peking University, Beijing 100871, China.

^{\$} Lead correspondence

^{*} These authors contributed equally to this article.

[#] To whom correspondence may be addressed. Email: chengshifeng@caas.cn; Noel.Ellis2@jic.ac.uk; Noam.Chayut@jic.ac.uk

Content

1. Supplementary notes	5
1.1. Overview.....	5
1.1.1. Pea, a model system	5
1.1.2. Mendel pea G2P project.....	7
1.2. The four cloned genes	8
1.2.1. Seed shape: round vs wrinkled, <i>R</i> vs <i>r</i>	8
Summary and discussion.....	10
1.2.2. Cotyledon colour: yellow vs green, <i>I</i> vs <i>i</i>	10
Summary and discussion.....	12
1.2.3. Flower colour: presence vs absence of anthocyanin pigmentation, <i>A</i> vs <i>a</i>	14
Summary and discussion.....	16
1.2.4. Internode length: tall vs dwarf, <i>Le</i> vs <i>le</i>	16
1.3. The remaining uncharacterized genes	17
1.3.1. Pod colour: green vs yellow, <i>Gp</i> vs <i>gp</i>	17
Genetic mapping and GWAS analysis of <i>Gp</i>	18
Summary and discussion.....	21
1.3.2. Pod shape: inflated vs constricted, <i>P</i> and <i>V</i>	23
<i>P</i>	23
<i>V</i>	24
Summary and discussion.....	26
1.3.3. Flower position: axial vs terminal (fasciation), <i>Fa</i> and <i>Mfa</i>	28
<i>Fa</i>	28
<i>Mfa</i>	29
<i>F</i> ₃ segregation test of the <i>F</i> ₂ discrepancies to further explain the <i>Fa</i> - <i>Mfa</i> model	31
Summary for the above analyses	34
Summary and discussion.....	36
1.4. Quantitative traits that Mendel commented on	40

1.4.1. Axil ring pigmentation: <i>D</i> vs. <i>d</i>	40
Summary and discussion.....	42
1.4.2. Organ size: pod width and seed weight, <i>PsOs1</i>	42
1.5. Agronomic traits to global breeding objectives	44
2. Supplementary Figures	46
Supplementary Fig. 1.....	46
Supplementary Fig. 2.....	47
Supplementary Fig. 3.....	49
Supplementary Fig. 4.....	51
Supplementary Fig. 5.....	52
Supplementary Fig. 6.....	54
Supplementary Fig. 7.....	56
Supplementary Fig. 8.....	57
Supplementary Fig. 9.....	59
Supplementary Fig. 10.....	61
Supplementary Fig. 11.....	63
Supplementary Fig. 12.....	65
Supplementary Fig. 13.....	66
Supplementary Fig. 14.....	67
Supplementary Fig. 15.....	69
Supplementary Fig. 16.....	71
Supplementary Fig. 17.....	73
Supplementary Fig. 18.....	74
Supplementary Fig. 19.....	75
Supplementary Fig. 20.....	76
Supplementary Fig. 21.....	78
Supplementary Fig. 22.....	80
Supplementary Fig. 23.....	82
Supplementary Fig. 24.....	83
Supplementary Fig. 25.....	85
Supplementary Fig. 26.....	86

Supplementary Fig. 27.....	88
Supplementary Fig. 28.....	89
Supplementary Fig. 29.....	90
Supplementary Fig. 30.....	91
Supplementary Fig. 31.....	93
Supplementary Fig. 32.....	94
Supplementary Fig. 33.....	95
Supplementary Fig. 34.....	96
Supplementary Fig. 35.....	97
Supplementary Fig. 36.....	98
Supplementary Fig. 37.....	99
Supplementary Fig. 38.....	100
Supplementary Fig. 39.....	101
Supplementary Fig. 40.....	102
Supplementary Fig. 41.....	103
Supplementary Fig. 42.....	104
Supplementary Fig. 43.....	105
Supplementary Fig. 44.....	106
3. References:	107

1. Supplementary notes

1.1. Overview

1.1.1. Pea, a model system

Pea (*Pisum sativum*), one of the oldest cultivated legumes, holds a unique position in genetics, ecology and agriculture¹. As a crop, peas contribute a rich source of plant protein, minerals, and vitamins to human diets worldwide². Their ability to fix atmospheric nitrogen through symbiosis with rhizobia helps maintain soil fertility, reducing the need for chemical fertilizers and promoting sustainable farming^{3,4}. Peas also have a distinct place in ornamental horticulture, prized for their fragrant blooms and decorative use in gardens. Archaeological evidence suggests that peas were domesticated at least 10,000 years ago in the Fertile Crescent, evolving from wild relatives with small seeds and pods that shatter easily⁵. Over time, humans selected for traits such as larger seeds, reduced pod shattering, and improved flavour. Understanding the genetic basis of these domestication traits remains a central research topic.

Peas are just as remarkable in science as they are in bloom. Their unique role in the history of genetics is legendary, owing primarily to Gregor Mendel's foundational experiments in the mid-19th century⁶. Mendel's crossing experiments of peas and subsequent analysis of inherited traits laid the groundwork for modern genetics⁷. In the early 20th century, many researchers followed in Mendel's footsteps, exploring pea's inheritance patterns for diverse traits, from seed colour to plant architecture⁸⁻¹². During this period, *Pisum* became one of the most important model systems for understanding classical genetics and plant breeding. However, around the 1990s, the focus shifted toward other model systems such as *Arabidopsis thaliana*, *Nicotiana benthamiana*, *Medicago truncatula*, *Lotus*

japonicus, and *Glycine max* (soybean)¹³⁻¹⁶. Multiple factors contributed to this shift. First, the pea genome is comparatively large (approximately 4.3 Gb) with a large proportion of complex repetitive elements that complicate sequencing and assembly. Second, pea has been challenging to stable genetic transformation and gene editing by CRISPR/Cas methods, slowing the pace of functional genomics. Third, *Arabidopsis* offered a much smaller genome (~135 Mb), faster life cycle, and more tractable transformation protocols, making it an attractive model system to the community. Recent advances in genomics, including more sophisticated sequencing technologies and improved bioinformatics tools, have overcome many of these obstacles¹⁷⁻¹⁹.

Pea has considerable diversity, both genetically and phenotypically²⁰. The nucleotide diversity in *Pisum*²¹, is about tenfold greater than that in the human population^{22,23}, reflecting bidirectional introgression between the cultigen and wild genotypes. Such diversity is invaluable not only for basic genetic studies, such as unravelling Mendel's seven 'riddles' in molecular genetics, but also for breeding programs aimed at improving yield, stress tolerance, and nutritional quality. *Pisum sativum* (meaning 'cultivated pea') is a subset of *Pisum* as a whole; wild peas designated *P. fulvum* are noticeably distinct^{19,24} and carry a translocation with respect to the rest of *Pisum*²¹, which creates a fertility barrier. Similarly, the independently domesticated *P. abyssinicum*^{24,25} differs in karyotype with respect to the rest of *Pisum*¹⁷, again presenting a fertility barrier. Thus, *Pisum* comprises four major divisions; the cultivated forms *P. sativum* and *P. abyssinicum* and the wild forms *P. fulvum* and *P. elatius*. Morphological diversity within *Pisum* has been documented since at least the 16th century, with Gerard²⁵ (p1045) illustrating four forms: *P. majus*, *P. minus*, *P. umbellatum* (fasciated) and *P. excorticatum* (parchmentless) and discussing several others, such as those with seeds "which being drie are cornered". If by this latter description, Gerard was referring to wrinkled peas, then three of the variant forms that Mendel studied¹⁷ viz.

peas with stem fasciation, parchmentless pods, and wrinkled seeds, were recorded nearly 300 years ago. White flowered forms, as previously noted, were described by about 1300²⁵.

1.1.2. Mendel pea G2P project

We launched the Mendel Pea G2P (Genomics to Phenomics) project (<https://mendelpea.com/>) as an international collaboration between China and the UK, aiming to explore *Pisum* population genomics and phenomics for trait and gene discovery. This study was based on a core set of 697 representative accessions selected from the JI *Pisum* Germplasm Collection (**Supplementary Table 1**). To connect genetic diversity with the phenotypic diversity, we combined whole-genome re-sequencing (20X) (**Supplementary Table 2**) with extensive phenotyping efforts conducted across the UK and China (**Extended Data Figure 8** and **Supplementary Table 37**), linking genomic and trait variation.

A high-quality genomic variation map, including SNPs, Indels, CNVs and SVs (**Fig. 1f** and **Supplementary Fig. 1** and **Supplementary Table 3-11**), was constructed, along with a linkage disequilibrium-based haplotype map (**Supplementary Table 12**). Concurrently, we performed a phenotyping survey of 79 agronomic, biochemical and morphological traits - spanning seeds, leaves, pods, flowers, to plant architecture - coupled with large-scale haplotype-phenotype association studies. Our focus included 9 traits in a Mendel-centric manner, emphasizing the seven contrasting pairs of traits as originally described by Gregor Mendel in his 1866 seminal paper. These 9 traits include Mendel's four characterized genes (*R*, *I*, *A* and *Le*) for which we prioritized the identification of novel alleles and the three remaining uncharacterized traits involving four genetic loci (*Gp* for pod colour, *P* and *V* for pod shape, *Fa* for flower position) (**Fig. 2** and **Extended Data Fig. 1** and **Supplementary Table 46**); for the latter group, we sought to elucidate gene identities and causal alleles. For

fasciation, we identified an additional genetic locus, *Mfa*. Additionally, we investigated two traits, relevant to Mendel's comments, identifying allelic variants in a *MYB* gene cluster corresponding to the *D* locus (axil ring pigmentation) and defining allelic variation regulating organ size, as reflected in both pod width and grain weight. We continued to explore the deletion haplotypes associated with the so-called 'semi-leafless' phenotype, a plant architectural trait essential to modern pea breeding, and grain quality traits (grain starch and protein content), which are of contemporary importance for crop improvement.

The key concept here was to identify candidate genes and novel allelic variation using whole genome trait association, locating them using existing and newly developed mapping populations, and using RNA-seq experiments, allelism tests, and mutagenesis coupled with genome sequence for functional genomic validation of the identified candidates. The results of these studies provide insights for pea fundamental research, education in biology and genetics, and applied breeding practices.

1.2. The four cloned genes

1.2.1. Seed shape: round vs wrinkled, *R* vs *r*

The first of Mendel's genes to be described at the molecular level was *R* corresponding to 'the difference in the shape of the ripe seed'. We associated the haplotype map with the phenotypic changes in seed shape across the entire diversity panel (**Supplementary Fig. 2** and **Supplementary Table 13**), showing a single strong but surprisingly broad signal, at the expected genomic position of *R*, encoding Starch Branching Enzyme I (*PsSBEI*)²⁶. This signal was not identified by SNP allelic variation (**Supplementary Fig. 2b**). The SNP variants found in the *r* allele are shared with *R* alleles, which lack the *Ips-r* insertion of *PsSBEI* Hap5. The breadth of this peak is not diminished if the presence /absence of *Ips-r* is subjected to GWAS analysis, suggesting that this is a structural feature of the allelic diversity

around this gene differentiating *R* vs *r* genotypes. Genetic differentiation between round and wrinkled types in breeding programmes could be the underlying reason for the broad GWAS peak: round types are predominantly field peas, grown for their dry seed, while wrinkled types are a class of peas grown for harvesting before maturity as fresh peas or for the freezing market²⁷. This single GWAS peak further indicates that there is little or no genetic heterogeneity associated with this phenotype within the set of lines we have examined. The only *r* allele corresponds to the presence of the *Ips-r* element, a non-autonomous type II transposon inserted in the last exon (#22) of the Starch Branching Enzyme I (*PsSBEI*) gene, as described by Bhattacharyya et al. (1990)²⁶. A previous study of 2792 *Pisum* accessions had also determined that all *r* mutations corresponded to the same insertion allele of *SBEI* (Rayner et al. 2017)²⁸.

Here we present the full-length sequence of the *Ips-r* element. It is a 1021bp non-autonomous *Ac*-like transposable element, within exon 22 of the *PsSBEI* coding sequence, the sequence of this allele predicts a truncated protein (from 922 aa to 890 aa) due to a premature stop codon^{26,28} (**Fig. 2c**), although multiple *r* transcripts have been detected²⁹. The sequence of the *Ips-r* element is very AT rich and bounded by the 8bp target site duplication, as previously described. The 1021 bp sequence is larger than the original estimate (0.8kb)²⁶, presumably reflecting anomalous migration in agarose gels of DNA fragments carrying this sequence, but similar to that determined by Rayner et al. (1024 bp)²⁸, where the repetitive nature of the insertion element was noted to hinder amplification and determination of its sequence.

Note that there are five known genes which, when mutant, confer a wrinkled seed phenotype (*r*, *rb*, *rug3*, *rug4* and *rug5*)³⁰ of which only the *r* and *rb* alleles are known outside mutagenesis projects³¹; the *rb* mutation arose in the 1930s³² and has limited representation in pea germplasm (12 *rb* lines identified among the 2792 accessions²⁸). The variety Kebby

(J12110) is wrinkled, but the genetic basis of this is complex²⁸. The phenotypic record of seed shape for each accession is summarized in **Supplementary Table 13** and can be viewed in **Supplementary Fig. 4**.

Summary and discussion

Our initial GWAS analysis of the entire pea diversity panel revealed a significant association between seed protein content and seed shape, with both traits linked to the *R* versus *r* allele, influencing seed protein content (% of dry weight). Pea protein has gained significant attention for its versatile benefits in food, nutrition and sustainability³³. The influence of the *R* gene on the ratio of the two major pea seed storage protein classes, as well as on overall protein content, has been well-documented³⁴. Further GWAS analyses conducted separately on round-seeded and wrinkled-seeded types uncovered several new loci (**Supplementary Fig. 3**), indicating the presence of other genetic factors that regulate seed protein content. As two complex traits influenced by many genetic elements and environmental conditions, pea seed starch and protein content remain challenging to predict and control with precision, reflecting the intricate interplay (see also the next section for cotyledon colour) of genetic and phenotypic diversity.

1.2.2. Cotyledon colour: yellow vs green, *I* vs *i*

We identified a single strong signal at the expected genomic position of *I*, a stay green gene encoding Mg-dechelatase^{21,35,36} (*PsSGR*) which catalyses the first step in chlorophyll degradation and underlies the genetic difference between green vs yellow cotyledons (**Supplementary Fig. 4, Supplementary Table 14**). This strong genetic effect with a significant signal can explain most of the heritable phenotypic difference – the green/yellow seed colour variation - in pea (**Supplementary Fig. 5**). The mutants fail to degrade chlorophyll and so remain green. In the *Ii* heterozygote, the enzymatic function is present

thus explaining the recessiveness of the green allele. We did not include JI2775 in the set of accessions we studied, and did not find the i^{JI2775} allele in any of the 697 accessions in our panel, so we conclude that the 6bp insertion event of the i^{JI2775} allele³⁵, is rare and therefore cannot be considered as the major cause of green cotyledons. Neither did we find the group 4 *sgr* variant described previously³⁵.

We found two classes of *i* alleles (**Fig. 2c** and **Supplementary Fig. 5**), which explain most of the green cotyledon mutants in this diversity panel. The more common mutant allele (designated as *i-1*) is an insertion of a 5,696 bp transposable element, a *Ty1*-copia retrotransposon (TAR) element, together with its 5 bp target site duplication; presumably this corresponds to the ca. 6.5 kb insertion of the group 3 allele as described by Sato et al.³⁵, but these authors did not present a sequence or characterize the frequency of this allele at the population level. The coding region of the *Ty1*-copia element is intact and the LTRs are identical, indicating an evolutionarily recent insertion^{37,38}.

The second allele we discovered is a novel 408 bp deletion in the promoter, 143 bp upstream of the ATG of the Mg-dechelataase gene (**Supplementary Fig. 6b**), which we designate as the '*i-2*' allele, explaining 15 accessions with green cotyledons. This deletion disrupts the region of the 5' UTR and therefore may interfere with the expression of this gene or the efficient production of an effective protein. These novel *PsSGR* (*I/i*) alleles discovered in this study are widely distributed in the *Pisum* diversity panel, and have strong explanatory power for most of the accessions with green cotyledons. Several minor peaks with a $-\log_{10}(p)$ value ~ 10 , distinct from the *I* locus, can be seen in the Manhattan plot (**Fig. 2b** and **Supplementary Fig. 5a**). Modifiers of cotyledon colour are well known and ten genetic loci that contribute to this effect have been identified³⁹, but their location with respect to these additional GWAS peaks could not be determined. The phenotypic record of green

cotyledons for each accession is summarized in **Supplementary Table 14** and can be viewed in **Supplementary Fig. 4**.

Summary and discussion

In our GWAS analysis, we detected two *i* alleles (*i-1* and *i-2*) associated with the presence of green rather than yellow cotyledons. However, not all *ii* genotypes were scored as having green cotyledons (**Supplementary Fig. 7**) and this could be taken as evidence that *i*, encoding Mg dechelatase, is not the only factor that is responsible for the green cotyledon character. The variability of cotyledon colour, in ostensibly green cotyledon lines, has been an issue of discussion for a long time. The matter was first discussed by Mendel in his 1866 paper where he wrote that “some care is requisite, since in some of the seeds of many plants the green colour of the albumen is less developed, and at first may be easily overlooked”. The issue was raised again by Weldon⁴⁰ who took this as evidence against Mendelian genetics, but Bateson 1902 argued against Weldon, noting that some varieties have cotyledons which are “*either yellow or green, with some piebalds*” p121⁴¹ (piebald: here that cotyledons are partly yellow *and* partly green), further commenting: “The variety with greatest *irregularity* (apart from regular dimorphism, i.e. being yellow *or* green) in cotyledon colour I have seen is a sample of ‘*mange-tout à rames, à grain vert,*’ but it was good deal injured by weevils (*Bruchus*), which always cause irregularity or change of colour.”

In the F₂ of the cross JI2822 x JI0816 (**Supplementary Tables 18-20**), cotyledon colour is segregating. JI0816 has yellow cotyledons and JI2822 is homozygous for the *i-1* allele (with the TAR element insertion) and has green cotyledons. One of the genetic markers, AX-183564806, which we used while creating the genetic map of this F₂ population (see **Supplementary Table 18**) is located within the Mg dechelatase gene (chr2, LGI), so we were able to identify F₂ individuals which were *i-1/i-1* homozygotes and examine their F₃

seeds. Within each packet of F₃ seeds, we removed some of the testa from 10 seeds in order to see the colour of the cotyledon ((**Supplementary Fig. 7a**). Nine F₃ seed packets had fewer than 10 seeds and, for these, all seeds were examined and counted. In total 833 F₃ seeds were examined, from 88 F₂ plants. Of these 96 F₃ seeds (from 33 different F₂ individuals) were scored as yellow (68) or pale green (28). We noticed that all but one of the seeds with yellow or pale green cotyledons had a strongly brown coloured testa, most likely coloured by oxidised polyphenols⁴². One insect damaged seed was partly yellow and partly green (see Bateson's comment above); this seed was not counted.

We next examined the genotypes of these F₂ individuals to see whether any genomic location could be assigned which may condition this yellowing of *i-1/i-1* homozygotes (**Supplementary Fig. 7b**). When we looked for an excess in the frequency of JI0816 homozygotes among the F₂ individuals with at least one seed with yellow or pale cotyledons, an excess was seen in LGIV (chr4) (arrowed). The significance of this association was not very strong ($\chi^2 \approx 6.6$, $p \approx 0.01$), given the number of χ^2 tests, even if they are not all independent. A region of LGIV (chr4) similarly seemed to be associated with this yellowing of the cotyledons ($\chi^2 \approx 5$, $p \approx 0.03$), but again not very significantly. These are different F₂ individuals that are homozygous for the JI0816 *i-1* allele than is the case for the LGII region, so the effects would be combined in the F₂:F₃ families.

However, we should note that the identification of *i-1/i-1* F₂ homozygotes with some yellow offspring is not accurate because only 10 (at most) F₃ seeds from F₂ homozygote individuals were examined. These observations are consistent with some factor, for example, prolonging the time to maturity, increasing the risk of damage to seed causing cotyledons which would have been green, to turn yellow independently of the action of Mg dechelatase.

Meanwhile, there is a wealth of literature on the involvement of lipoxygenase (Lox) enzymes (in conjunction with hydroperoxide lyases) in the loss of chlorophyll from plant

tissues. In pea, *Lox-3*, located on chr4 (LGIV) but distant from the position marked on **Supplementary Fig. 7b**, has been implicated strongly⁴³, though it is clear that there are many different Lox isozymes, including those localised to plastids. Some of these have been mapped genetically^{44,45}. The genetics of resistance to loss of seed colour in pea identified lots of QTL and concluded that genetic and environmental factors were involved; environmental effects are also supported by other work^{21,45}. Some effort is needed to equate the QTL identified with candidate genes but one maps genetically to the position of *Lox-10*³⁹.

The pathway of chlorophyll loss from seeds is known to be regulated by ABA through the B3 domain TF, ABI3, which has a critical role in seed maturation, conferring desiccation tolerance⁴⁶. In addition, an EMS-induced green cotyledon mutant in pea, which was not allelic to *i*, named *gc* (green cotyledon), has been obtained and the mutation mapped to chr6 LGII (not shown). The mutant was clearly impacted in desiccation tolerance and we concluded that the locus was distinct from that reported by LeSignor et al.⁴⁷, for mutants also showing green cotyledons and a similar phenotype.

In summary, it seems likely that environmental factors lead to oxidative damage which can convert otherwise green cotyledons to a pale green or yellow colour. Genetic factors could prolong this exposure or increase the sensitivity of the seed to this damage. So, there are indeed *ii* mutant seeds where a cotyledon may be green or yellow and occasionally both colours can be seen in an individual seed (possibly this is what Weldon observed), There is clearly more complexity to the green cotyledon phenotype than a simple dechelation step.

1.2.3. Flower colour: presence vs absence of anthocyanin pigmentation, *A* vs *a*

The difference in the colouration of the seed coat is the third character for which Mendel studied the pattern of inheritance; he noted that this was pleiotropic. In one of his three factor crosses, he referred to this as ‘flower colour’, which is the more usual description today. Our haplotype-phenotype association study of pigmented vs white flowers (**Supplementary Table 15**) revealed a single strong signal spanning a genomic region consistent with the location of *A*, which encodes a *bHLH* transcription factor⁴⁸ (**Fig. 2**) required for the expression of chalcone synthase in epidermal tissues⁴⁹, thereby enabling anthocyanin pigmentation. Consequently, *a* mutants lack anthocyanin pigmentation throughout the plant. In the initial GWAS analysis based on the Caméor v1a genome reference (**Supplementary Fig. 8b**), we were surprised to find two GWAS peaks for white vs pigmented flowers. This turned out to correspond to a mis-assembly as revealed by mapping the Caméor v1a assembly to the JI0281 x Caméor recombinant inbred population⁵⁰. For this reason, when the ZW6 assembly became available, we re-mapped our SNP matrix with respect to ZW6 (**Supplementary Fig. 8a**) and used this assembly in the Manhattan plots for other traits.

We discovered several novel haplotypes within the structural *bHLH* gene (**Supplementary Fig. 8d-g**). The wild types (*A*) with pigmented flowers were assigned to haplotype (Hap) 1 based on the distribution of functional variants, but Hap1 is remarkably diverse at other positions. The two most common *a* alleles correspond to Hap5, carrying the splice donor site variant (G to A), originally identified in Caméor, and Hap2, with an additional ‘A’ nucleotide insertion in exon 6 creating a premature stop codon, as originally identified in JI1987⁴⁸. Two new variants (Hap3, with four accessions and Hap4, with one accession) are deletions of part (the first two exons), or almost the entirety (the first six exons) of the gene.

For the accessions that have the wild-type splice donor site (of Hap1), 359 are clearly pigmented with anthocyanins; although 5 are very pale they show axil ring

pigmentation, corresponding to the *am* mutation. Of the Hap1 accessions, 9 appear white but the genetic basis for this is not clear; none of the accessions in this panel was known to carry the *a2* mutation. A further 35 accessions carry the α^{JI1987} allele (Hap2). Four samples including JI0735 are white-flowered while having the wild-type splice donor site, but harbour a deletion of exons 1 and 2. The splice donor site allele could not be called in JI1775 (Hap4) which is white flowered and has a deletion of exons 1 to 6. The gene *A* and its *a* alleles are thus entirely consistent with this *bHLH* transcription factor being required for anthocyanin pigmentation. JI0233 has the splice donor site mutation of most *a* alleles, but is pigmented. This allele is a rare intragenic suppressor mutant that restores the functionality of this transcription factor.

Summary and discussion

JI0233 is an accession with coloured flowers that carries the splice donor site mutation which should render the gene dysfunctional. This JI0233 allele (in Hap5), restoring the exon 7 reading frame, has an insertion of an additional T nucleotide in what would be the sixth intron of the wild type allele. The insertion lies between the wild-type splice donor site and the splice donor site used in the Caméor allele, adding nine nucleotides to the transcript, one more than in the $\alpha^{Caméor}$ allele (this has been validated by cDNA sequencing and read mapping of JI0233, **Supplementary Fig. 9**). This single nucleotide insertion is therefore an intragenic suppressor mutation. The phenotypic record of flower colours for each accession is summarized in **Supplementary Table 15**.

1.2.4. Internode length: tall vs dwarf, *Le* vs *le*

Variation in internode length in our analysis corresponds to Mendel's plant height character (**Supplementary Table 16** and **Supplementary Fig. 10**). Several factors affect plant height in pea, of which variation in internode length under the control of gibberellic acid is a major

contributor. Mendel described the segregation of plant height, but did not specifically refer to internode length. The GA-related genes, either involved in GA signalling or biosynthesis, are well-known to be responsible for genetic changes in plant height, such as in the semi-dwarfing wheat (*Della*)⁵¹ and rice (*sd1*)⁵², equivalent to *La* and *Cry* in pea⁵³. Our GWAS analysis of internode length (stem length, or plant height) identified a significant peak (chr5: 620824850-652929960) at the end of chromosome 5, at the location of *Le*, encoding GA 3-oxidase1 (*Psat05G0825300*, also called GA 3 β -hydroxylase) (**Fig. 2**); this is therefore the major determinant of the difference in stem (internode) length in pea. This location does not extend to *Lh*⁵⁴ (*Psat05G0840800*, chr5:650785676-650788204), another gene conditioning plant height, closely linked to *Le*. That the GWAS approach finds this single peak suggests that variation at other known loci involved in regulation of the type and abundance of plant hormones affecting plant height, or internode length, do not contribute significantly to natural phenotypic variation in this trait. We found only one *le* allele (a G to A substitution in the coding region), consistent with previous reports^{55,56}. We identified one tall plant (JI0019) among the *le* genotypes, this corresponds to a ‘cryptodwarf’⁵⁷. There are several accessions carrying the *Le* alleles but showing short internodes. The haplotype clusters and the haplotype-phenotype analysis are summarized in **Supplementary Table 16**.

1.3. The remaining uncharacterized genes

1.3.1. Pod colour: green vs yellow, *Gp* vs *gp*

The yellow colour of *gpgp* pea pods, as Mendel noted, is one aspect of a wider syndrome affecting several parts of the plant, young leaf tips, the flowering inflorescence and the sepals as well as the pods, which is illustrated by the comparison between *GpGp* and *gpgp* near-isogenic lines in **Fig. 3a** and in **Supplementary Fig. 11**. The disturbance of the thylakoid membranes in leaflets of *gpgp* suggested the overall photosynthetic activity of these plants,

and not just pod photosynthesis, might be impaired⁵⁸ (**Fig. 3b,c** and **Supplementary Fig. 12**). Pod dimensions and seed weight were all affected in *gp* plants, suggesting that plant productivity was lower in *gpgp* compared with the *GpGp* plants (**Supplementary Fig. 13**). Not all peas with yellow pods are *gpgp* homozygotes; these others can have a generally yellow or golden appearance such as chlorotic mutants (e.g. JI2764 *chlorotica*), the phytochrome deficient mutant *pcd1* (JI3172), or Goldkonig (JI0799), which has an overall yellowish appearance, a phenotype distinct from *gp*. Similar variant pea plants, such as cv. Doré (Vilmorin 1856, https://en.wikipedia.org/wiki/Louis_de_Vilmorin) were available at the time of Mendel's work but, as White⁵⁹ noted, crosses involving Goldkonig show that the genetic determinant of this trait is distinct from *gp* or *i*.

There are nineteen yellow-podded lines in total in the JI *Pisum* germplasm collection: JI0013, JI0015, JI0046, JI0056, JI0058, JI0073, JI0128, JI0132, JI0136, JI0816, JI1341, JI1343, JI2082, JI2273, JI2659, JI2716 and JI3541, and 12 representatives were selected for genome sequencing in this study; crosses between these 19 lines show that all are allelic to *gp*. Pellew and Sverdrup (1923)⁶⁰ reported a spontaneous origin of this phenotype in the offspring of cv. Duke of Albany (JI0313) and JI0013 is recorded as 'Pellew's *gp*', suggesting that it might have been an independent *gp* allele, which if true would have been very useful.

Genetic mapping and GWAS analysis of Gp

As described in Online Methods and summarized in **Supplementary Fig. 14**, the markers PsCam005046, PsCam056084, AX-183865165, AX-183571028 and AX-183571050 do not co-segregate with *Gp* in the cross JI0015 x JI0399. However, in the JI2822 x JI0816 F₂ population (**Supplementary Tables 18-20**), the marker AX-183879077 (ZW6 Chr3:324762848) does co-segregate with *Gp*. These data are consistent with *ChlG* co-segregating with *Gp* and primers designed to test *ChlG* confirmed that this is indeed the case

in the JI0015 x JI0399 RIL population (**Supplementary Tables 21-22**). JI0399 has green pods (genotype *Gp/Gp*) and JI0015 has yellow pods (genotype *gp/gp*). Thus, *Gp* is located within the 4.5Mb interval defined by AX-183571050 and AX-183571028 (ZW6 Chr3:321020350-325580858). Two genes, corresponding to 3' exonucleases, have previously been shown to be associated with the *Gp* vs *gp* allelic difference²². These genes were located to a scaffold in the Caméor v1a assembly, which is part of the chromosome 3 assembly of ZW6 at a location close to our GWAS peak (**Supplementary Fig. 14**). Genomic analysis of this region in genome assemblies of JI0015 and JI2366 revealed a ca. 100kb deletion with the coordinates ZW6 Chr3:324692139-324798134 (**Supplementary Fig. 15**).

From our association genomics analysis, we can see that in the SNP-based GWAS, there is an unexpected shoulder that extends to most of the short arm of this chromosome, which requires further exploration (**Fig. 2**, **Supplementary Fig. 14**, and **Supplementary Table 17**). The SNP-based GWAS gives a 10Mb broad genomic interval (Chr3, 315-325Mb) with high significance, whereas our haplotype-based GWAS gives a single strong and narrow interval with only 1Mb (324-325Mb). The 10Mb interval identified in the SNP-based GWAS analysis includes the 4.5Mb region identified in the genetic analysis. SV (structural variation)-based GWAS detected a ca. 100kb deletion within the minimum interval defined for *Gp* in the genetic analysis described above, as shown in **Supplementary Fig. 14-Supplementary Fig. 15**. This ca. 100kb deletion was found in all 19 *gp* accessions but in no *Gp* accessions (**Supplementary Fig. 15**). All genes within the deletion, *Psat03G0413800*, *Psat03G0413900*, *Psat03G0414000* plus a transcript within the LTRs of an *Ogre* element (at chr3:324696949-324696718 in ZW6), which has not been annotated in the ZW6 genome, together with the truncated *TIR-NBS-LRR* (ZW6 *Psat03G0414100*, Caméor *Psat3gl26360*) gene, are therefore candidates for *Gp*. The gene adjacent to the deletion is annotated as

Chlorophyll Synthase (*ChlG*) and a disruption to the transcription of this gene is expected to result in a yellow phenotype.

A VIGS experiment silencing *ChlG* did indeed result in the yellowing of leaves, consistent with its annotation (**Supplementary Fig. 17**); however, this phenotype is not consistent with the green leaflets of *gpgp* mutants. While the VIGS experiment is not consistent with the *gp* phenotype, it would be consistent with a reduction in *ChlG* activity in leaves as an explanation for the inefficient thylakoid membrane assembly shown in **Fig. 3b, c and Supplementary Fig. 12**, and also the yellowness of the pods. This suggested that a study of the transcription of *ChlG* in *Gp* and *gp* pods and leaves could be informative.

Aberrant transcripts were found in *gpgp* genotypes, including three fused transcripts from *TIR-NBS-LRR* gene (*Psat03G0414100*) to *ChlG* (*Psat03G0413700*) by transcriptional intergenic read-through (t1-t3), which were confirmed by RNA-seq read mapping and the junction reads (**Fig. 3d and Supplementary Tables 23-25**). One transcript (t4) with breakpoint read-through was confirmed by the RNA-seq read mapping. Two transcript variants (t5, t6) with read-through of intron 1 of *ChlG* were confirmed both by RNA-seq mapping, and by RT-PCR for *gp* RIL segregants, along with the expected wild-type *ChlG* transcript which is also evident in *Gp* RILs (**Supplementary Fig. 16-17**). It was noted that the predicted t4 transcript is not an intergenic readthrough because it does not read through from one gene into another, nor does it splice two genes together like the t1-t3 transcriptional fusions do. Instead, it reads across the deletion breakpoint into the intergenic (non-coding) region. So the broken exon 5 becomes elongated by a predicted 55 amino acids. A premature stop codon is evident within the *ChlG* intron 1 read-through. The presence of the intron 1-containing transcript was shown to cosegregate with *gp* (**Supplementary Fig. 16**), indicating that intron 1 read-through is not an effect of a separate locus acting in trans, and is probably a consequence of the overall rearrangement of transcription in this region. The

original *ChlG* transcript (T0) was also detected in the *gp* mutant line, but with much lower transcript abundance compared with its counterpart in the wild-type *Gp* line (**Fig. 3e** and **Supplementary Fig. 19**). The relative transcript abundance from each of the exons of both *TIR-NBS-LRR* and *ChlG* was assessed in leaves and pods of the *gp* accession JI2366 as shown in **Supplementary Fig. 19**. The aberrant transcript isoforms detected are indicated in **Fig. 3d** and **Supplementary Tables 26-27**. These aberrant transcripts (fused transcripts, breakpoint read-through and intron read-through) either break the intergenic and promoter region of its downstream *ChlG* gene, or disrupt the transit peptide of the *ChlG* transcript, making mutant *ChlG* isoforms which would not insert into the thylakoid membrane (**Supplementary Fig. 19**). However, the biological functions of these aberrant transcripts are unknown.

We tested whether *ChlG* is implicated in the *gp* phenotype. A TILLING mutant in *ChlG* was obtained from the pea TILLING resource at INRAe⁶¹. This mutant was W121* corresponding to a premature stop codon between the first and second predicted transmembrane helices of *ChlG*. The mutant was obtained as a heterozygote and no homozygous mutants were obtained upon selfing of this line. This is consistent with the W121* non-functional *ChlG* being an embryonic lethal. Note that there was no difference in the transmission rate of the mutant allele from the male or female gametophyte, indicating that *ChlG* is not required for gametophytic viability. A complementation/allelism test was performed by crossing the heterozygous TILLING mutant (as female) with JI0015 *gpgp* as male and the phenotype of the F₁ was examined (**Fig. 3h-i**). The failure of *ChlG*^{W121*} - *WT* to complement *gp* indicated that the TILLING mutant does not harbour a wild-type (Caméor) *Gp* allele; rather the TILLING allele *gp*^{lethal} is a lethal allele recessive to *gp*^{JI0015}. This establishes that *ChlG* is allelic to *Gp*.

Summary and discussion

These results suggest that a reduction in the abundance, and the presence of novel aberrant isoforms, of *ChlG* transcripts account for the yellow tissues in *gp* mutants.

The elevated transcript abundance in leaves vs pods correlates with the green colour of expanded leaflets. Several different aberrant transcripts were discovered from the *TIR-NBS LRR-ChlG* region, and they are of different types. One is an aberrant transcript that includes sequences from both genes, which has been validated by our PCR and sequence analysis. One is an aberrant transcript with breakpoint transcriptional readthrough from exon5 of *TIR-NBS LRR* into the intergenic region that would encode a truncated NBS-LRR protein of 753 amino acids. Another type of aberrant transcript includes intron read-through. This intron read-through aberrant transcript co-segregates with *gp*, demonstrating that it is determined by the *gp* allele. This strongly suggesting that this type of transcript is not the consequence of another gene acting *in trans*.

Therefore, we propose that the aberrant transcripts were the major reason for lower *ChlG* transcript abundance in *gp* pods compared with leaves (**Supplementary Fig. 19**). We propose that aberrant transcripts carrying an in-frame early stop codon before the final exon of *ChlG* would be expected to enter into the nonsense-mediated decay pathway, resulting in an increase in the degradation rate of *bona fide ChlG* transcripts. Nonsense-mediated decay would therefore be one of the mechanistic connections between the aberrant transcripts and the lower transcript abundance. On the other hand, these unusual transcripts may have altered stability. We suspect that when Pol II is reading from the *TIR-NBS-LRR* gene to *ChlG*, the polymerase will prevent access to the normal *ChlG* promoter simply because it is occupying that position during translation. In that sense, there will be interference with the normal initiation of translation at *ChlG*. It is also possible that the deletion in *gp* impacts an extended promoter or enhancer element upstream of the *ChlG* promoter. This model is hypothetical as the *ChlG* promoter or enhancer region has not been defined; there is ~ 3.7kb between the

deletion boundary and *ChlG*. Further studies of the regulation of *ChlG* transcription from the wild type and *gp* mutant allele are needed to resolve these issues.

Here we provide evidence for aberrant *ChlG* transcripts of various types, and lower overall transcript abundance, suggesting that the yellow pod phenotype of *gp* mutants is likely caused by some transcriptional interference mechanism.

1.3.2. Pod shape: inflated vs constricted, *P* and *V*

It is not clear whether Mendel studied the segregation of *P* or *V*. The gene *P* has been favoured because *V* is linked to *Le* (tall vs dwarf), but Mendel reported only one cross in which this phenotype segregated together with tall vs dwarf; this was not in his 1866 paper, but relates to a set of seeds, derived from a four-factor cross, which he sent to his collaborator Carl Nägeli⁷. The F₂ genotypes of the list of these F₃ seeds includes one population segregating for tall vs dwarf and inflated vs constricted pods, and the genotypic frequencies are consistent with *P* vs *p* rather than *V* vs *v* segregation⁷.

However, the similarity in the segregation ratio of tall versus dwarf plants and the presence versus absence of pod parchment in Mendel's crosses, used to determine the proportion of heterozygotes in the F₂ generation, could suggest that these segregating crosses involved the same plants in a multifactor cross⁶² where *Le* and *V* were in repulsion phase. The issue therefore remains equivocal, and Mendel may have had some *pp* and some *vv* plants; we do not know of him having performed an allelism test.

Within our panel of accessions, both *pp* and *vv* genotypes were present as is revealed in the GWAS analysis, where significant associations were found at the expected position of *P* (chromosome 1 Linkage group VI) and *V* (chromosome 5 linkage group III) (**Extended Data Figs. 2, 3**).

P

The GWAS interval corresponding to *P* was initially defined as an approximately 11 Mb interval, chr1: 374,889,451-385,871,425 in the ZW6 genome¹⁸(**Supplementary Tables 28-29**). This included the previously identified candidate *PsPSI* (*Psat01G0417600*, chr1:379,493,242-379,495,200)⁶³ encoding a pectate lyase superfamily protein. We investigated this further in the F₂ segregating population of the cross JI2822 x JI0816 (N=404, **Supplementary Tables 18-20**). *P* was mapped between the markers AX-183,563,747 and AX-183,563,750 (chr1: 380,049,894-380,967,975) within the GWAS peak, narrowing the interval a little more than 10-fold: chr1 380Mb-381Mb, a 1Mb interval (in JI2822 x JI0816). This interval excluded *PsPSI* as a candidate for *P*.

There are in total 9 genes in this interval; we screened each of these 9 for haplotype variation that could account for *p* and found that the gene *Psat01G0420500*, related to CLE41/44 or TDIF (tracheary element differentiation inhibitory factor) had one haplotype that included an in-frame premature stop codon (235A>T, R79*). The R79* allele was designated haplotype 3, and all haplotype 3 accessions (JI0033, JI0066, JI0134, JI0340, JI0466, JI0467, JI0794, JI0806, JI1041, JI1071, JI1086, JI2100, JI2273, JI2366, JI2693 and JI3120) had parchmentless pods (**Extended Data Fig. 2**). *PsCLE41* (*Psat01G0420500*) was exclusively expressed in pod, endocarp and mesocarp tissue of pod in pea (**Extended Data Fig. 2**). From the microscopic imaging of the anatomical slides of the pea lines with *PPVV*, *ppVV*, *PPvv*, and *ppvv* genotypes, we can clearly see that the *pp* mutant lines abolish the formation of the mesocarp, which is consistent with the functional role of P/*PsCLE41* in cell proliferation and differentiation (**Extended Data Fig. 4**).

V

The second major GWAS signal for the presence or absence of pod parchment defines a broad ca. 50 Mb region at chr5:594,837,588-644,213,386 (**Extended Data Fig. 3**), which

overlapped with the 1.25Mb interval defined by an F₂ population⁶⁴, corresponding to the most significant interval in our GWAS as well as the expected position of *V*⁶⁵. When the accessions carrying the R79* haplotype of *Psat01G0420500* (*P* candidate) were removed, this GWAS peak remained in the Manhattan plot and exhibited an elevated significance (**Extended Data Fig. 3**). There are in total 19 protein-coding genes in this 1.25Mb genomic interval, among which the gene *Psat05G0804500* exhibited both the highest gene expression level in pod tissue (**Extended Data Fig. 3h**) and the most significant differential gene expression in the pods of the wild-type (*PPVV*) and mutant (*PPvv*) lines (**Extended Data Fig. 3f**); in the mutant the gene is significantly down-regulated, presumably because of its mutation. Interestingly, gene *Psat05G0804500* is also exclusively expressed in pod and endocarp tissues in pea (**Extended Data Fig. 3f**). Surprisingly, all of the accessions without the *P* candidate gene R79* mutation (JI1865, JI0136, JI1863, JI2683, JI1132, JI0077, JI3605, JI2262, JI0074, JI0122, JI0061, JI2177, JI2583, JI3120, JI0066, JI0134) which scored as parchmentless carried a 23kb *Ogre*-type LTR insertion, which is 35kb upstream of the gene *Psat05G0804500*. Three accessions (JI0066, JI0134 and JI3120) carrying the R79* allele of *Psat01G0420500* also had the upstream 23kb *Ogre*-type LTR insertion of the *V* candidate, consistent with their being *ppvv* double mutants. JI0066 is a representative accession of the cultivar ‘English Sabel’, WBH33, a designated *ppvv* double mutant⁶⁶; no independent information is available for JI0134 and JI3120.

We initially considered several different candidate genes as possibly involved in pod sclerenchyma development, like the fasciclins gene cluster (*Psat05G0769200*, *Psat05G0769000*, *Psat05G0768700*, *Psat05G0768600*) possibly regulated by TDIF / PXY-1 representing *P*, the pea homologue of the interfascicular fibre mutant *Iff*⁶⁷ (*Psat05G0771900*), the gene *WRKY*⁶⁸ (*Psat05G0839000*), and *Psat05G0805200* encoding a cell wall invertase or β -fructofuranosidase. The fasciclin domain is an extracellular domain

involved in cell adhesion; and in Arabidopsis these have been shown to be involved in regulating cell wall elasticity and one class has been shown to be involved in the regulation of cell expansion. However, the haplotype clustering, haplotype-phenotype association analyses, comparative gene expression analysis, and functional investigation for all of these genes excluded them as the candidates (**Supplementary Fig. 21**). However, our haplotype-phenotype analysis suggested that the 23kb *Ogre*-type LTR insertion upstream of the *PsMYB26* (*Psat05G0804500*) can explain the parchmentless phenotype of *v*. This was further supported by our qRT_PCR, comparative RNA-seq analysis and DNA methylation analyses (**Supplementary Fig. 24**).

Summary and discussion

We have identified the R79* allele of a CLE41 peptide which fully explains the *p* allele. Therefore, we propose that *P* encodes this class of regulatory peptide, which is responsible for cell proliferation and differentiation in the development of endocarp tissue in pea. This process is governed by the CLE41/TDIF-PXY-WOX signalling pathway⁶⁹. Removing accessions carrying the R79* allele of a *PsCLE41* from the GWAS analysis abolished the GWAS peak at this position and enhanced the significance of the peak corresponding to *V*, suggesting that *P* and *V* are two independent genetic loci controlling pod parchment development.

In our *Pisum* diversity panel, we have identified 26 *vv* mutant lines with parchmentless pods (*PPvv* and *ppvv*). Only these 26 lines carried the 23kb *Ogre* insertion upstream of *PsMYB26* (*Psat05G0804500*). Furthermore, we identified 13 *ppVV* lines, all of which carried a mutation in *PsCLE41* (*Psat01G0420500*) that resulted in a premature stop codon (R79*). Both *PsMYB26* and *PsCLE41* are expressed in the pods of pea, particularly in the endocarp, as confirmed by qRT_PCR and comparative transcriptome (RNA-seq)

analysis, but their transcripts have not been detected in other organs (**Extended Data Fig. 4**).

There is a hierarchical gene regulatory network governing sclerenchyma formation and secondary cell wall biosynthesis⁷⁰. The *CLE41/TDIF-PXY-WOX* signalling pathway, well-established in *Arabidopsis*, is known to regulate stem cell fate and cambium development⁷¹. Meanwhile, *MYB26* has been reported as a master transcription factor that directly activates *NAC*-domain transcription factors: *NST1* (NAC SECONDARY WALL THICKENING PROMOTING FACTOR1) and *NST2*, as well as other *MYB* transcription factors, which are critical for inducing secondary cell wall thickening⁷². These factors activate genes involved in cellulose and lignin biosynthesis. Additional genes, such as *NST3/SND1* (*SECONDARY WALL-ASSOCIATED NAC DOMAIN PROTEIN1*), *VND6*, and *VND7*, also contribute to this pathway, with *WRKY* acting as a negative regulator of *MYB26*⁷³.

In *Arabidopsis thaliana*, *MYB26* is essential for promoting secondary cell wall thickening and anther lignification, critical for anther dehiscence and pollen release⁷⁴. Mutation in *MYB26* leads to a lack of secondary-wall thickening and impaired anther-dehiscence phenotype, while its overexpression results in ectopic secondary wall deposition⁷². In legumes, *MYB26* is required for sclerenchyma formation in pods, contributing to the loss of pod dehiscence during domestication. Loss-of-function mutations or genome-editing in *MYB26* may reduce sclerenchyma development, enhancing resistance to pod shattering^{75,76}.

Our study showed that *PsMYB26* expression was significantly downregulated in *vv* mutant lines (*PPvv*, *ppvv*) and in *pp* mutant lines (*ppVV*), whereas *PsCLE41* expression was unaffected in *vv* mutants (*PPvv*) (**Extended Data Fig. 4** and **Supplementary Tables 30-31**). These findings indicate that the *CLE41/TDIF-PXY-WOX* module functions as an upstream signalling pathway regulating the downstream *MYB26*-mediated processes. Both *PsCLE41*

and *PsMYB26* are key gene targets for domestication of additional legume species and have considerable breeding potential. The loss of the sclerenchyma formation results in parchmentless pod, which is a key feature for vegetable pods eaten fresh or lightly cooked.

1.3.3. Flower position: axial vs terminal (fasciation), *Fa* and *Mfa*

Mendel scored ‘the position of the flowers’ on the pea stem; he used the terms axial vs terminal flowers although what he was discussing was fasciation, a condition that has been known and used in pea varieties at least since Gerard’s herbal of 1597⁷⁶. The term ‘terminal flower’ is more appropriately used to refer to the *det* mutant where the usual apex (I_1) is replaced by the secondary (terminal) inflorescence (I_2); sometimes this terminal inflorescence may have a terminal flower, but usually it terminates with the I_2 stub. Although several genes are known which, when mutant, confer a fasciated phenotype in pea, only *Fa* (chromosome 4 linkage group IV) and *Fas* (chromosome 5 linkage group III) are considered candidates for the gene which Mendel studied⁷⁷. Within our panel of accessions, GWAS analysis revealed only one strong peak on the Manhattan plots (**Fig. 2** and **Extended Data Fig. 5**) corresponding to the expected peak of *Fa*; *Fas* was not obvious in these analyses. These results indicate that *Fa* rather than *Fas* is the most common determinant of fasciation.

Fa

Previously, the gene *Psat4g010240* annotated as *AP2* domain was considered as a candidate for *Fa*⁷⁸. The authors report a mutant “which displays premature termination of the shoot meristem and differentiation of the stem cells”. However, this *AP2* gene is located outside of the genetic interval and is thus excluded as a candidate. The *sym28* mutant is deficient in nodulation (leading to supernodulation) and also confers a fasciated phenotype, but is not the gene studied by Mendel⁷⁹.

Among our panel of accessions, the line JI2671 has been shown to be allelic to *fa*⁸⁰ and this is consistent with the identification of the strong GWAS signal at the end of chromosome 4 as being *Fa*. Analysis of the segregation of *Fa/fa* in an F₂ population (Caméor x JI0814) positioned *Fa* in an interval corresponding to Chr4 5-20Mb of ZW6. Fasciation also segregates in the F₂ of the cross JI2822 x JI0816 (Supplementary Tables 18-20). JI0816 is fasciated and JI2822 is wild type. The segregation of this phenotype is largely explained by a genetic locus on chromosome 4 linkage group IV which we assume to be *Fa* (**Extended Data Fig. 5a**). In the F₂, almost all fasciated plants carry the JI0816 allele (*fafa*) from a region of chr4 LGIV between the markers AX-183636277 and AX-183633456. This places *Fa* between chr4: 18,180,969 and 19,506,907 of ZW6. Combining with the data from the cross between Caméor and JI0814, we map *Fa* to the interval between 18.18 and 19.51 Mb on chr4 (**Supplementary Tables 34-35**). There are 35 genes within the *Fa* region; one of these is *Psat04G0031700* at chr4: 18,548,435 – 18,551,779, annotated as *Clavata3 insensitive receptor kinase 3*, for which it was found that all but one of the fasciated accessions have a 5bp deletion within the coding sequence (JI0208, JI0568, JI0780, JI0782, JI0795, JI0814, JI0815, JI0817, JI0819, JI0820, JI0825, JI0853, JI0958, JI1713, JI2671), the exception being JI1713 (the phenotype of JI1713 can be explained by *Mfa* as described below). The predicted truncated protein lacks the tyrosine and serine/threonine kinase domain (**Extended Data Fig. 5h**), and would thereby be unable to phosphorylate other proteins and to transmit signals between *CLV* and *WUS*, which would result in abnormal development of the shoot apical meristem and a fasciated phenotype.

Mfa

There is no position in the genetic map of JI2822 x JI0816 that co-segregates exactly with *Fa*; nevertheless, a chi-square (χ^2) test for its association with AX-183635452 (within the AX-

183636277 and AX-183633456 interval and close to ZW6 *Psat04G0031700*, **Supplementary Table 18**) is highly significant $\chi^2 = 253$, $p \sim 0$. An initial and obvious explanation for this would be mis-scoring of the fasciated phenotype. If the mis-scores were random errors then there should be no other non-random association of markers with respect to fasciation, which is not the case. There is a clear non-random association with a region of linkage group II (chromosome 6, see JI2822 x JI0816 mapping population in **Supplementary Table 18-20**). This suggests a model whereby an allele of a gene in this region masks the appearance of the fasciated phenotype.

We therefore proposed that this region carries a *Modifier of fasciation (Mfa)* such that the double recessive *fa/fa mfa/mfa* is not fasciated but wild type in appearance, or in which the signs of fasciation are subtle and easily missed. *Mfa* could not be located precisely because relatively few individuals are informative. The genetic map of the cross JI2822 x JI0816 shows that, adjacent to *Mfa*, there is a region of identity by descent (IBD) in JI0816 and JI2822 which positions one border of the *Mfa* interval at about chr6: 244,689,457 and the mapping position of the other at about chr6: 253,701,016. Interestingly, a minor peak at this position is also seen in the Manhattan plots of the GWAS analysis and G' plots from the Caméor x JI0814 F₂ population (**Fig. 2b, Supplementary Fig. 28**). The **Extended Data Fig. 7a** shows the genotyping which corresponds to these data. The fasciated lines JI0816 (WBH 1185) and JI0817 (Wisconsin-711) are genetic stocks from Stig Blixt formerly at the Nordic Gene Bank, while JI0814 is from GA Marx in the USA.

This model proposes that JI0816 is *fa/fa Mfa/Mfa* and that JI2822 is *Fa/Fa mfa/mfa*, but in the progeny of this cross *fa/fa mfa/mfa* genotypes are predicted to be wild type. In the **Extended Data Fig. 7b, c**, there is clearly a deficiency of *fa/fa mfa/mfa* genotypes scored as fasciated, and most *fa/fa mfa/mfa* genotypes scored as wild type. There are, however, several F₂ individuals which do not agree with this model, which can be further explained by mis-

scoring or by other unknown genetic interactions (for example, the flowering time that could affect the observation of fasciation phenotypes) and this was tested in the F₃, as discussed below. Note that we do not know why *Mfa* is required for fasciation to be seen; it could be that *Mfa* directly interacts with *Fa*, or it could be that it causes the plant to mature before the onset of fasciation. *PsWUS*⁸¹ is *Psat02G0115700* at chr2:74,831,848-74,832,478, which is outside this interval, and therefore is firmly ruled out as a candidate for *Mfa*.

F₃ segregation test of the F₂ discrepancies to further explain the Fa-Mfa model

We examined F₃ plants derived from selected F₂ individuals. These results are summarised below.

There are 15 F₂ individuals categorised in **Supplementary Fig. 28a** that disagree with our *Fa-Mfa* model, so we undertook F₃ progeny testing as shown in **Supplementary Fig. 28b**. For simplicity we refer to the *Fa* and *Mfa* genotypes. These are derived from marker genotypes around the mapped positions of *Fa* and *Mfa*: they are not dependent on an assumption about what DNA sequence corresponds to *Fa* or *Mfa*.

The JI0816 alleles are *fa* (Y) and *Mfa* (Y) while the JI2822 alleles are *Fa* (G) and *mfa* (G). The single letter codes correspond to the parent of origin of the allele and the genotype of homozygotes. Heterozygotes *Fa/fa* or *Mfa/mfa* are represented by H. The wild-type phenotype, which is dominant, is coded as (D) and the recessive fasciated phenotype is (R). *Fa/Fa* and *Fa/fa* plants are always wild type, and wild type is dominant to fasciated. However, *fa/fa* is not always fasciated; plants that are also *mfa/mfa* can appear to be wild type. However, *Fa/_ Mfa/_* are wild type. The allelic state of *Mfa* is not, on its own, telling us anything about the phenotype; all it does is qualify whether or not *fa/fa* plants are fasciated.

The detailed analyses were described below (**Supplementary Fig. 28**):

1. The F₂ plant F1-7A B2 (this is the category ‘a’ as designated in **Supplementary Fig. 28**) was scored as wild type and the markers indicate it was: *fa/fa Mfa/mfa*, i.e. Y H. The selfed F₃ of this plant is expected to segregate 3 : 1 fasciated to wild type where the wild type plants are *fa/fa mfa/mfa*:

The F ₃ of this plant is expected to segregate	3:1	<i>fa/fa Mfa/_</i>	:	<i>fa/fa mfa/mfa</i>
Phenotypically	3:1	R fasciated	:	D WT
Observed		2		8
Expected		7.5		2.5

2. The F₂ plant F1-3 D5 was scored as *fasciated* (this is the category of ‘b’ as designated in **Supplementary Fig. 28**), but the markers were: G H indicating *Fa/Fa* and *Mfa/mfa* so this should have been wild type.

The F ₃ of this plant is expected to segregate	3:1	<i>Fa/ Fa Mfa/_</i>	:	<i>Fa/ Fa mfa/ mfa</i>
Phenotypically	3:1	D	:	D

All ten F₃ plants were wild type as expected. This is consistent with all these plants being *FaFa* homozygotes so not fasciated. We conclude that there was an error in the F₂ phenotype score for this individual.

3. The F₂ plants F1-3 C4 and F1-3 C10 were scored as fasciated (this is the category of ‘c’ as designated in **Supplementary Fig. 28**), but the markers were: G Y indicating *Fa/Fa* and *Mfa/ Mfa* so this should have been wild type. As the F₂s were homozygous there should be no segregation in the F₃ and all F₃ plants should be wild type. All 20 F₃ plants were wild type as expected. We conclude that there was an error in the F₂ phenotype score for these individuals. *Fa* is not segregating and all plants are *FaFa* homozygotes so not fasciated.

4. The F₂ plants F1-3 A8, F1-3 D1 and F1-3 F4 were scored as fasciated (this is the category of 'd' as designated in **Supplementary Fig. 28** – the 'd' is one plant which provided no F₃ seed), but the markers were: H Y indicating *Fa/fa* and *Mfa/Mfa* indicating this should have been wild type. This genotype is expected to segregate 3 : 1 wild type to fasciated in the F₃.

The F ₃ of these plants is expected to segregate	3:1	<i>Fa/_ Mfa/Mfa</i>	:	<i>fa/fa Mfa/Mfa</i>
Phenotypically	3:1	D	:	R
Observed		16		12
expected		21		7
$\chi^2 = 4.76, p = 0.03$				

There is a slight excess of fasciated plants in the F₃, but the important point is that fasciation is segregating and therefore the F₂ plants were not *fa/fa Mfa/Mfa*. This is consistent with the genotypic scores, suggesting that the F₂ individual was mis-scored phenotypically. In this group *Mfa* is not segregating. What we see is consistent with 3:1 segregation *Fa_* : *fafa*, as expected.

5. Ten F₂ plants F1-10 C1, F1-10 G3, F1-3 B2, F1-3 F3, F1-3 H4, F1-4A B11, F1-4A B8, F1-4A E11, F1-5 E4 and F1-7B H10 were selected for further analysis from 54 F₂ individuals scored as fasciated and which were genotyped as Y H, i.e. these were genetically *fa/fa Mfa/mfa* consistent with them being fasciated. We predicted that a quarter of the F₃ should be double recessive and therefore wild type.

The F ₃ of these plants is expected to segregate	3:1	<i>fa/fa Mfa/_</i>	:	<i>fa/fa mfa / mfa</i>
Phenotypically	3:1	R	:	D
Observed		79		15
expected		70.5		23.5
$\chi^2 = 4.1, p = 0.04$				

There is a slight excess of fasciated plants in the F₃, but the important point is that some wild types were observed, so fasciation is segregating. In other words, because of the segregation of *Mfa* vs *mfa*, fasciation is not true breeding in the F₃.

All of these F₃ plants are *fa/fa* homozygotes, so should be fasciated, however, a quarter of the F₃ plants are expected to be *mfa/mfa* homozygotes. If *fa/fa mfa/mfa* homozygotes are scored as wild type then we expect 3 : 1 segregation of fasciated to wild type. There is a slight deficiency of wild types, but this could be due to chance, or due to *fa/fa mfa/mfa* homozygotes sometimes showing signs of fasciation.

Summary for the above analyses

It is clear that all fasciated plants are *fa/fa* homozygotes. Some F₃ plants that are offspring of *fa/fa* F₂ homozygotes appear to be wild type. This surprising result is consistent with the suggestion that *mfa* is a modifier of fasciation and that the *fa/fa mfa/mfa* double recessive can appear to be wild type.

The next group of F₃ plants to be discussed are all derived from F₂ *fa/fa mfa/mfa* double recessives, but do not always have the expected appearance (expected to be wild type phenotype).

We need to recognise that what we are calling fasciated has several different phenotypic features. One is the bunching of flowers at the apex rather than having them distributed along the stem. Another is the broadening and flattening of the stem, associated with the production of extra vascular strands, and sometimes the formation of extra lateral organs (leaves or inflorescences). There may be a degree of independence in the conditioning of these phenotypes. For example, early flowering genotypes may form flowers at lower nodes, but this will not happen in late flowering genotypes (we know the JI2822 early flowering locus *lf-a* is segregating in this population). Similarly, early

flowering genotypes may mature before the distortion of the behaviour of the apex becomes apparent.

F₃ plants derived from *fa/fa mfa/mfa* double recessive F₂ individuals are discussed below:

6. The seven F₂ plants (e in **Supplementary Fig. 28**) F1-10 H3, F1-3 H6, F1-4A C2, F1-4A F8, F1-6 C2, F1-6 H1 and F1-7A B1 were scored as fasciated but were predicted to be wild type by their genotype. The F₃ phenotypes were therefore of interest. As the F₂ plants were homozygous recessive at both loci there should be no segregation in the F₃ generation, which is expected to be entirely wild type based on the hypothesis of the role of *mfa*. The F₃ phenotypes are tabulated below:

All these F ₃ plants are expected to be				<i>fa/fa mfa / mfa</i>
Phenotypically		R	:	D
Observed		51		14

This observation suggests that among these F₃ plants the double recessive is capable of generating wild type plants, but mostly they were fasciated.

7. Ten plants F1-10 A10, F1-3 A2, F1-4A A11, F1-4A G4, F1-4B F5, F1-4B H5, F1-6 A2, F1-6 B10, F1-6 E11 and F1-7A H11 were selected from the 30 double recessive F₂ plants, predicted to be wild type and which were scored as wild type in the F₂. For the two loci of interest these plants are genetically the same as in 6 above (*fa/fa mfa/mfa*), but had different F₂ phenotypic scores. The F₃ phenotypes are given below:

All these F ₃ plants are expected to be				<i>fa/fa mfa / mfa</i>
Phenotypically		R	:	D
Observed		70		18

Again this observation suggests that among these F₃ plants the double recessive is capable of generating wild type plants, but mostly they were fasciated. The proportion of

wild type and fasciated plants is not statistically different in groups 6 and 7 and is not significantly different from a 3:1 ratio.

There is no marker anywhere else in the genetic map which is heterozygous in all 17 of these F₂ plants, so segregation of an unlinked marker cannot explain the segregation of fasciation in F₃ progeny.

Summary and discussion

In our GWAS analysis, we found no strong evidence for the gene *Fas*, reportedly on chromosome 5 linkage group III. The line JI2771, reportedly carrying *fas*, was not included in these studies⁷⁷.

The *fa* allele is strongly associated with a 5bp deletion in the gene *Psat04G0031700*, annotated as *Clavata3 insensitive receptor kinase 3*, designated here as *PsCIK2/3 (fa)*. The Arabidopsis orthologue of *Psat04G0031700* (*PsCIK2/3*) is closely related to *AtCIK2/3*, which plays an important role in *CLV3*-mediated regulation of apical meristem homeostasis⁸². The *atc1 2 3 4* quadruple mutant also has a fasciated phenotype⁸³. Through transient expression of the pea orthologue in leaves of *Nicotiana benthamiana* and subcellular localization, it was found that the *PsCIK2/3* is located on the cell membrane, as expected for co-receptor kinases (**Supplementary Fig. 27**). Based on differential gene expression patterns by qRT_PCR and RNA-seq analysis, and as well as yeast two-hybrid experiments among *PsCIK2/3*, *PsWUS*, *PsCLV1*, *PsCLV2*, and *PsCLV3* in various tissues in pea, we can propose that *PsCIK2/3* may interact at least with *CLV1* and *CLV2* (**Supplementary Fig. 27**) that are located on the membrane, thereby regulating the expression pattern of *WUS*. CLAVATA (*CLV*)–WUSCHEL (*WUS*) plays an important role in the proliferation and differentiation of stem cells maintained by a negative feedback loop⁸⁴⁻⁸⁶. Therefore, a *CLV3-WUS* conserved regulatory network can be proposed in pea

in which *PsCIK2/3* is involved, the biochemical genetics of which requires further investigation.

However, the penetrance of this mutant phenotype appears to be modified by an unknown gene on chromosome 6 linkage group II which we have designated *Mfa*. This model suggests that *Mfa* may be involved in meristem maintenance although the *mfamfa* mutant, in an otherwise wild-type background, is phenotypically wild type.

The analysis of the F₃ generation of this population was undertaken to check our interpretation on the *Fa-Mfa* model and some discrepancies (**Supplementary Fig. 28**). Six F₂ plants were scored as fasciated, but genotyping suggested that these should be heterozygous at either *Fa* or *Mfa*; F₃ progeny tests were consistent with these being phenotypic mis-scores and in agreement with the F₂ genotypes. Ten fasciated F₂ plants were genotypically *fa/fa Mfa/mfa* and should therefore segregate for fasciation in the F₃, due to the segregation of *Mfa* vs *mfa*; this was confirmed.

According to the marker data, seven F₂ plants were predicted to be *fa/fa mfa/mfa* (which should be wild type in appearance based on our proposed *Fa-Mfa* model) but were scored as fasciated. Their F₃ progeny segregated 51:14 for fasciated vs wild type, suggesting that *mfa/mfa* homozygotes do not always mask the fasciated phenotype of *fa/fa*. Conversely, ten F₂ plants predicted to be *fa/fa mfa/mfa* scored as wild type, which were expected to have only wild type offspring in the F₃, but again fasciation segregated (70:18 for fasciated vs wild type) at F₃; both results are consistent with the effect of an additional complex genetic locus (designated as the third unknown locus). Nevertheless, it is clear that *fa/fa mfa/mfa* genotypes can appear to be wild type, but the effect of *mfa* is either incompletely penetrant or dependent on the third locus, as is suggested by the approximately 3:1 segregation (121:32, $\chi^2 = 1.36$, $p = 0.24$).

There are several different possible explanations to interpret the dataset and analysis results.

Firstly, incomplete penetrance, and/or variable expressivity of *mfa*. *fa/fa mfa/mfa* genotypes have a tendency to appear wild type, but do not always do so. This may be purely by chance, or it may be influenced by other genetic loci, for example in the regulation of node to flower initiation. The fasciated plants in the F₃ of Groups 6 and 7 derived from *fa/fa mfa/mfa* F₂ individuals consistently exhibited floral inflorescences at nodes below the apex; up to six such nodes were found, but typically there were 4. This gives the appearance of ‘axial’ rather than ‘terminal’ flowers and, without careful examination of the upper part of the plant, the appearance of fasciated structures would be missed, so they might easily have been wrongly scored as wild type. Furthermore, obvious signs of fasciation (e.g. a pair of leaves at a node, two inflorescences at a node) were rarely seen. In plants with multiple shoots, fasciation was sometimes not seen on all shoots. Thus, it is possible that some plants simply do not develop fasciated structures, although they may be capable of doing so. This suggests that homozygous *mfa/mfa* delays the expression of the fasciated phenotype, in which case, plants are scored as wild type, but when expression of fasciation is not sufficiently delayed, double homozygotes are scored as fasciated.

Secondly, we may have failed to locate *Mfa* correctly. *Mfa* maps close to a region devoid of genetic markers distinguishing JI0816 and JI2822. We have interpreted this to mean that this segment of the genome is identical by descent (IBD), and therefore that JI0816 and JI2822 carry the same alleles throughout the region. We have rejected the possibility that *Mfa* lies within this region because *mfa* vs *Mfa* segregates: according to our hypothesis, there are two alleles. However, it is possible that *Mfa* lies within this region and is in fact different between JI0816 and JI2822. For the 17 F₂ parents of groups

6 and 7, all F₃ plants carry the JI2822 allele where we presume *mfa* to be located. Of these, 11 also have the JI2822 allele at the other side of the IBD region, so we infer that these carry the JI2822 allele throughout the region. The six remaining plants are heterozygotes at the other side of the IBD region. If we propose the existence of heterozygosity at some position within the 15cM IBD region in these 11 plants, this means there must be double recombination events on both chromosomes within the IBD region (11 F₂ individuals, 22 recombination events). Double recombination is very unlikely to be found within a 15cM interval, especially in all 11 individuals. For the six other plants, the double recombination event is needed on only one of the two chromosomes. If we say that 15 cM corresponds to a recombination fraction of about 20% (0.2), then the chance of there being two recombination events would be $0.2^2 = 0.04$ (probably lower because of crossover interference). For this to happen in 11 plants there would have to be two recombination events within the IBD region on both chromosomes. The probability of this occurring is 0.04^{11} or ca. 2×10^{-31} . Therefore, it seems highly unlikely that *Mfa* lies within the IBD region.

Thirdly, there may be some genetic instability, such as the inheritance of an epigenetic mark. If the difference between *Mfa* and *mfa* is an epigenetic mark, then it may not be accurately reflected by the Axiom marker genotype. However, we should note that in the genotypes of Group 1 to 5, the genetic markers do accurately reflect the behaviour predicted for *Fa* and *Mfa*. Therefore, the F₃ progeny tests broadly confirmed the *Fa-Mfa* model, but not all *fa/fa mfa/mfa* genotypes were wild type. Intriguingly, the *fa/fa mfa/mfa* genotypes segregate 3 : 1 in favour of fasciation (vs wild type). Whether this is a coincidence or indicative of an additional segregating gene is not known. We can say that all fasciated plants are *fa/fa* homozygotes, but whether all *fa/fa* homozygotes are fasciated depends on other factors (such as the complex genetic interaction or unknown

environmental factors). Many F₃ plants that were clearly fasciated had flowers at multiple nodes, and these could easily be mistaken for wild types rather than fasciated forms. Furthermore, several fasciated F₃ plants were branched with several stems, not all of which showed signs of fasciation, further indicating that *fa/fa* homozygotes can sometimes maintain a normal meristem size.

1.4. Quantitative traits that Mendel commented on

Mendel's work was mostly driven by his quest to uncover the rules of heredity, but his findings in genetic inheritance, as well as the traits he chose for the hybrid experiments, have great practical breeding values. Many of the divergent traits Mendel studied correspond to market classes crucial for pea improvement and breeding. Mendel also commented upon several quantitative traits with genetic complexity.

1.4.1. Axil ring pigmentation: *D* vs. *d*

Mendel noted that axil ring pigmentation is absent in white flowered varieties. A dominant allele of the gene *D* is required for the presence of axil ring pigmentation⁶⁶. In *Pisum abyssinicum* flowers are pigmented, but axil ring pigmentation is absent; this taxon is a naturally occurring *d* mutant^{50,66}. There however are several alleles of *D*: *D^w* *D^{co}* *D^{tet}* *D^{ma}* and *d* with distinct phenotypes, as illustrated in **Supplementary Fig. 29**.

Our GWAS study uncovered a single major peak in Manhattan plots, at chr2:91435801-110725232 in the ZW6 assembly (**Supplementary Fig. 30, Fig. 4c**). This complements a genetic analysis of the segregation of *D* vs *d* in a recombinant inbred population of a *P. abyssinicum* x *P. sativum* cross which placed *D* in an interval corresponding to Chr2:97476978 (AX-183581090) and Chr2:113478796 (AX-1835643150), overlapping with the GWAS signal (**Fig. 4d**). This region includes several genes encoding *MYB* transcription factors⁵⁰, forming a *MYB* gene cluster.

A VIGS experiment was undertaken with the gene *Psat02G0138300* annotated as *MYB*-type HTH DNA-binding domain profile in the Caméor genome assembly. This is within the GWAS interval but is excluded as a candidate for *d* by the genetic mapping and Fast Neutron mutant analysis. Nevertheless, this experiment provided an interesting further insight into ring pattern formation as shown in **Supplementary Fig.** , showing that the *MYB* gene regulating the outer ring of pigment is different from that regulating the presence or absence of this pigment.

In addition to the *D* allelic variation in the diversity panel (**Supplementary Table 41**), we obtained three Fast Neutron induced mutants (FN1091/4, FN1218/6, and FN2073/5), in the JI2822 background⁸⁷, which exhibited the *d* phenotype (**Supplementary Fig. 32**). Allelism tests were performed among these and with respect to JI0073 (a multiply marked line) and JI2202 (*P. abyssinicum*). The phenotypes of the axil rings of these accessions and F₁ plants are shown in **Supplementary Fig.** ; the results indicate that FN1091/4 and FN1218/6 were allelic to *d*, but FN2073/5 was not.

These studies have defined *d* as corresponding to either *PsMYB104* or *PsMYB106*. The line FN2073/5 remains to be characterised, but this line has a deletion nearby that includes *Psat02G0252700*, a *MYB* gene at chr2:221575409-221576389 in the ZW6 assembly. (This gene was not described by Yang et al. 2022b)¹⁸. Świącicki (1990)⁸⁸ obtained a genetic map position for *And* with a recombination fraction corresponding to 15.5 ± 4.5 cM with respect to *D*. The positions of *Psat02G0252700* and *PsMYB106* on the genetic map of JI0281 x Caméor correspond to a distance of ca. 22 cM, so it is possible that a mutation in *Psat02G0252700* corresponds to this *d*-like phenotype and an allele of *And*. The deletion event was confirmed by PCR and read depth variation analysis (**Supplementary Fig. 33**). The allelic haplotype and association analysis of each of the *MYB* genes were given, with *PsMYB104* highlighted in **Supplementary Fig. 34**.

Summary and discussion

The results presented here suggest that *PsMYB104* and/or *PsMYB106* are required for axil ring pigmentation as is another unknown gene defined by FN2073/5. Furthermore, the *MYB* related gene *Psat02G0138300* has a role in patterning this pigmentation. It has been noted⁶⁶ that not only are there multiple alleles of *D*, but that there are also many spontaneous conversions from one allelic form to another. These spontaneous interconversions may represent recombination or conversion events within this close association of functionally related genes.

1.4.2. Organ size: pod width and seed weight, *PsOs1*

A correlation analysis within our panel shows a significant and previously unrecognised positive relationship between pod width and seed weight ($r = 0.65$). In addition, we analysed an F₂ population from the cross between JI1995 (narrow pod and small seed, female) and JI0074 (wide pod and large seed, male), which identified a QTL on chromosome 2 that aligned with the GWAS peak (**Fig. 4f** and **Supplementary Fig. 35-Supplementary Fig. 36**) and in the approximate location of a previously identified seed size QTL⁸⁹. Fine mapping of the pod width QTL narrowed down the interval to a ca. 1Mb region containing 11 genes in the ZW6 reference genome (**Fig. 4g**). Transcriptome data in pods and qRT-PCR analysis revealed that *Psat02G0011300* was the most significantly differentially expressed gene in a comparison between the parental lines (**Supplementary Fig. 37**).

Transverse sections of pods (12 days post-flowering) in JI0074 and JI1995 showed that cell size in JI0074 was notably larger than that in JI1995 (**Supplementary Fig. 38**). Flow cytometry revealed a relative increase in the number of cells in the G2/M phase in JI0074 compared to JI1995, indicating cell cycle suppression (**Supplementary Fig. 38**). Flow cytometry also revealed a decreased proportion of cells in the G2 phase in *PsOs1*-

silenced lines. These studies suggest that *PsOs1* inhibits the cell cycle at the G2/M transition and induces endoreduplication, a process potentially crucial for rapid fruit growth by increasing cell size^{90,91}. The number of higher ploidy nuclei and endoreduplication indices were significantly reduced in VIGS-*PsOs1* silenced lines as compared to the control VIGS-Con lines (**Supplementary Fig. 39**).

We introduced the coding sequence of *PsOs1* from the broad-pod variety, JI0074, into Arabidopsis (Col-0) under control of the CaMV 35S promoter, resulting in stable transgenic plants. The siliques of the overexpression lines were not only significantly wider than those of the wild type but also featured mature seeds that were larger and heavier, albeit with a shorter silique length compared to the wild type (**Supplementary Fig. 40**). Flow cytometry data demonstrated an increased proportion of higher ploidy cells and a significant elevation in the endoreduplication index in the overexpressing lines. The relationship between pod and seed size is moderate for pea ($r = 0.65$). The transgenic Arabidopsis lines suggest that both traits are altered together under the influence of *Psat02G0011300*.

In light of significant differences in the expression of *PsOs1* between the parents, we performed Sanger sequencing of the *PsOs1* coding region from JI0074 and JI1995, and cloned a 3kb promoter region upstream of the coding sequence from both parents. We did not find any functional variation within the coding sequences but identified multiple differences in the promoter region (**Supplementary Fig. 41**), notably at positions chr2:5261351 and chr2:5261587, with the most significant SNPs detected in both pod width and seed weight GWAS analysis ($7.701225E^{-19}$ and $1.44E^{-19}$ respectively) (**Fig. 4**). We classified our pea diversity panel into two haplotypes based on pod width: Narrow (N) including JI1995, and Wide (W) including JI0074. Pod width and seed weight were significantly greater in W than N (**Supplementary Fig. 41**). qRT-PCR results showed that *PsOs1* transcripts were more abundant in wide-pod samples compared to N

(**Supplementary Fig. 41**). Variation in the promoter region of *PsOs1* was consistent with its role in driving these phenotypic changes. Evaluation of *PsOs1* promoter activity using β -glucuronidase (GUS) constructs and expression analysis revealed that the JI0074 promoter exhibited higher activity than the corresponding JI1995 promoter. Expression analysis by qRT-PCR indicated predominant expression of *PsOs1* in pea pods (**Supplementary Fig. 42b**), further confirmed by GUS staining in transgenic *Arabidopsis* plants expressing *PsOs1* promoter-driven GUS in roots, trichomes, sepals, and siliques, but not in seeds (**Supplementary Fig. 42**).

1.5. Agronomic traits to global breeding objectives

In this paper, our narrative focuses on the traits originally studied by Gregor Mendel. These are the traits and studies that laid the foundation of the field of modern genetics. These traits, we believe, are of special interest to scientists and educators across many fields and disciplines. Naturally, when working on Mendel's traits, the complexity that we reveal and discuss (newly discovered genes or novel allelic variations of previously known genes) is augmented by their historical significance. These findings are not just current discovery, but draw together work since the dawn of genetics with state-of-the-art in functional genomics. Nevertheless, our work is also a foundation for a revolution in contemporary pea crop breeding.

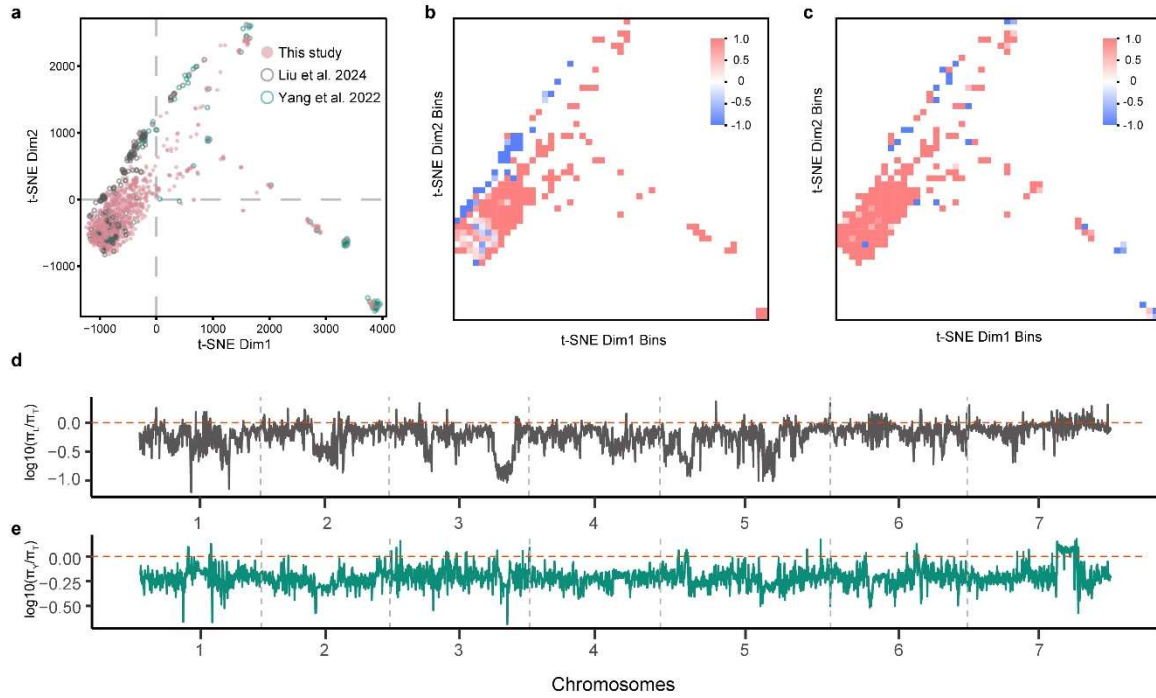
We deliver open-source genomic data in unprecedented depth and breadth, tightly linked to open access germplasm resources. The capability to associate a genomic region with a measured trait is key for genomic assisted breeding and genomic predictions, including AI-driven algorithms. While underlying causal genes are of secondary importance for breeders, our novel dataset enables easy association of a trait with a region spanning, on average, 12 Mb or ca. 150 protein-coding genes. This was shown by ~80 trait associations (**Extended Data Fig. 8, Fig. 4, Supplementary Tables 37-38**), dozens of which are taken

from the protocols of the International Union for the Protection of New Varieties of Plants. These traits are pivotal for defining new pea varieties and foreshadow future genomics-based Distinctness, Uniformity, and Stability (DUS) variety testing.

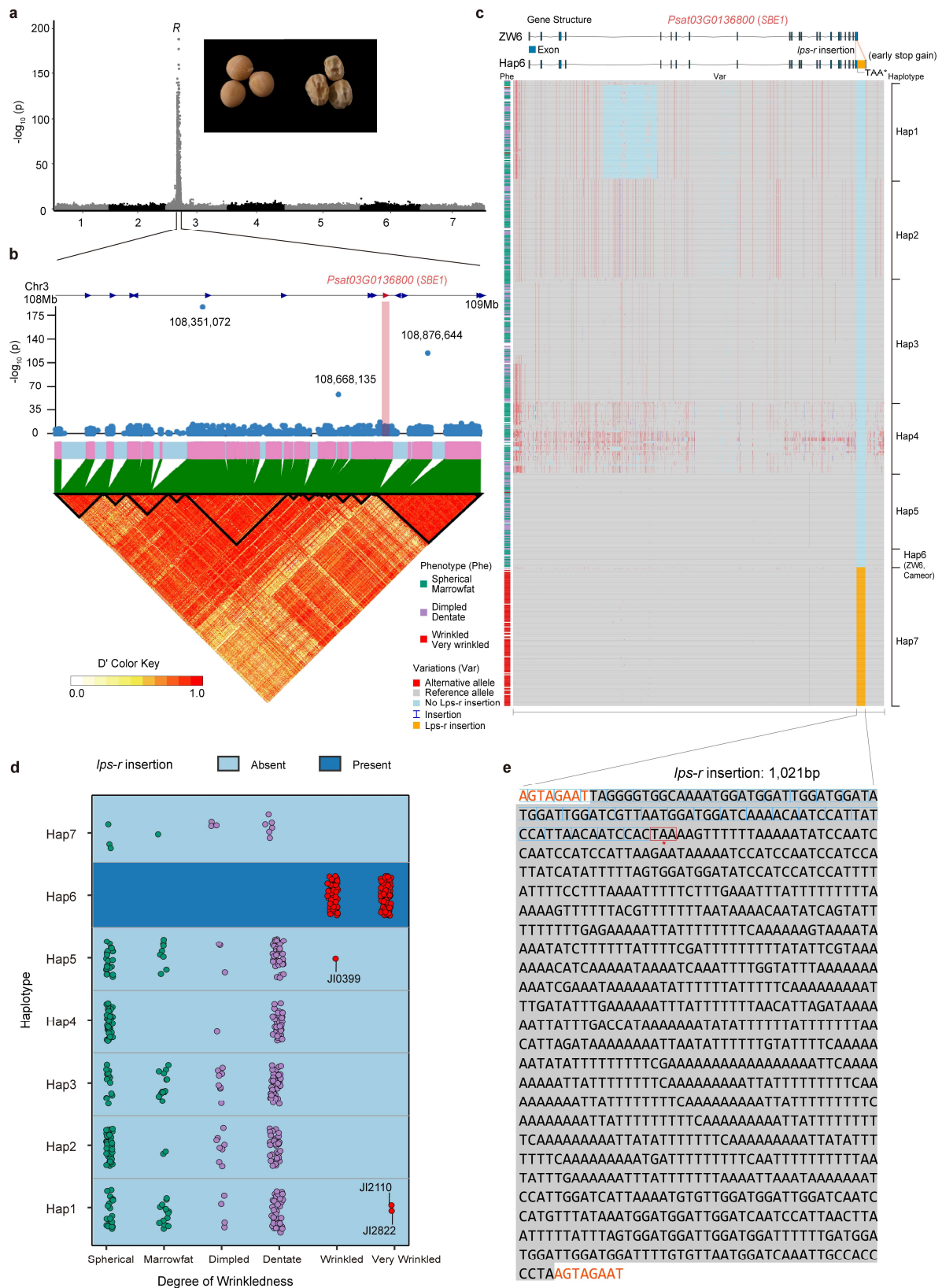
In the current manuscript, we illustrate this utility for breeding as described above and in the Main Text. For example, we identified the clear endpoints of deletion haplotypes with different combinations of deleted genes associated with the ‘semi-leafless’ phenotype, a trait essential in modern pea breeding. Additionally, we discovered a novel organ size gene that controls the size of the mature seed. This is important for breeding: 1. As a seed component for yield improvement; 2. As separate market classes in end use applications; 3. For consumer preference.

We also report on trait association with Seed Protein Content (Supplementary Fig. 3). Arguably, protein production is the most important trait for pea (as for other crops) in contemporary breeding to improve global sustainable food production and mitigate the negative effect of agriculture on climate change. Taking Europe as an example, legumes are key to future agriculture for their potential to reduce carbon emissions by assimilating atmospheric N₂. N₂ fertilisers contribute to 25% of the total agricultural greenhouse gas emissions in the EU⁹². Hence, intensifying legume production would lower carbon pollution in Europe (and globally). However, in 2018⁹³, the EU reported a failure in achieving this goal: despite their agronomic and environmental benefits, legumes still account for only 3% of EU agricultural land area. Our work has high (and industrially recognised) potential to intensify industrial pea breeding and therefore legume production.

2. Supplementary Figures

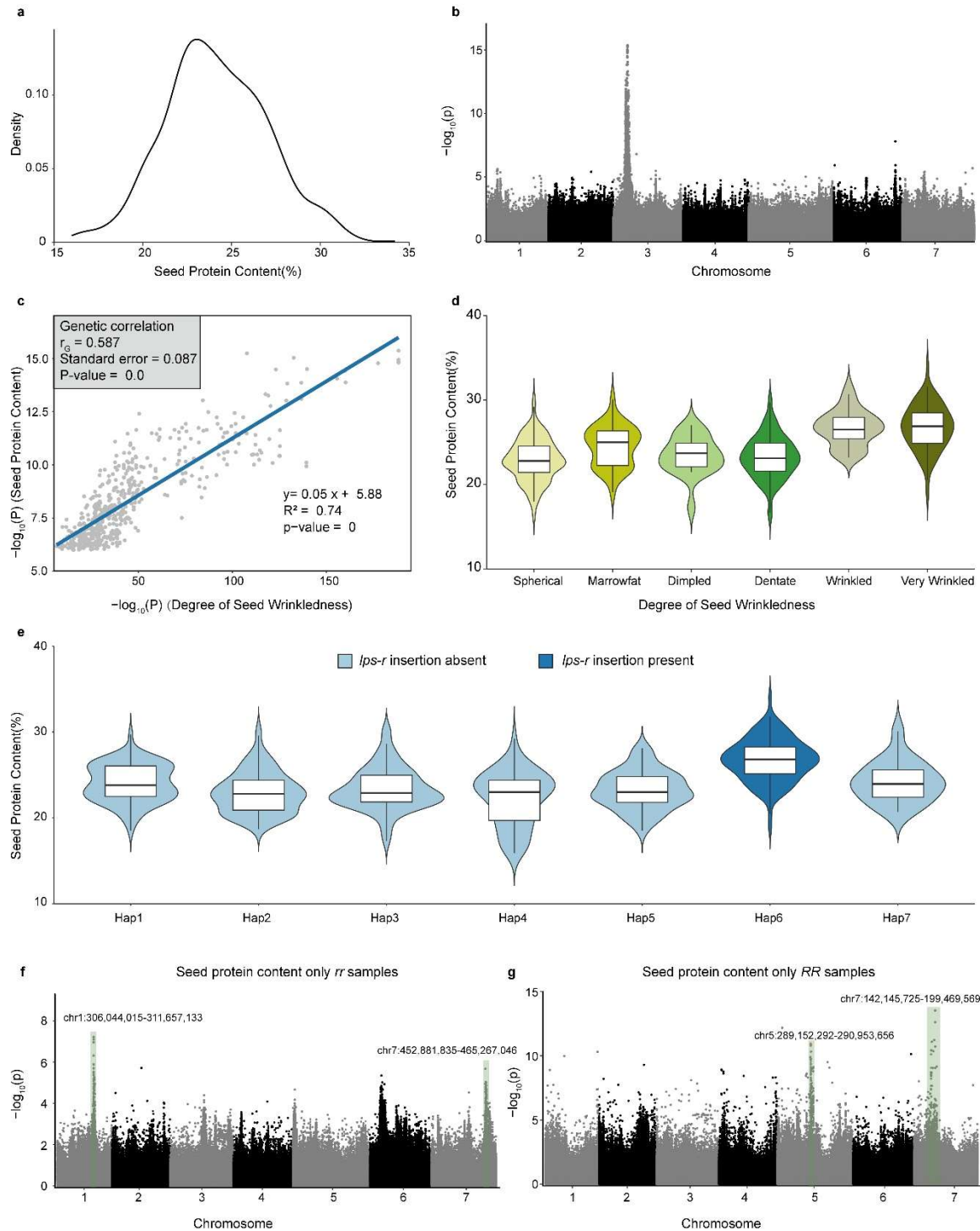


Supplementary Fig. 1 | Comparison of genetic diversity (π). (a) t-SNE plots showing Dim1 and Dim2 on the merged whole-genome SNP matrix from this study (Mendel pea G2P), Liu et al. (2024)¹⁹, and Yang et al. (2022)¹⁸. (b) Two-dimensional t-SNE results from panel a are divided into 50 x 50 rectangles. For each rectangle, a difference ratio (DI) was calculated using the formula $DI = (\text{This study} - \text{Liu et al. 2024}) / (\text{This study} + \text{Liu et al. 2024})$. Rectangles with $DI > 0$ are shown in red, indicating higher values for this study, while those with $DI < 0$ are shown in blue, indicating higher values for Liu et al. (2024). This allows a clearer view of the difference distribution between the two groups. (c) As in panel b, but using the formula $D2 = (\text{This study} - \text{Yang et al. 2022}) / (\text{This study} + \text{Yang et al. 2022})$. Rectangles with $D2 > 0$ are shown in red, while those with $D2 < 0$ are shown in blue, providing a clearer view of the difference distribution between this study and Yang et al. (2022). (d) Sliding window analysis (bin size = 100 kb) comparing genetic diversity (π) across all seven chromosomes between this study and Liu et al. (2024). A value of $dI = \log_{10}(\pi(\text{This study})/\pi(\text{Liu et al. 2024}))$ was calculated for each window. Positive dI values ($dI > 0$) indicate higher diversity in this study. (e) As in panel d, but comparing genetic diversity between this study and Yang et al. (2022) using the formula $d2 = \log_{10}(\pi(\text{This study})/\pi(\text{Yang et al. 2022}))$. Positive $d2$ values ($d2 > 0$) suggest higher diversity in this study.



Supplementary Fig. 2 | Haplotype-phenotype association study for seed shape (round vs. wrinkled). (a) Manhattan plot of the genome-wide association study (GWAS) using the ZW6

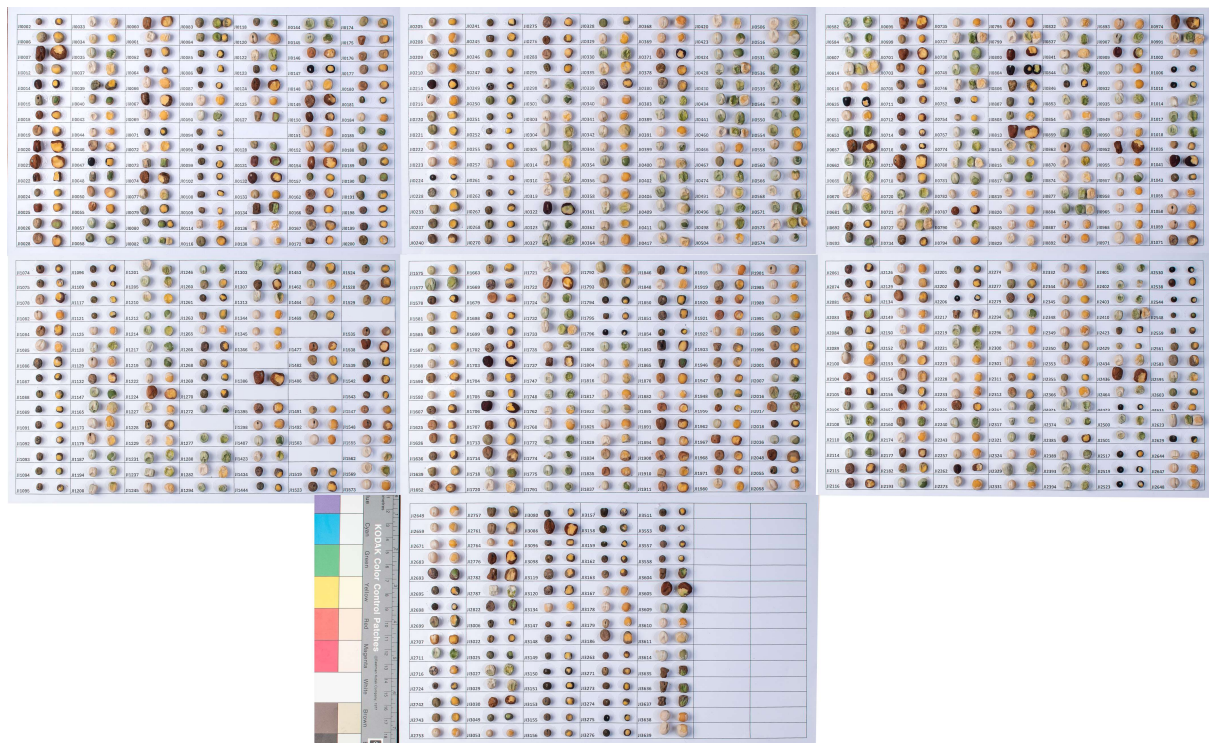
genome assembly as the reference to identify genetic variants explaining the round vs. wrinkled phenotype. The single strong signal corresponds to the *R* locus located on the long arm of chromosome 3 (Chr3), as shown in the illustration. **(b)** Local detail of the significant genomic interval shows the gene list, with *Starch Branching Enzyme 1* (*SBE1*, *Psat03G0136800*, chr3:108,732,329-108,770,718) gene highlighted in red. Below, the local linkage disequilibrium (LD) map shows the data points corresponding to *SBE1* (at 108,732,329-108,770,718) marked in red. The quite low significance values of these data points suggest that the associated SNP variants here are not causative and presumably do not distinguish the *r* alleles from the wild type *R* progenitor. **(c)** Population-based haplotype clustering and haplotype-phenotype association analysis of *SBE1*. Most accessions with wrinkled seeds cluster within Haplotype 6, consistent with the causal variation involving the *Ips-r* insertion event (marked in orange) in the last exon. Note that this haplotype lacks unique SNPs. A few accessions with wrinkled seeds distributed elsewhere are caused by the *rb* allele^{28,94} (JI0399, JI2822) or the previously described variant in JI2110, cv Kebby (see also panel **d**). **(d)** Distribution of the different seed shape phenotypes recorded in this study, categorized into six classes: Spherical (green dots), Marrowfat (green dots), Dimpled (purple dots), Dentate (purple dots), Wrinkled (red dots), and Very Wrinkled (red dots)). These phenotypes correspond to the haplotypes identified in panel **c**. **(e)** Full sequence of the *Ips-r* element (1,021 bp) with the 8 bp site duplication (AGTAGAAT) sequences bounding the insertion event are highlighted in red. The extension of exon 22 into *Ips-r* is indicated by the blue boxes around the codons and the premature termination codon is boxed in red.



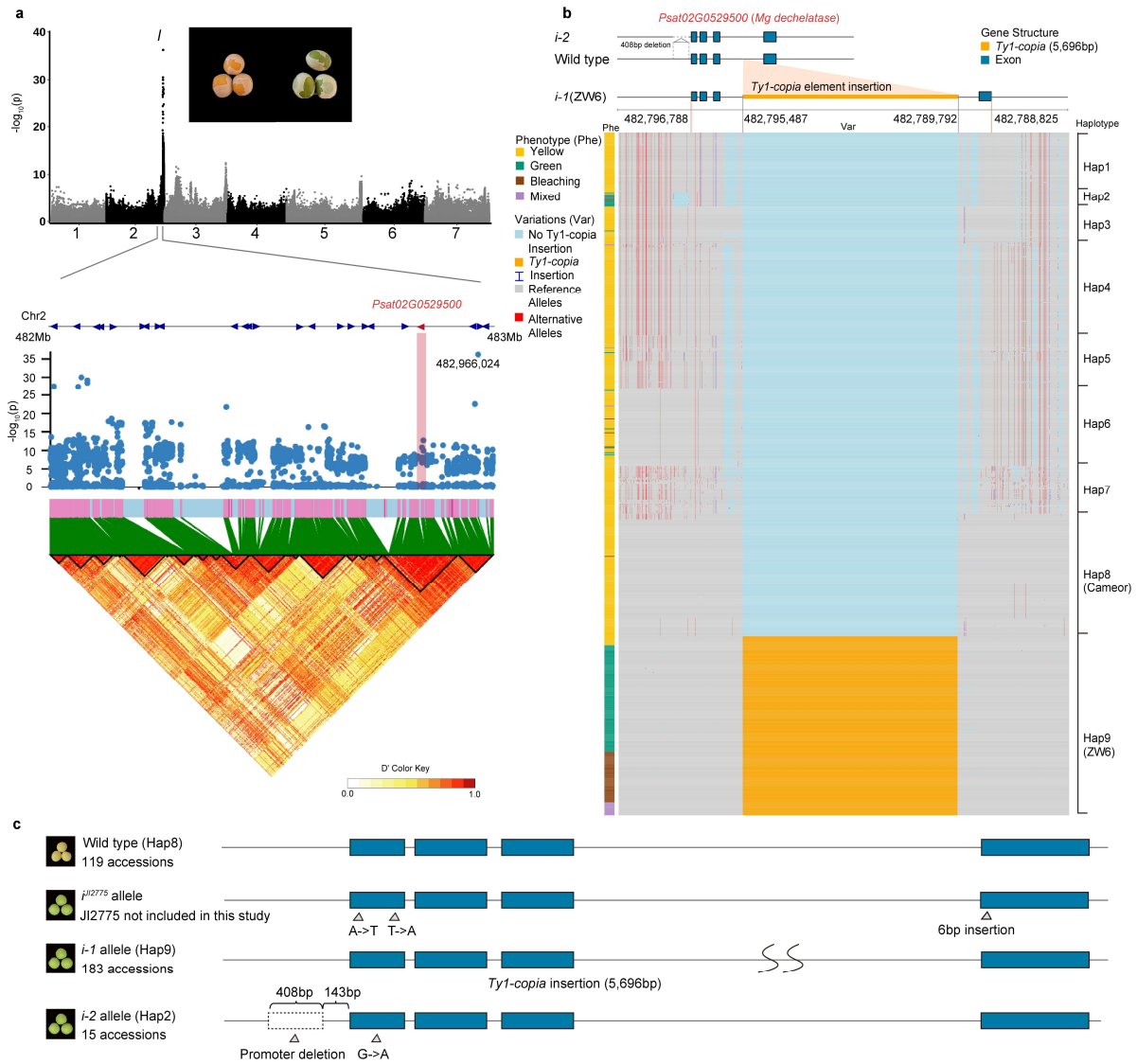
Supplementary Fig. 3 | Genetic correlation between seed shape and seed protein content.

(a) Frequency distribution of seed protein content (%) measured across the diversity panel, shown as a density plot; **(b)** Manhattan plot of the GWAS analysis for seed protein content using the entire diversity panel in this study; **(c)** Regression analysis of the candidate genomic interval on Chr3 (90-140 Mb) identified by GWAS for both seed shape (wrinkledness) and seed protein content. The grey box summarises the genetic correlation between the two traits;

(d) Distribution and comparison of the seed protein content (%) across groups with different levels of seed wrinkledness; (e) Distribution and comparison of seed protein content (%) among the haplotypes identified in **Supplementary Fig. 2c** (Hap1-Hap7), using only accessions of *Pisum sativum*; (f) Manhattan plot of GWAS for seed protein content using only accessions with *rr* (wrinkled seeded) alleles; and (g) Manhattan plot of GWAS for seed protein content using only accessions with *RR* (round seeded) alleles. Subset GWAS analyses in panels **f** and **g** reveal additional loci associated with seed protein content (highlighted with green shading).

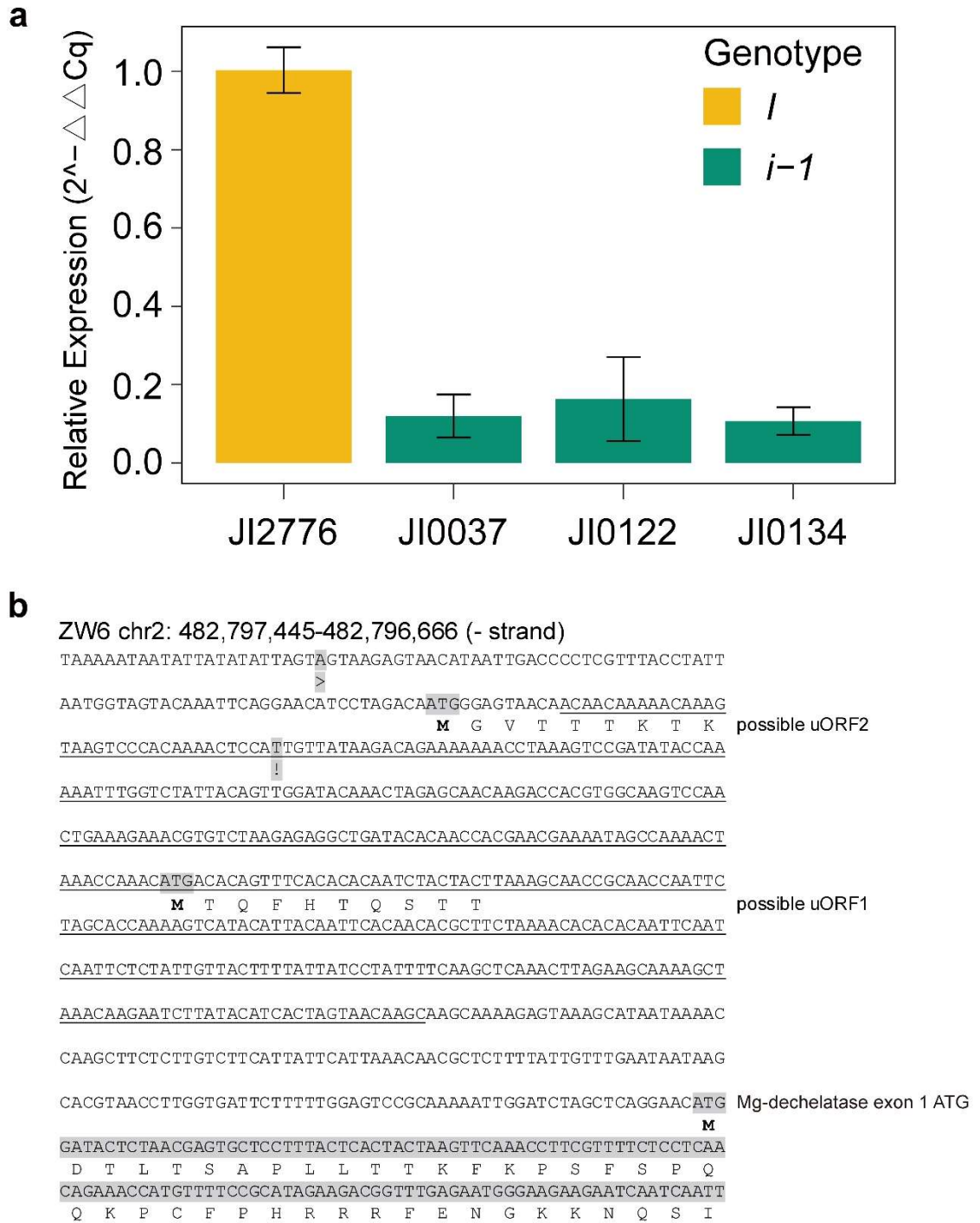


Supplementary Fig. 4 | Photo gallery to illustrate the phenotypic survey of seed characters within the diversity panel. The phenotypic records for the rest of Mendel's pea traits are summarized in Supplementary Table 7. The overall phenotypic dataset is summarized in Supplementary Table 37.



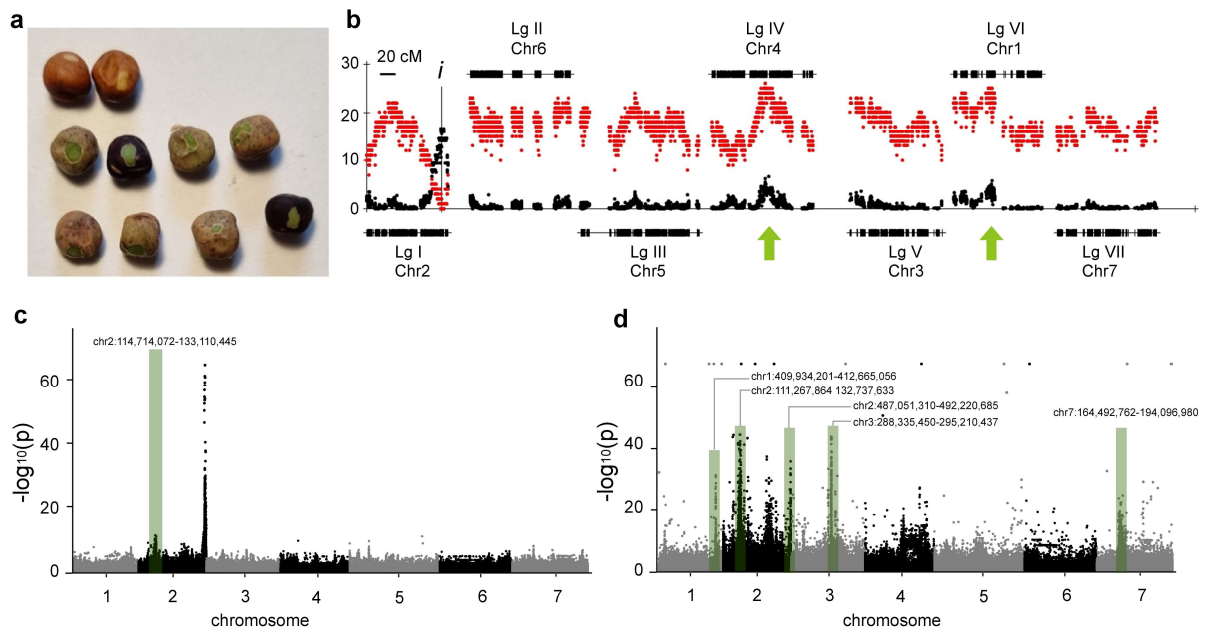
Supplementary Fig. 5 | Haplotype-phenotype association study for cotyledon colour (yellow vs. green). **(a)** Manhattan plot of GWAS using the ZW6 genome assembly as the reference. The strongest signal corresponds to the *I* locus located on the short arm of Chr2. The associated genomic interval is narrowed down into a 1 Mb region surrounding the *Stay Green* gene (*SGR*, *Psat02G0529500*) gene, highlighted in red. The local linkage disequilibrium (LD) map is shown below. The significance value of the data point from the local GWAS corresponding to *I* is low; the causative variation is an indel and the SNPs do not distinguish the *i* allele from its *I* progenitor. **(b)** Population-based haplotype clustering and haplotype-phenotype association analysis of *SGR*. Most accessions with green seeds cluster within Haplotype 9, associated with a *Ty1-copia* insertion event in the last intron between exon 4 and exon 5. Note that this haplotype lacks unique SNPs. Haplotype 2 corresponds to a promoter deletion event that presumably disrupts the expression of *SGR*. Some of the other accessions with green seeds distributed elsewhere are possibly caused by other genes or by premature

maturation of the seeds. The TAR (*Ty1-copia* element) is marked in yellow both within the gene structure (top) and in the haplotype heatmap (in the middle of Hap9). **(c)** *SGR* gene structures identified in this study and previously³⁵. In this study only two allelic forms of *i* were found; *i-1* in Haplotype 9 and *i-2* in Haplotype 2.



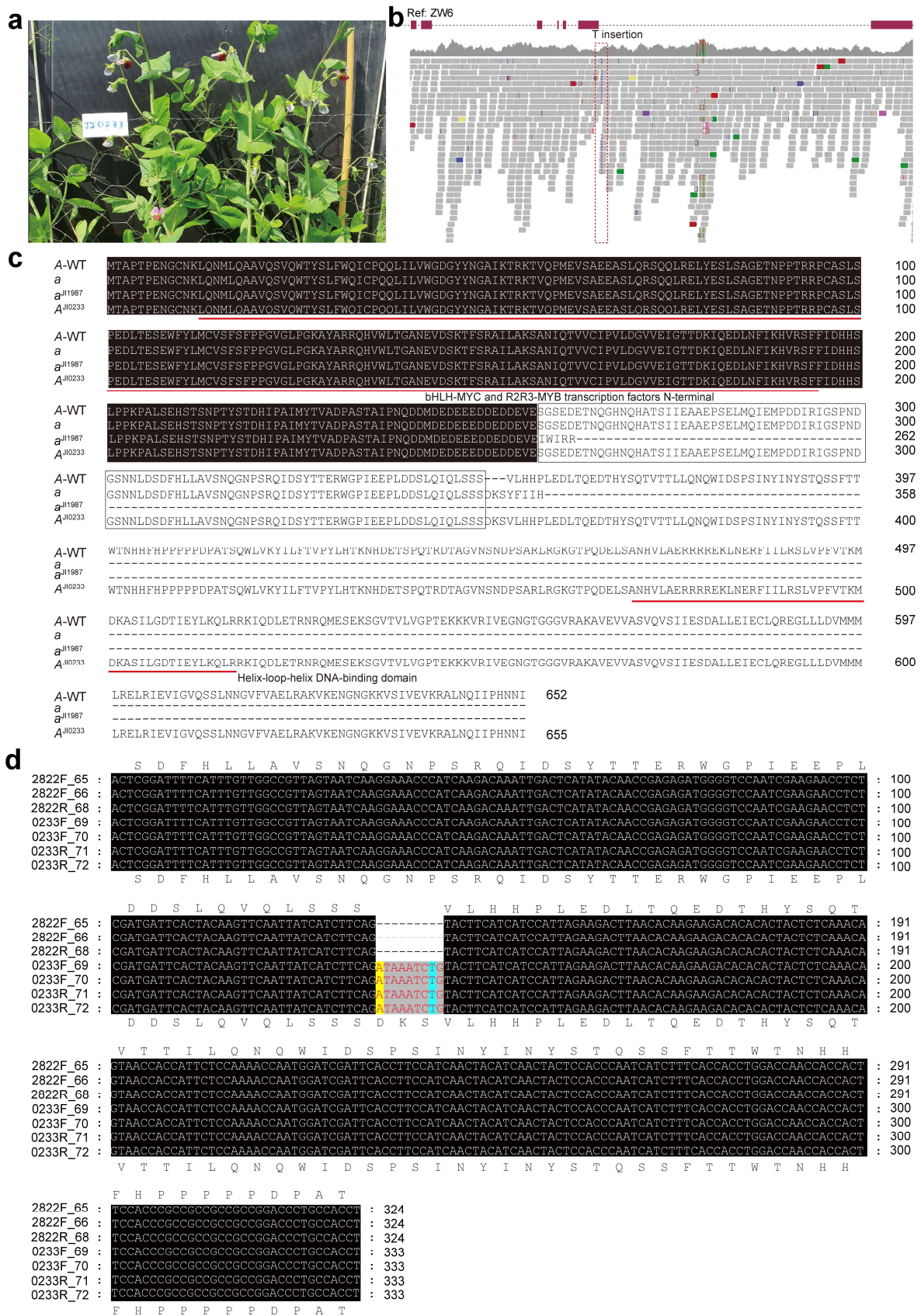
Supplementary Fig. 6 | Transcriptional effects of the *i-1* allele and *i-2* alleles. (a) qRT_PCR analysis comparing the expression levels of the *Stay Green* (*SGR*, Mg-dechelata) gene between the wild-type (*I*) and *i-1* allele carriers. The analysis includes accessions JI2776 (*I*), JI0037 (*i-1*), JI0122 (*i-1*), and JI0134 (*i-1*). Three biological replicates for each sample (cotyledons, i.e seeds without testas) were obtained 15 days post flowering. The results highlight reduced transcriptional activity associated with the *i-1* allele. (b) Illustration of the

408bp deletion upstream of the Mg-dechelatase sequence, as revealed in the *i-2* allele in this study. The region immediately upstream of *SGR*, *Psat02G0529500* of ZW6 is presented. The 408 bp deleted in haplotype (Hap) 2 (*i-2* allele as presented in Supplementary Fig. 5c) is underlined. Two possible regulatory upstream open reading frames are indicated (these are in the same frame as exon 1). The proposed start of the 5' UTR in the Caméor annotation¹⁷ is at the A marked '>' immediately after a potential TATA box. In the same assembly the most 5' nucleotide in a transcript is at the T marked '!'.



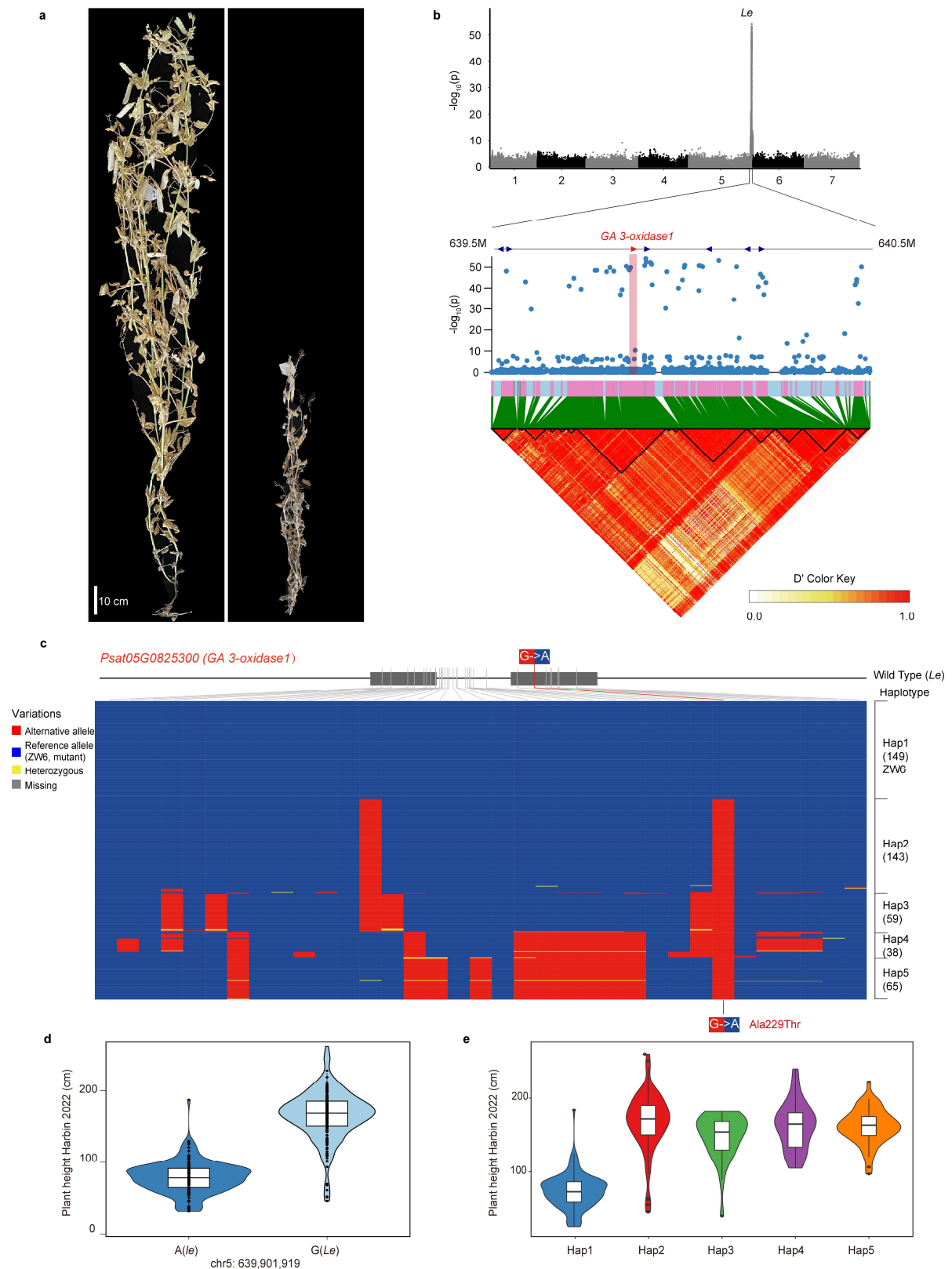
Supplementary Fig. 7 | Phenotypic, genetic, and GWAS analysis of accessions with unexpected cotyledon colours. (a) Ten F₃ seeds derived from the selfing of an *i-l/i-l* F₂ individual, showing two seeds with yellow cotyledons and eight with green cotyledons; (b) Investigation of the allele scores for 33 of the 88 F₂ *i-l/i-l* homozygotes F₂ plants from the cross JI2822 x JI0816. These 33 F₂ plants produced F₃ seeds, some of which were yellow, contrary to the expected green phenotype. If these F₂s were heterozygous at some other locus affecting seed colour, then this may account for the appearance of yellow F₃ seeds where only green seeds were expected. The plot shows the number of heterozygous scores for each marker (red) and the corresponding χ^2 value (black). The x-axis is the genetic map position of the marker. Above and below the graph regions of identity by descent are shown as thin (rather than thick) horizontal black lines. The position of *i* is marked and a 20cM scale is shown. The green arrows indicate regions of the map discussed in the main text; (c) GWAS analysis only using accessions of *Pisum sativum*; (d) GWAS analysis only using accessions carrying the *II* alleles (removing accessions carrying the *i-l* and *i-2* alleles). GWAS signals of interest are highlighted with green shading.

reference, showing a single strong signal which is consistent with the *A* locus located on the short arm of Chr6. **(b)** Manhattan plot of the GWAS with respect to the Caméor v1a genome reference, showing a second signal (marked with a red star) found in the middle of Chr6 which, with reference to the genetic map⁵⁰, was attributed to an assembly error. **(c)** Local detail of the genomic interval, showing the gene list with the basic Helix-Loop-Helix (*bHLH* transcription factor, *Psat06G0169800*) highlighted in red. The local linkage disequilibrium map is shown below. **(d)** Population-based haplotype clustering analysis of the *bHLH* gene. The upper panel represents the whole gene, and the lower panel focuses on functionally relevant SNPs. Five haplotypes are identified. Most of the purple flowered lines (as shown in the left bar) belong to Hap1 carrying the G of the wild-type intron 6 splice donor site⁴⁸. Haplotypes 2-4 are mutant types with white flowers: The Hap2 allele has an additional A in exon 5, as originally described for JI1987⁴⁸; Hap3 corresponds to the deletion of exons 1 and 2. Hap4 has a deletion of exons 1 to 6. Hap5 (including the ZW6 reference genome) is the most common mutant type with white flowers and carries the G to A substitution at the intron 6 splice donor site first described in Caméor⁴⁸. However, within Hap5, there is one exception (JI0233) which carries this G to A substitution but has fully pigmented flowers (see main text). It is worth mentioning that the coding sequence (the *bHLH* transcript) is on the – strand in the reference genome (ZW6); here we standardize the comparison between the wild-type and the mutant types. **(e)** Gene structures corresponding to the variants described in panel **d**. **(f)** Frequency distribution of phenotypes (pigmented/purple vs. white) for each of the haplotypes. **(g)** Read mapping to confirm the exon deletion events that disrupt gene function conferring white flowers of Hap3 and Hap4 as shown in panel **d/e**.



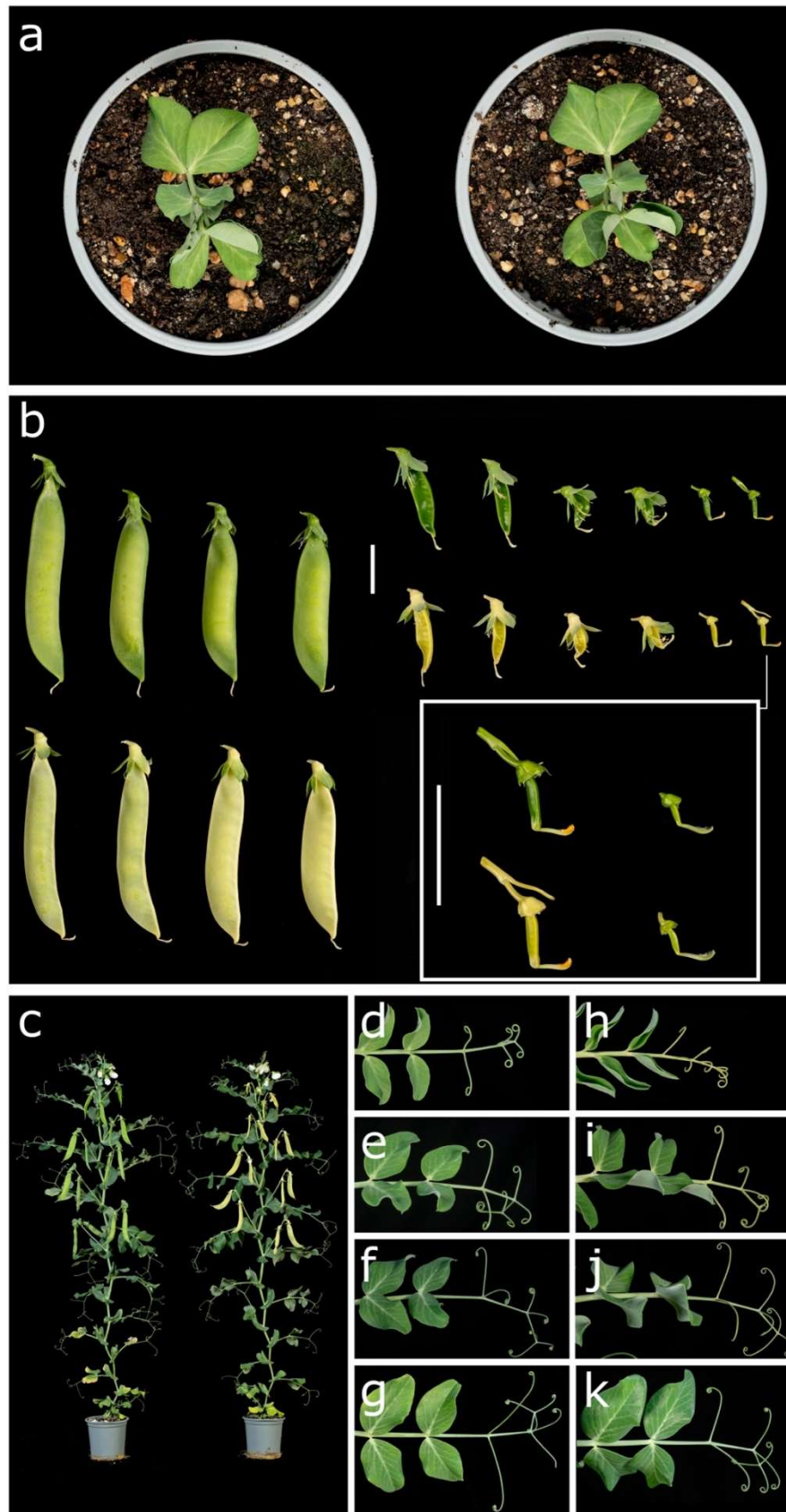
Supplementary Fig. 9 | The intragenic suppressor mutation, A^{J10233}. (a) The phenotype of a JI0233 mature plant, clearly showing the purple flower. **(b)** Read mapping shows the insertion

of a nucleotide (T in the coding strand) in JI0233 with respect to the Caméor *a* allele. (c) Alignment of the predicted amino acid sequences of the wild type *A* allele (JI2822, *A*-allele) and the *a* alleles of Caméor (*a*) and JI1987 (*a*^{JI1987}) compared to the *a*-derived *A* allele of JI0233 (*A*^{JI2033}) (**Supplementary Fig. 8e**); *A*^{JI0233} has an insertion of an additional nucleotide (T in the coding strand) in what would be in the sixth intron of the *bHLH* transcription factor corresponding to the wild type *A* allele. With respect to the progenitor *a* allele this results in restoration of the wild-type reading frame in JI0233, adding nine nucleotides to the transcript as described in the main text. (d) cDNA sequence of the transcript from JI2822 and JI0233 showing the most common G to A SNP in the splice donor site (yellow-shaded) and the T insertion (blue-shaded) in JI0233, which restores the wild type reading frame. This confirmed that the *A*^{JI0233} allele uses the same splice donor site as in the *a* allele of Caméor.



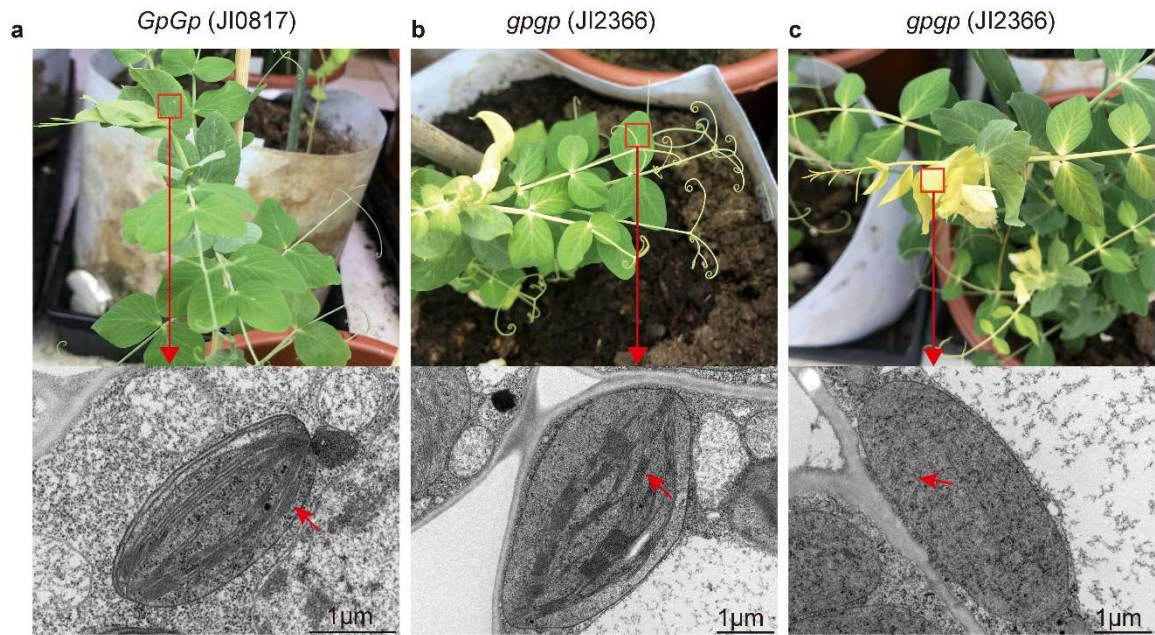
Supplementary Fig. 10 | Haplotype-phenotype association study for plant height (tall vs. dwarf). (a) Representative images showing the contrasting traits of stem length (tall vs. dwarf). **(b)** Manhattan plot of GWAS, identifying a single strong signal corresponding to the *Le* locus

located on the short arm of Chr5, based on the ZW6 genome reference. Local detail of a 0.5 Mb region within this genomic interval including the GA 3-oxidase1 gene^{55,56}, *Psat05G0825300*, highlighted in red. The local linkage disequilibrium map is shown below. **(c)** A population-based haplotype clustering analysis of the GA 3-oxidase1 gene, showing five different haplotypes. Most accessions with short stem length are clustered into Hap1 carrying the previously described G to A mutation^{55,56}; all the other haplotypes (Hap2-5) have the G nucleotide. The reference genome (ZW6) belongs to the mutant type with the Ala229Thr substitution (nucleotide G to A). **(d)** Distribution of the phenotypes (plant height, Harbin location) corresponding to accessions carrying the mutant (A, *le*) or the wild type (G, *Le*) allele; **(e)** Distribution of the phenotypes (plant height, Harbin location) corresponding to the different haplotypes (Hap1-5).

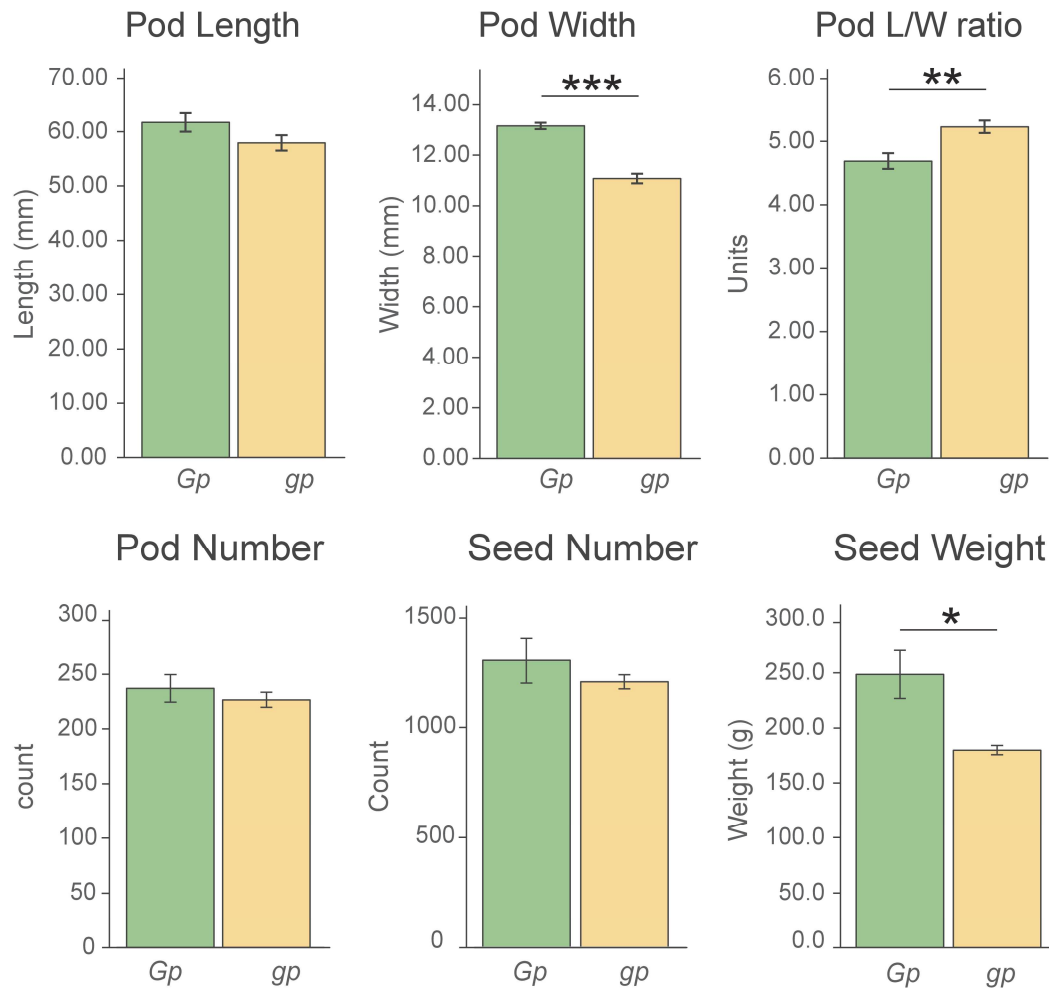


Supplementary Fig. 11 | Plant phenotype of the *gp* mutant. (a) Phenotype of seedlings 15 days after sowing (DAS), BC6 *Gp* (left) and BC6 *gp* (right). At the early vegetative stage of

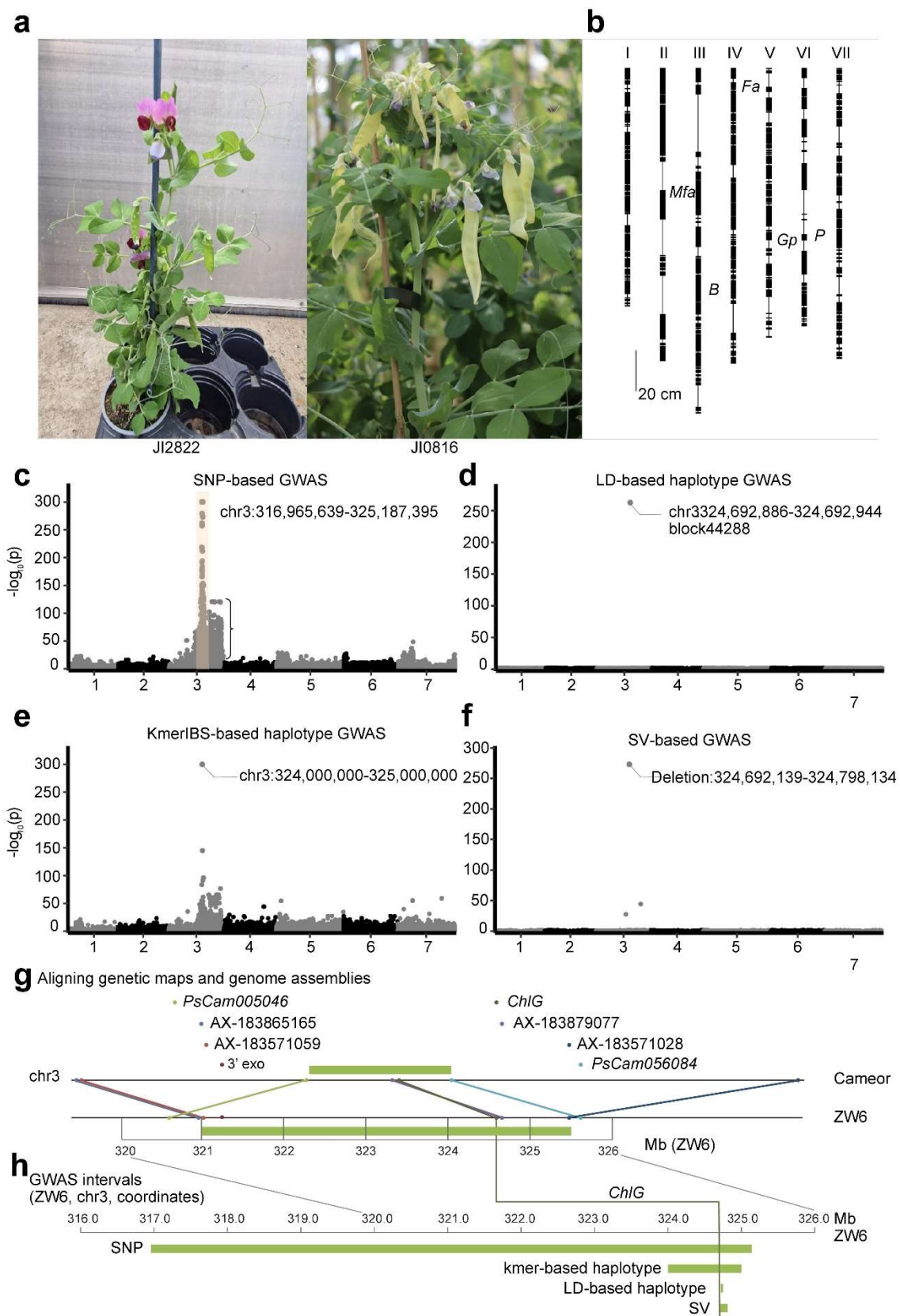
development, *gp* mutant and wild type seedlings are indistinguishable. **(b)** Phenotype of BC6 *Gp* pods (upper) and BC6 *gp* pods (lower) showing pod colour at successive stages of development. Stages are 13 days post anthesis (DPA), 11 DPA, 9 DPA, 7 DPA, 5 DPA, 4DPA, 3DPA, 2DPA, 1DPA, 0DPA from left to right. Inset shows a magnified view of *Gp* carpels (upper) and BC6 *gp* carpels (lower), post-fertilisation (left) and pre-fertilisation (right). At all stages shown, *gp* mutant carpels and pods are distinguishable from wild type by their yellow colour. **(c)** Phenotype of mature plants 72 DAS, BC6 *Gp* (left) and BC6 *gp* (right). Stage 72 DAS is the middle frame of a 7-frame time-lapse series ([7frames_BC6Gp3gp21.gif](#)), showing the development of these plants at 29 DAS, 44 DAS, 58 DAS, 72 DAS, 86 DAS, 105 DAS and 177 DAS. **(d – k)** Uppermost four leaves (each one including the rachis, two pairs of leaflets, three pairs of tendrils, the terminal tendril, and excluding stipules) on BC6 *Gp* (**d – g**), and BC6 *gp* (**h – k**) plants shown in **c**. **(d, h)** Uppermost upright leaves; **(e, i)** Fully unfolded leaves at node below uppermost; **(f, j)** Fully expanded leaves located two nodes below uppermost; **(g, k)** Fully expanded leaves located three nodes below uppermost. Leaves at the two uppermost nodes (and leaf primordia, not shown) of the BC6 *gp* mutant have yellow tendrils, rachis and leaflet midribs and petiolules compared to more mature, expanded leaves at lower nodes, which are fully green. In contrast, leaves at all nodes of BC6 *Gp* wild type plants are green. Pots in panels **a** and **c** are 9 cm in diameter, scale bars in panel **b** are 2 cm.



Supplementary Fig. 12 | Comparison of transmission electron microscopic observations (HT7800/HT7700, HITACHI) of the normal or abnormal thylakoid structure (short red arrows) between JI2366 (*gp*) mature green leaves (**b**), JI2366 (*gp*) young yellow leaves (**c**), and JI0817 (*Gp*) green leaves (**a**). Regions of leaf samples analysed are highlighted by red boxes and long arrows.

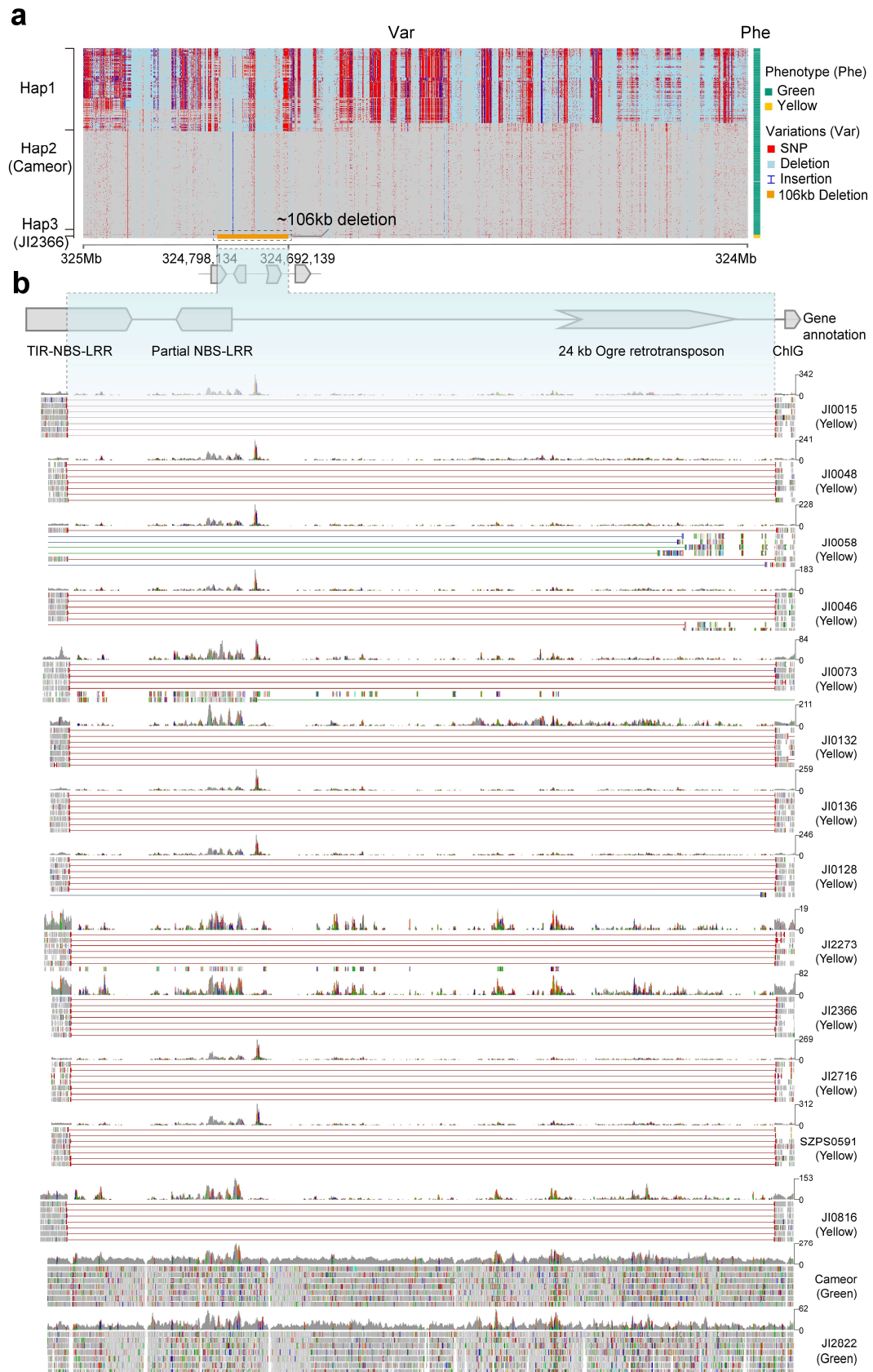


Supplementary Fig. 13 | *Gp* vs *gp* phenotypes. Pod length (t test, $P = 0.096$, $n = 25$), pod width (t test, $P < 0.0001^{***}$, $n = 25$), ratio of pod length to width (t test, $P = 0.0017^{**}$, $n = 25$), pod number /plot (t test, $P = 0.5124$, $n = 3$), seed number/plot (t test, $P = 0.418$, $n = 3$) and seed weight/plot (t test, $P = 0.036^{*}$, $n = 3$) of BC6 S3 *GpGp* and *gpgp* genotypes derived from the Caméor x JI0015 cross grown in 1 m x 1 m field plots.



Supplementary Fig. 14 | Identification of the position of *Gp*. (a) The phenotype of JI2822 (*Gp*) and JI0816 (*gp*) showing different pod colour. (b) The genetic markers including those

associated with *Mfa*, *B*, *Fa*, *Gp* and *P* locus are shown on the linkage groups I - VII. **(c)** SNP-based GWAS. **(d)** LD-based haplotype GWAS. **(e)** kmer IBS-based haplotype GWAS. **(f)** SV-based GWAS. A significant single signal, corresponding to a deletion variant located at Chr3: 324,692,139-324,798,134 has been identified in these different approaches. **(g)** Based on JI0015 x JI0399 RILs and the two RIL cross (Supplementary Table 21), genetic markers flanking *Gp* are positioned with respect to each other on a 10 Mb region of the assemblies of two pea genomes: Caméor v1a (Kreplak et al. 2019)¹⁷, and ZW6 (Yang et al. 2022)¹⁸. In addition to the genetic markers discussed above, the positions of chlorophyll synthase *ChlG* and the gene encoding a 3' exonuclease identified by Shirasawa et al. (2021)²² that lies within this region are indicated. **(h)** The genomic intervals identified by different approaches of GWAS; the location where *ChlG* is situated is highlighted.



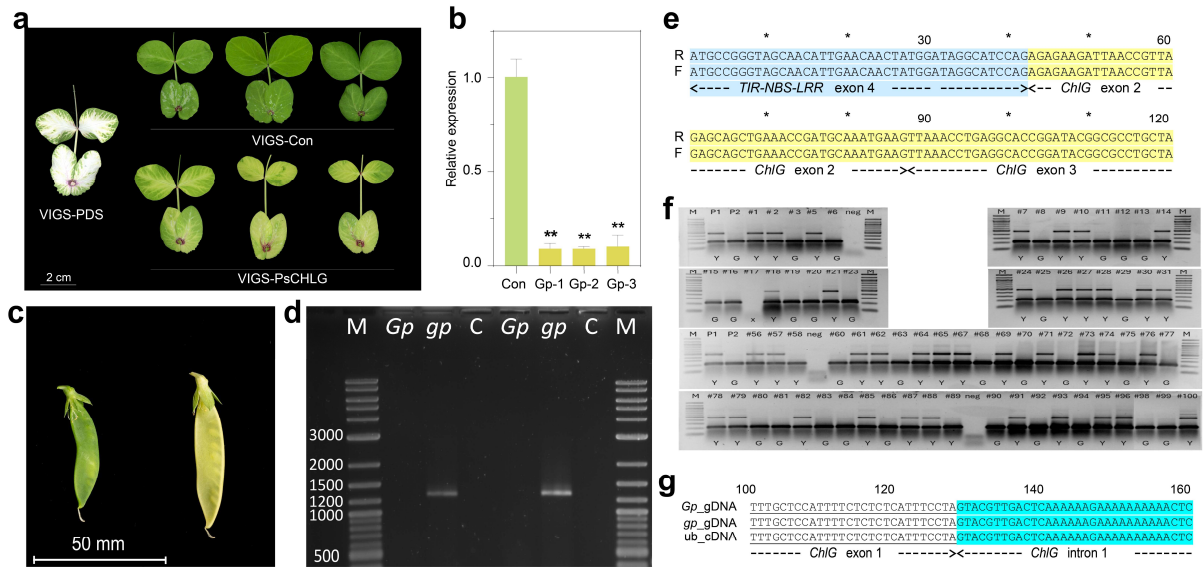
Supplementary Fig. 15 | (a) Haplotype clustering and haplotype-phenotype association analysis near the large deletion region across the entire diversity panel in this study. In total,

three major haplotypes were identified (Hap1, Hap2, Hap3) with respect to the ZW6 genome reference, in which the green podded-line Caméor (*Gp*) belongs to Hap2 and the yellow-podded mutant lines JI2366/JI0015 (*gp*) belong to Hap3. All the *gp* lines carrying the ca. 100kb deletion (marked in orange, coordinate: 324,692,139 – 324,798,134 at Chr3) are clustered into Hap3, affecting one ~24Kb *Ogre* LTR retrotransposon, one ORF encoding a partial *NBS-LRR* gene, and part of exon 5 and the whole of exons 6 and 7 from a gene encoding a TIR-NBS-LRR (NLR, Past03G0414100) protein. The distance between the deletion boundary and the *ChlG* coding sequence is approximately 3.7 kb (Figure 3d). Note this map is in the opposite orientation to that shown in Fig. 3. **(b)** Top row the exon structure and position of the genes within or adjacent to the deletion. IGV display of short-read mapping against the ZW6 reference genome, showing the ca. 100 kb deletion event, found in all 12 yellow podded lines (*gp**gp*), but in no *GpGp* lines within our panel. (*Gp*, Caméor and JI2822 for comparison at the bottom).

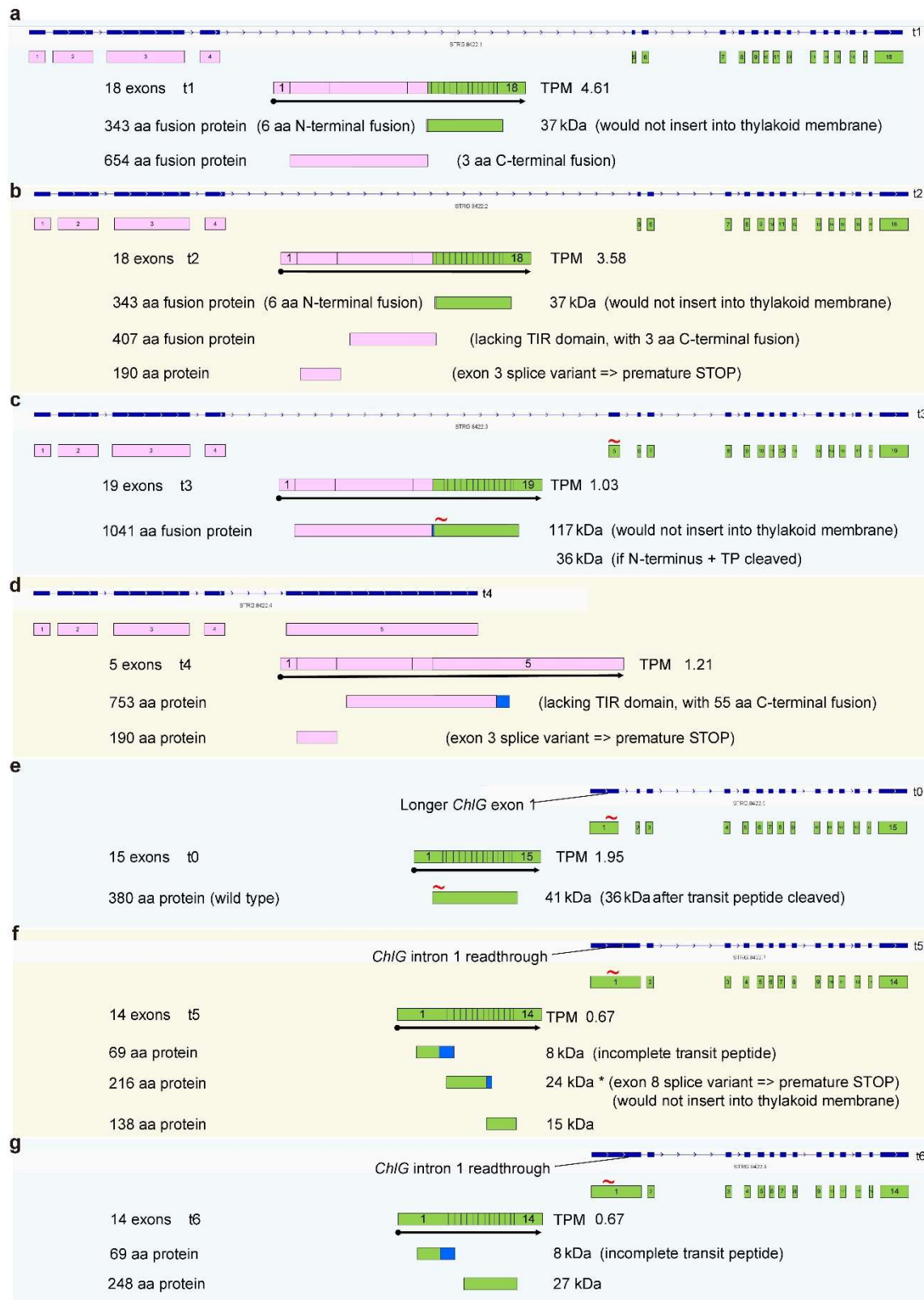


Supplementary Fig. 16 | Details of the aberrant transcripts in *gp* pods in comparison with the normal transcripts in the wild-type lines from the *NLR-ChIG* region. (a) Predicted transcripts from wild-type (JI0399) green pods and three RIL green pod pools, mapped vs JI2822 (*Gp* allele), in a 150 kb region encompassing the *TIR-NBS-LRR* gene (*Psat03G0414100*) and *ChlG* (*Psat03G0413700*). An *Ogre* retrotransposon with LTRs represented by black arrows, and an expressed captured sequence in its 3' LTR (corresponding to ZW6 chr3:324695299-324696949), lies upstream of *ChlG* as indicated. Data are displayed in IGV with RNA-Seq read coverage and StringTie-predicted transcripts shown. Circled transcripts are annotated according to the ZW6 genome (Yang et al 2022)¹⁸. TPM values of five predicted *TIR-NBS-LRR* (*Psat03G0414100*) transcripts, one predicted *ChlG* transcript and one predicted *Ogre* LTR2 transcript are marked with black numbers. (b) Predicted transcripts from *gp* mutant (JI0015) yellow pods and three RIL yellow pod pools, mapped vs JI0015 (*gp* allele), in a 15 kb

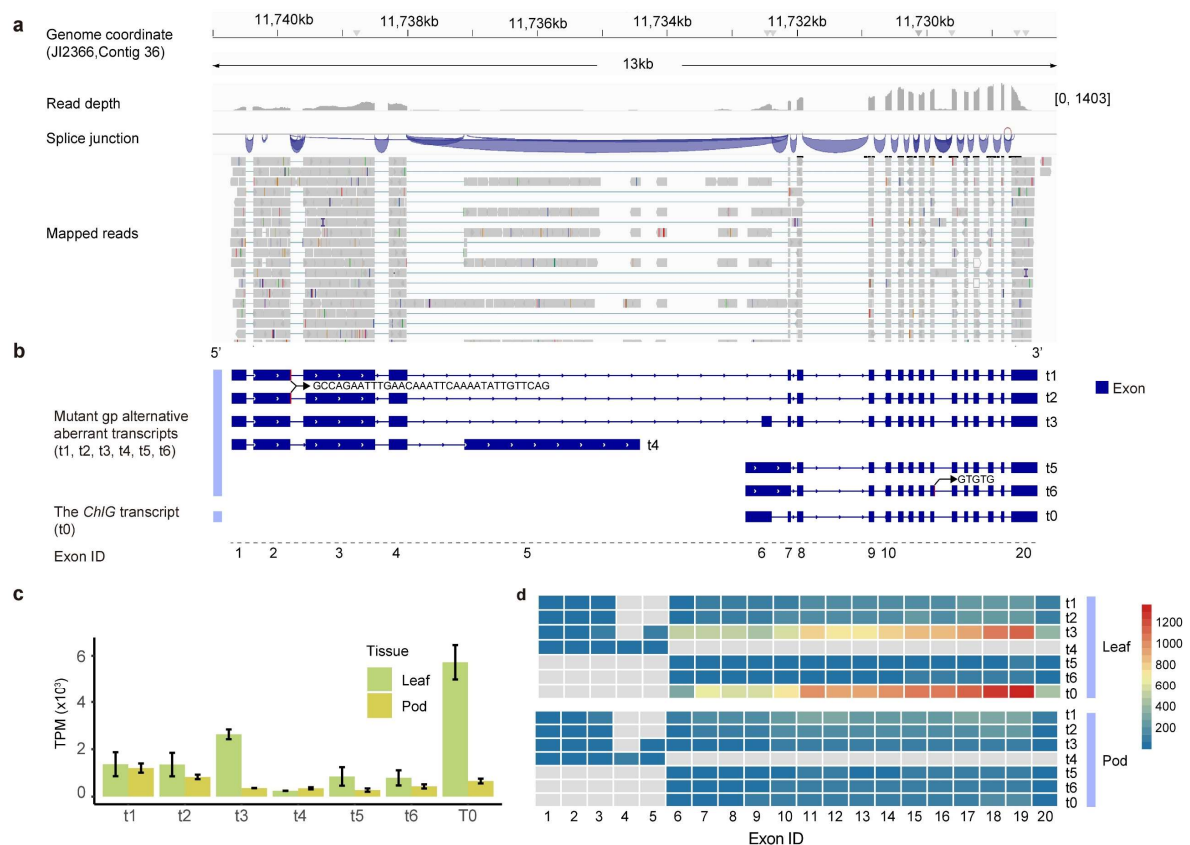
region spanning the truncated *TIR-NBS-LRR* gene (*Psat03G0414100*) and intact *ChlG* (*Psat03G0413700*). The large deletion, which removes the 3' terminal 11 bp of *Psat03G0414100* exon 5 and ca. 100 kb downstream, is represented by a red polygon. The *Ogre* retrotransposon lying upstream of *ChlG* has been deleted, as have three annotated genes (*Psat03G041400*, *Psat03G0413900* and *Psat03G0413800*). Data are displayed in IGV with RNA-Seq read coverage and StringTie-predicted transcripts shown. TPM values of seven predicted *gp* allele-specific transcripts are marked with black numbers. Three predicted aberrant *TIR-NBS-LRR-ChlG* transcripts, splicing *TIR-NBS-LRR* exon 4 to either *ChlG* exon 2 (2 isoforms), or *ChlG* exon 1 (1 isoform), are shown. A predicted transcriptional fusion of *TIR-NBS-LRR* exon 5 with deletion-adjacent non-coding sequence to create a novel predicted translational fusion product (1 isoform) is shown; however this is not predicted in the three yellow pod pools. Predicted aberrant intron 1 read-through *ChlG* transcripts (2 isoforms) with extended 5' UTRs are shown. One conventionally spliced *ChlG* isoform with extended 5' UTR is shown.



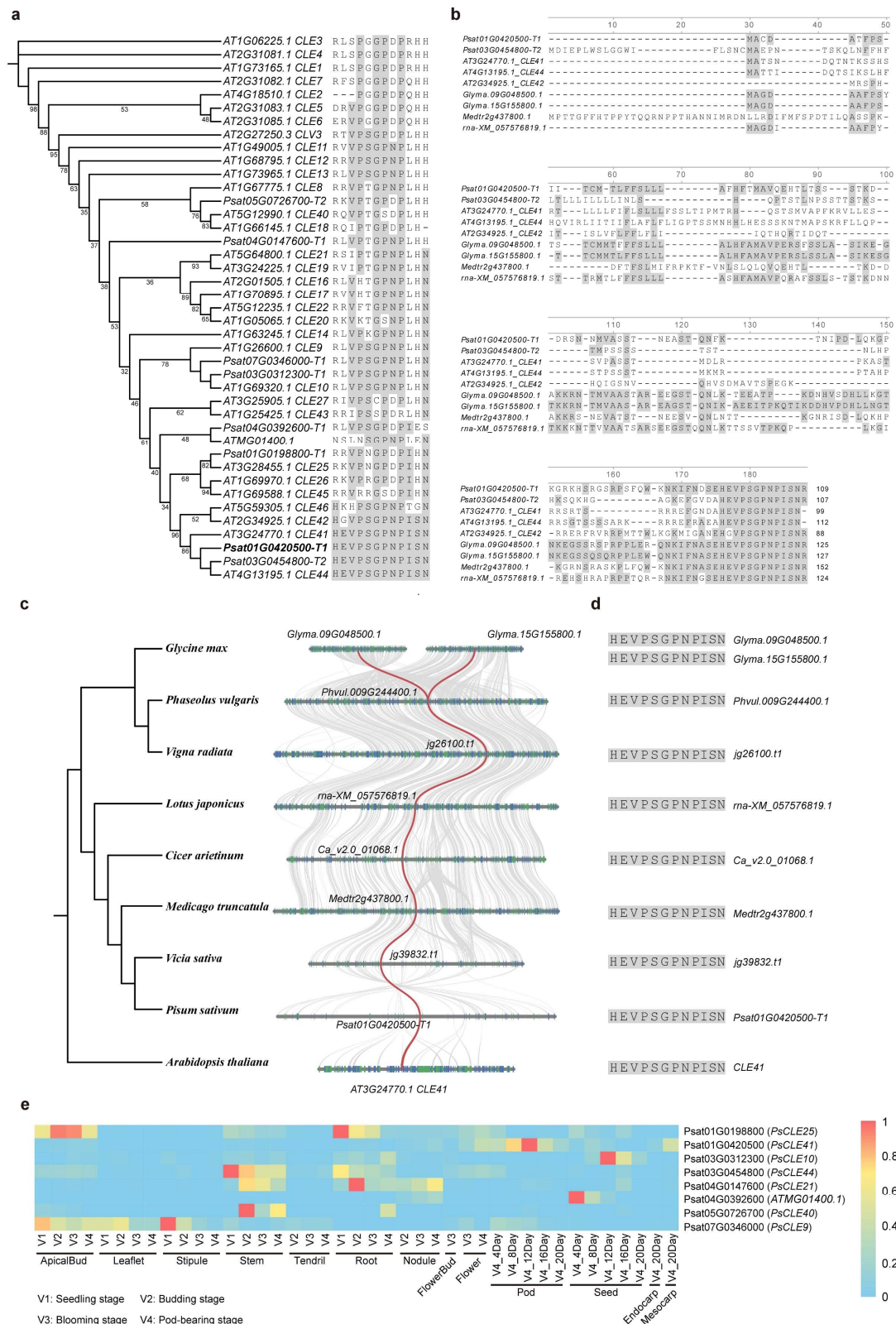
Supplementary Fig. 17 | Gene and transcript validation experiment. (a) Silencing of *PsCHLG* by VIGS, with VIGS-con and VIGS-PDS serving as negative and positive control, respectively, after 20 days of infection using *Agrobacterium* with different VIGS vectors (scale bar = 2 cm). (b) Measurement of expression levels of *PsCHLG* in the leaves of VIGS assayed plants after 20 days of infection using *Agrobacterium* with VIGS vector (** indicate a significant level at $P < 0.01$ in a Students' t-test). (c) Examples of BC6 *Gp* (left) and *gp* (right) pods 5 days after anthesis. Scale bar is 50 mm. (d) RT-PCR amplification of predicted aberrant transcripts from BC6 *Gp* and *gp* pods 5 days post anthesis, in duplicate, with water negative controls (C). Primers TIRexon3F and ChlGexon4R were located in *Psat03G0414100 TIR-NBS-LRR* exon 3 and *Psat03G0413700 ChlG* exon 4 (Supplementary Table 47). Sizes of markers (M) shown on left. (e) Sanger sequenced 396 bp PCR products re-amplified from samples in (d) using primers DISRES25994F (F) and ChlGR15 (R) showing *TIR-NBS-LRR* exon 4 spliced to *ChlG* exon 2. (f) RT-PCR amplification of pod cDNA in the JI0015 (*gp*) x JI0399 (*Gp*) RIL population. A 210 bp PCR product amplified by primers ChlGF18 and ChlGR7 (Supplementary Table 47), corresponding to conventionally-spliced *ChlG* exons 1 and 2, is seen in all RILs. A 459 bp PCR product, corresponding to an aberrant intron 1 read-through transcript (upper band), occurs only in yellow podded lines of the population. RILs (#1, #2, #3 etc.) with green pods (G) and yellow pods (Y) are indicated. The last three RILs (#98, #99 and #100) were run on a separate gel. M, 100 bp ladder lanes; P1, JI0015 parent 1 of the RIL population; P2, JI0399 parent 2 of the RIL population; x, reaction failed. (g) Sanger sequenced isolated 459 bp upper band (ub) from JI0015 yellow pod cDNA in (f) aligned with JI0015 and JI2822 genomic DNA (gDNA). *ChlG* intron 1, boxed in green, is unspliced in the isolated JI0015 yellow pod cDNA upper band.



Supplementary Fig. 18 | Predicted protein sequences translated from the different transcript isoforms in the *NLR-ChlG* region.

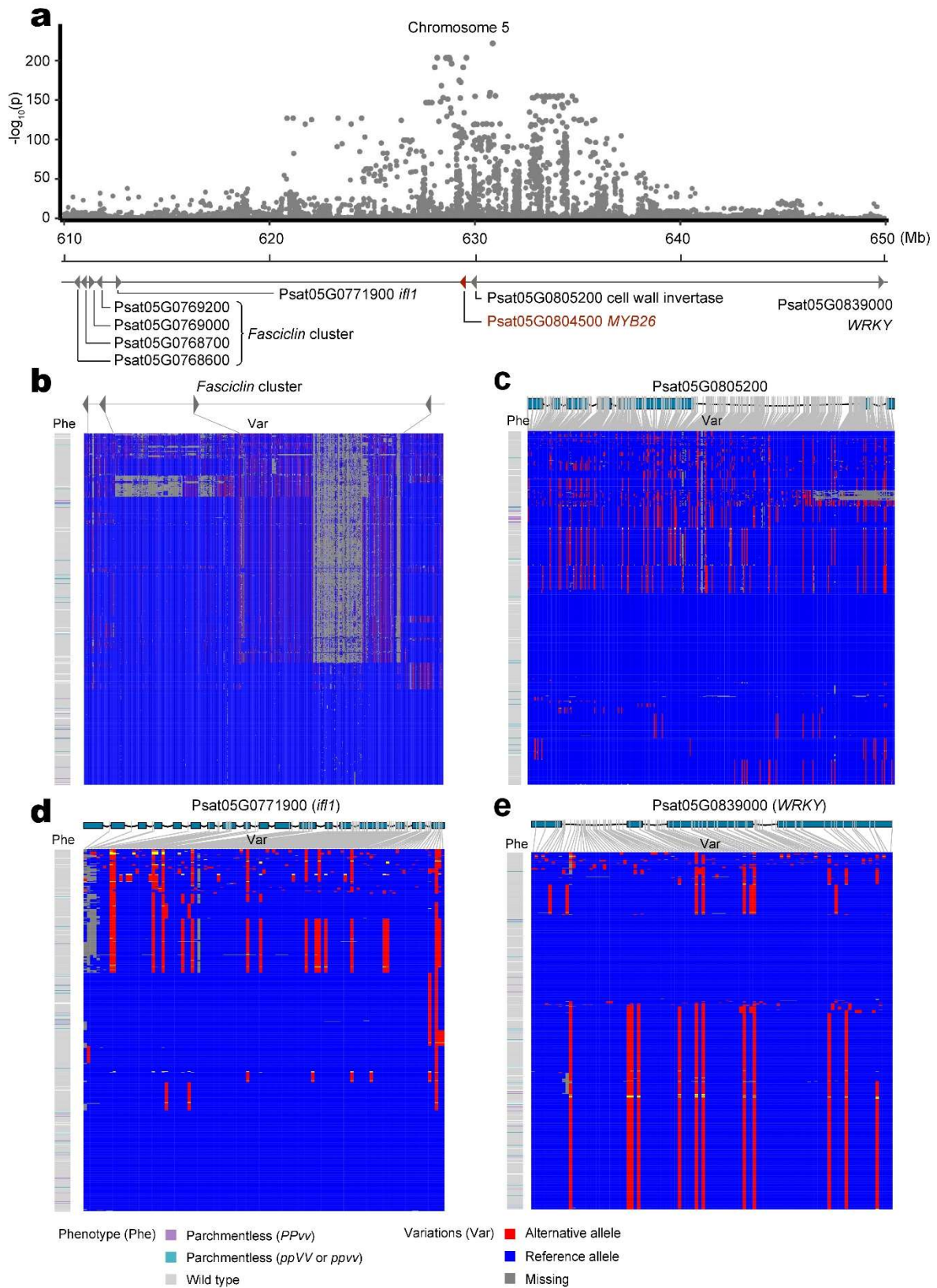


Supplementary Fig. 19 | The gene model and transcript annotation for JI2366 (*gp*) in the *NLR-ChlG* region near the ca. 100 kb deletion, and relative gene expression of different transcripts in leaf and pod. (a) Integrated Genomics Viewer (IGV) snapshot showing RNA reads aligned to the JI2366 (*gp*) genome, including genome coordinates, read depth, splice junctions, and mapped reads. **(b)** Illustrates the structures of 4 different transcripts along with exon numbering. **(c)** Quantifies the expression levels of these transcripts in *gp* leaf and pod tissues. **(d)** Provides a detailed visualization of read depth by RNA-seq short-read mapping across each exon for both *gp* tissue types.



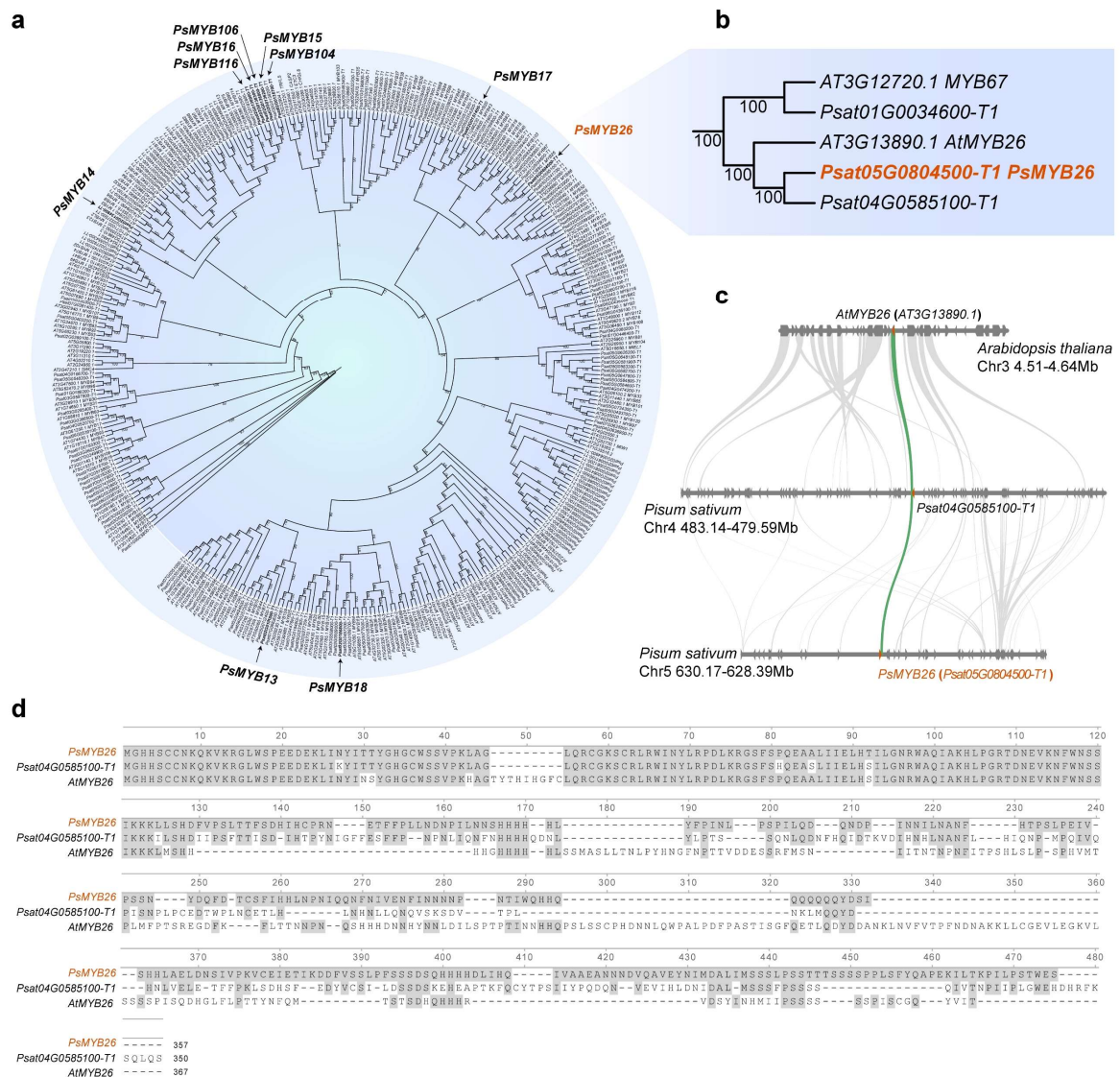
Supplementary Fig. 20 | Orthologue inference and characterization of *PsCLE41* (*Psat01G0420500*). (a) Phylogenetic analysis of homologous genes from *Arabidopsis thaliana* and *Pisum sativum*. The analysis reveals that *Psat01G0420500* (designated here as *PsCLE41*)

is most closely related to *Arabidopsis AtCLE41*, and *Psat03G0454800-T2* (designated here as *PsCLE44*) is close to *Arabidopsis AtCLE44*. While the non-TDIF sequence evolved very quickly, the mature 12-amino acid signalling peptide sequences deduced from *PsCLE41*, *AtCLE41* and *AtCLE44* ('HVPSPGNPISN', in the right panel of the tree) are 100% identical to each other. Note that the TDIF peptide of *AtCLE42* peptide ('HGVPSPGNPISN') has a mutation in its second amino acid. The phylogenetic tree was constructed using iqtree2 based on their conserved C-terminal sequences which are the most informative sites for the *CLE* gene families, and the MFP mode and bootstrap support was calculated using the parameter, -bb 1000; **(b)** Multiple sequence alignment of full-length amino acid sequences from *Psat01G0420500-T1* (*PsCLE41*), *Psat03G0454800-T2* (*PsCLE44*), *AT3G24770.1* (*AtCLE41*), *AT4G13195.1* (*AtCLE44*), *AT2G34925.1* (*AtCLE42*), *Glyma.09G048500.1* and *Glyma.15G155800.1* (*CLE41* orthologues from soybean), *Medtr2g437800.1* (*CLE41* orthologue from Medicago), and *rna-XM_057576819.1* (*CLE41* orthologue from Lotus), performed using MAFFT, showing the conserved C-terminal sequence and the TDIF motif. The N-terminal encoding region of the gene is quite dynamic. **(c)** Synteny analysis of the genomic region containing *CLE41* orthologues across different legume genomes and *Arabidopsis*, showing conserved collinear blocks between species; **(d)** Comparison of the TDIF 12 amino acids of *CLE41* orthologues from panel c, showing identical peptide sequences across different legume genomes and *Arabidopsis*; **(e)** Expression profiles of the identified 8 *CLE* peptides in pea across different developmental stages and organs of Caméor.



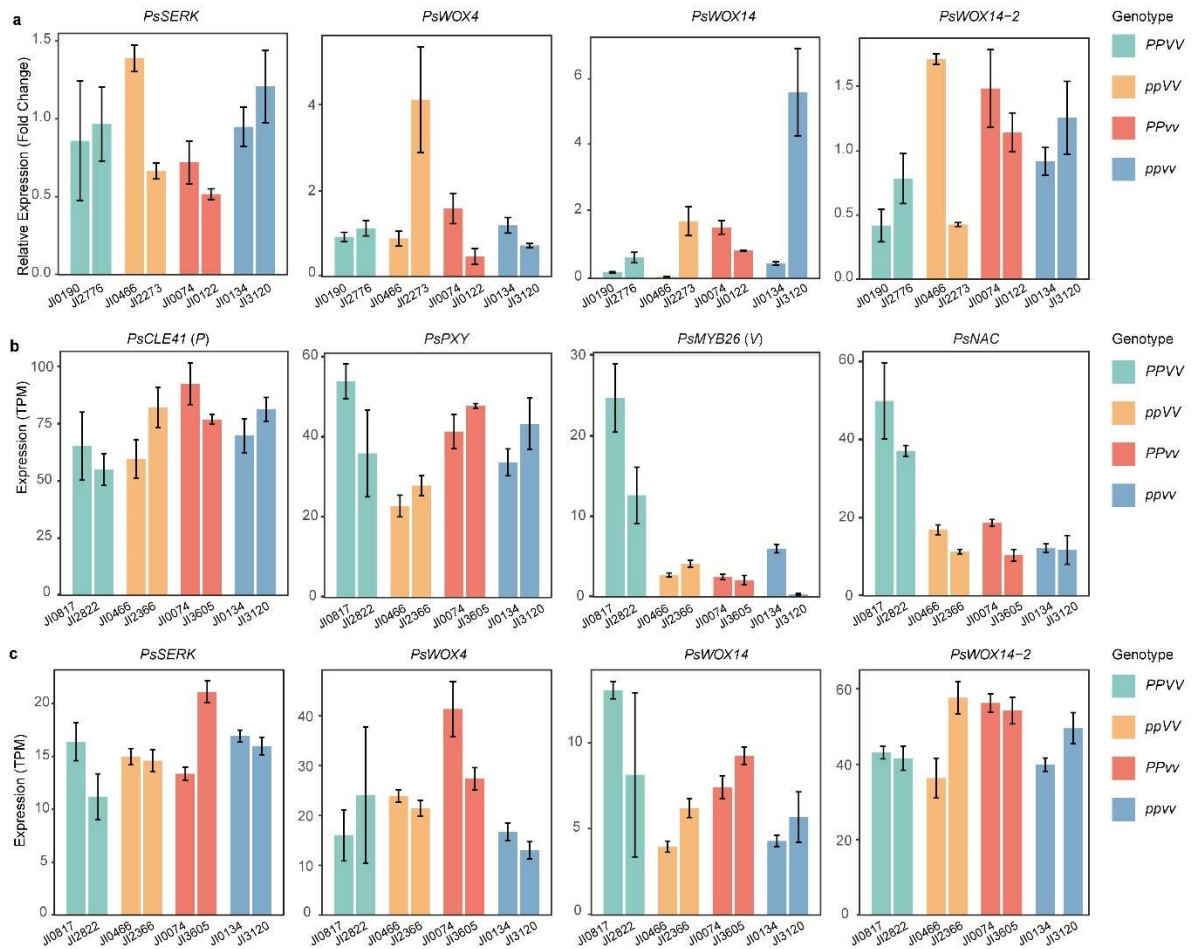
Supplementary Fig. 21 | Haplotype analysis of proposed alternative candidate genes for *V*. (a) GWAS local Manhattan plot showing a strong signal corresponding to *V*. The genes located within this region are shown in the bottom and *Psat05G0804500*, a *MYB26*-like

transcription factor (coloured in orange) was proposed to be the candidate gene of *V*. Haplotype clustering and haplotype-phenotype association analysis for: **(b)** The entire fasciclin gene cluster. **(c)** Gene *Psat05G0805200* annotated as cell wall invertase, ~650kb distance away from the *MYB26* gene. **(d)** *Psat05G0771900* (*ifl1*) and **(e)** *Psat05G083900* (*WRKY*). The phenotypes and variations are colour coded, as indicated below the figure.

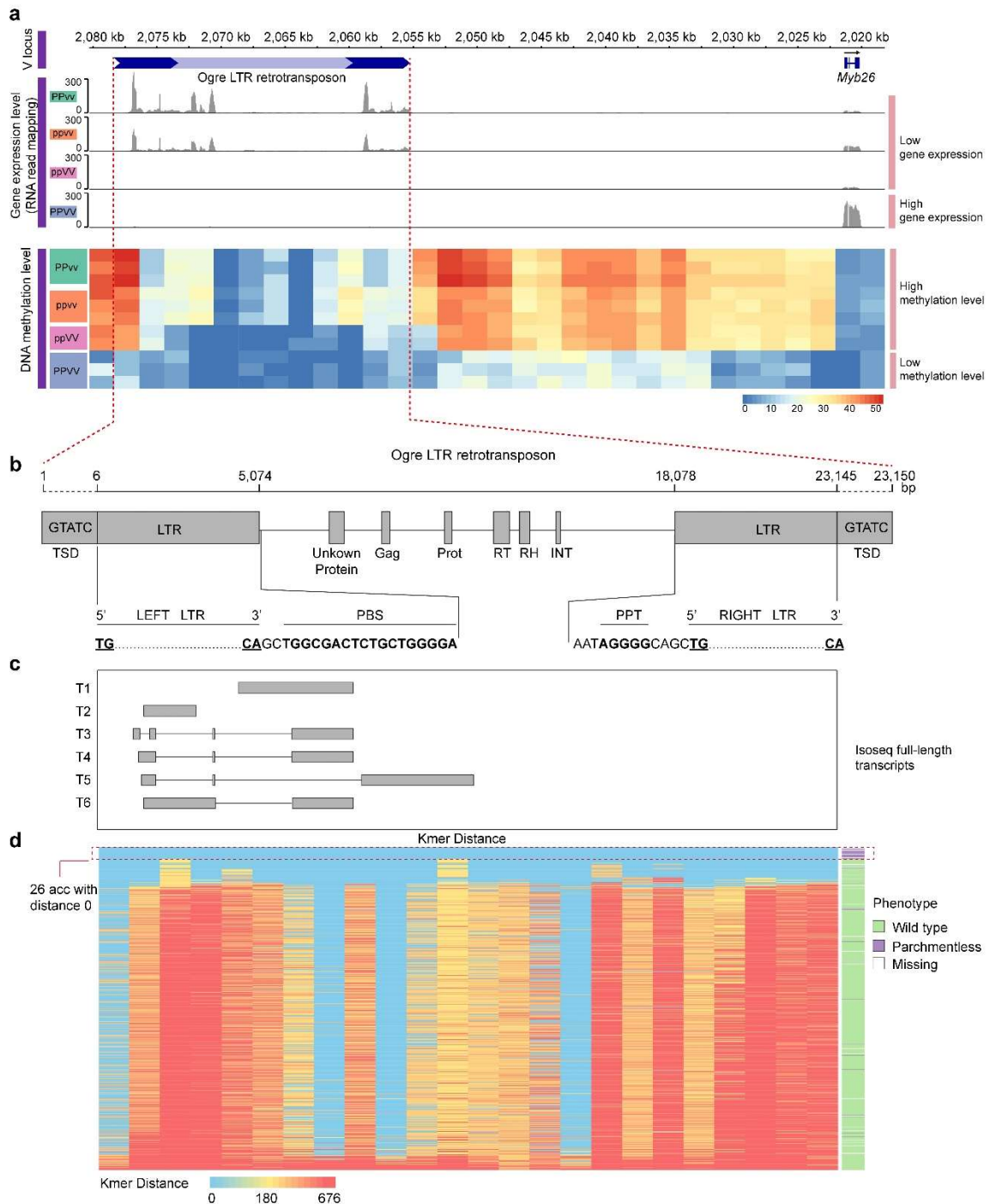


Supplementary Fig. 22 | Analysis of MYB-like gene families associated with *Psat05G0804500* (*PsMYB26*). (a) The phylogenetic tree reconstruction indicates that the *Arabidopsis* MYB26 (*AtMYB26*) is the closest orthologue to pea *Psat05G0804500* (designated as *PsMYB26* in this study); protein sequences were aligned using MAFFT, and the tree was constructed with IQ-TREE2 employing the VT+G4 model. Bootstrap support was calculated using the parameter -bb 1000. (b) Two genes (*Psat05G0804500* and *Psat04G0585100-T1*) were identified as duplication to the orthologues of *AtMYB26*, whereas the gene expression pattern is distinct: only *PsMYB26* is uniquely expressed in pod (Extended Data Fig. 3h). (c) Syntenic and collinear blocks confirm the paralogous (*Psat05G0804500* and *Psat04G0585100-T1*) and orthologous (*AtMYB26*) relationship between *Arabidopsis* and pea.

(d) Full-length protein sequence alignment between *Psat05G0804500* (*PsMYB26*), *Psat04G0585100-T1* and *AtMYB26*.

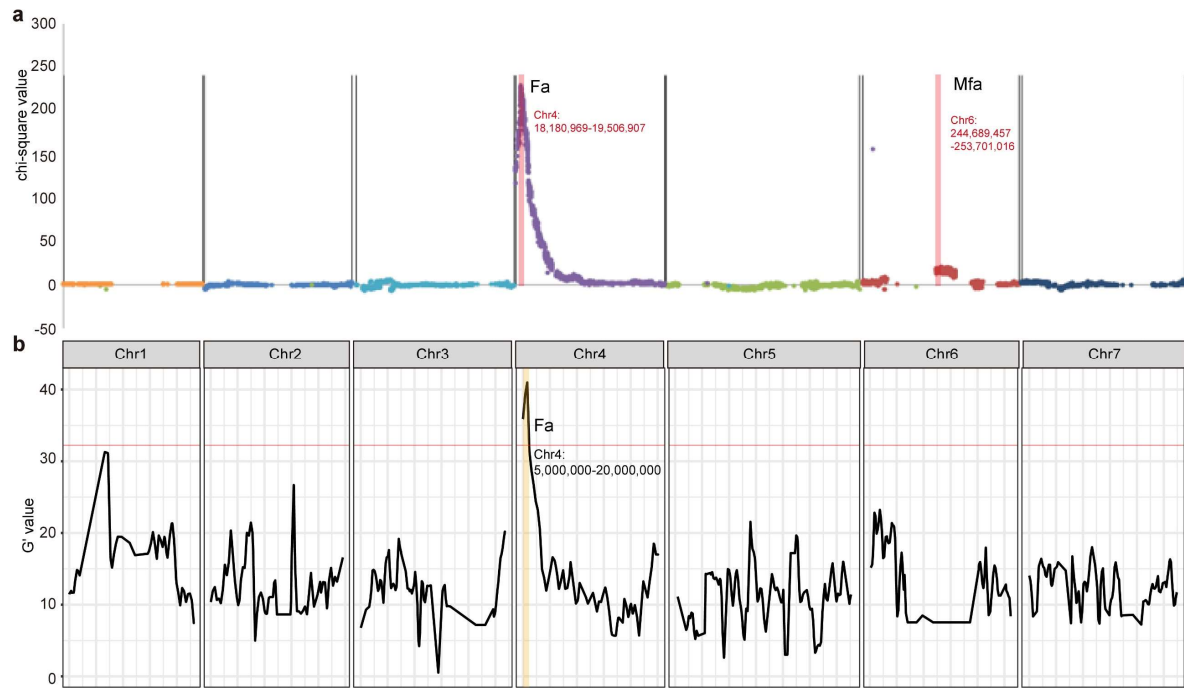


Supplementary Fig. 23 | Gene expression profiles for some of the key elements involved in the formation of pod parchment. (a) qRT_PCR for *PsSERK*, *PsWOX4*, *PsWOX14*, and *PsWOX14-2* involved in the *CLE41-PXY-WOX* signalling pathway (corresponding to Extended Data Fig. 4). **(b)** RNA-seq for *PsCLE41*, *PsPXY*, *PsMYB26*, and *PsNAC*. **(c)** RNA-seq for some of the key genes, including *PsSERK*, *PsWOX4*, *PsWOX14*, *PsWOX14-2*. Comparison was made between accessions of *PPVV*, *ppVV*, *PPvv*, *ppvv*. Three biological replicates of the selected lines for the target gene were used in each experiment.

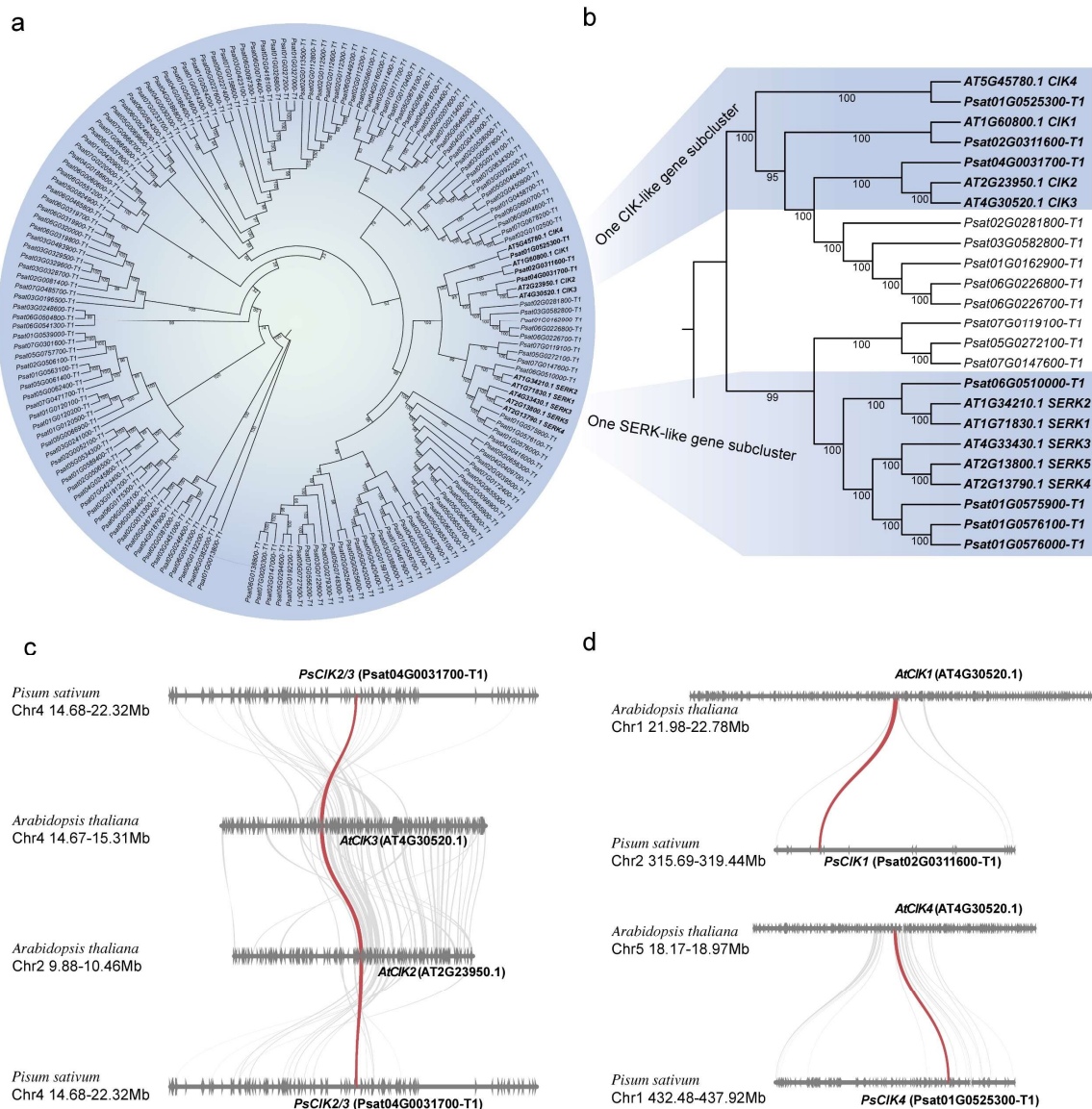


Supplementary Fig. 24 | The *PsMYB26* and upstream sequence together with gene expression and DNA methylation patterns are shown. (a) Top: genomic structure of the *MYB26-Ogre* region in a *vv* line; middle: RNA-seq read mapping showing the transcript abundance of different elements in this region; bottom: DNA methylation level estimation by formula: methylation level (%) for each window = 100 * the sum of the methylation levels of the C sites in each window size / the total number of methylated C sites of this type in the region. Three biological replicates for each sample. Window size (bin) = 2kb. **(b)** The detailed

structure of the *Ogre* LTR element; **(c)** The construction of the full-length transcripts determined by the PacBio Iso-seq full-length transcriptome sequencing; **(d)** IBS-based K-mer distance analysis showing that all the 26 *vv* lines clustered into a single haplotype corresponding to the parchmentless phenotype.

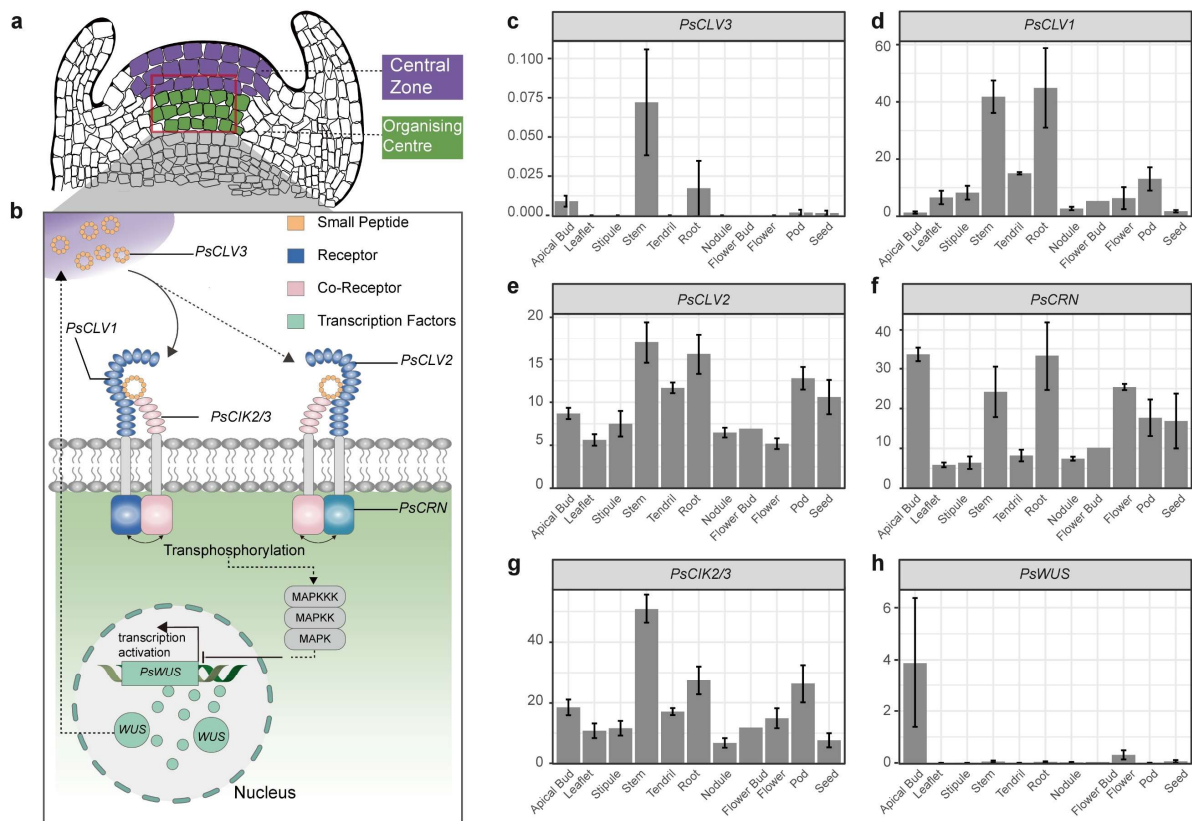


Supplementary Fig. 25 | Chi-square test and Bulk Segregant Analysis (BSA) for the F₂ population derived from the cross between JI2822 (*FaFa*) and JI0816 (*fafa*) to locate fasciation genetic locus. (a) Chi-square value for the 3x2 contingency test of G, H, or Y vs Dominant vs Recessive phenotype, where G is the homozygous JI2822 allele, H is the heterozygote and Y is homozygous for the JI0816 allele (Supplementary Table 18). Where there were fewer than expected of the recessive, JI0816, homozygote combination then we made the sign of the chi-square negative. The candidate intervals of *Fa* and *Mfa* are marked in red in the figure. *Fa*, narrowing the target region into about 1Mb at the end of Chr4. **(b)** BSA analysis of the F₂ population derived from the cross of JI2822 (*FaFa*) and JI0816 (*fafa*). The loci for *Fa* (Chr4) is marked by yellow bars and the regions were shown on the plot.



Supplementary Fig. 26 | Phylogenetic tree reconstruction and synteny analysis of the *CIK*-like gene family. It shows that our candidate gene to *Fa* (*Psat04G0031700*, designated in this study as *PsCIK2/3*) is an orthologue of the *Arabidopsis AtCIK2/3*. **(a)** A phylogenetic tree showing the relationship of the homologous gene sequences obtained from both *Arabidopsis thaliana* and *Pisum sativum*. The homologous gene sequences were obtained using a few core gene members from *Arabidopsis* (*AtCIK1*, *AtCIK2*, *AtCIK3*, *AtCIK4*, and *AtSERK1-5*). The homologues were searched by blastp followed by Orthofinder gene clustering. MAFFT was used to align protein sequences. iqtree2 was implemented using the WAG+I+G4+F model to construct a phylogenetic tree. The bootstrap calculation parameter is set to -bb 1000. **(b)** Two sub-clusters, where the *AtCIK1-4* and *AtSERK1-5* were located, are expanded to show the relationships of genes from *Pisum sativum* with their orthologues to *Arabidopsis thaliana*. **(c)**

Syntenic and collinear blocks analysis demonstrate the orthologous relationship between *PsCIK2/3* (*Psat04G0031700*), *AtCIK2* (*AT2G23950.1*) and *AtCIK2* (*AT4G30520.1*), where a clear duplication event occurred within *Arabidopsis* between *AtCIK2* and *AtCIK3*. (d) The orthologue of *PsCIK1* (*Psat02G0311600*) is *AtCIK1*; the orthologue of *PsCIK4* (*Psat01G0525300-T1*) is *AtCIK4*. No collinear block was observed between *PsCIK2/3* and *AtCIK1* or *PsCIK4*.



Supplementary Fig. 27 | Differential gene expression and the proposed regulatory module for *PsCIK2/3*. (a) Illustration of the shoot apical meristem (SAM) with the central zone and the organising center highlighted. The SAM is composed of the upper central zone, the organizing center underneath, and the flanking peripheral zones (orange). (b) A proposed genetic regulatory module for the roles of *PsCIK2/3*, *PsWUS*, and *PsCLV3* in the *CLV3-WUS* pathway⁹⁵ in pea. The *CLV3* signal peptide binds to receptor kinases such as *CLV1*, and also binds to receptor-like kinases such as *CIK*, which may transmit signals to the cell nucleus through the phosphorylation pathway, promote the transcriptional activation of the *WUS* gene, and thus maintain the homeostasis of stem cells. Comparison of the gene expression profile by RNA-seq across different developmental stages in Caméor for (c) *PsCLV3*; (d) *PsCLV1*; (e) *PsCLV2*; (f) *PsCRN*; (g) *PsCIK2/3*; (h) *PsWUS*.

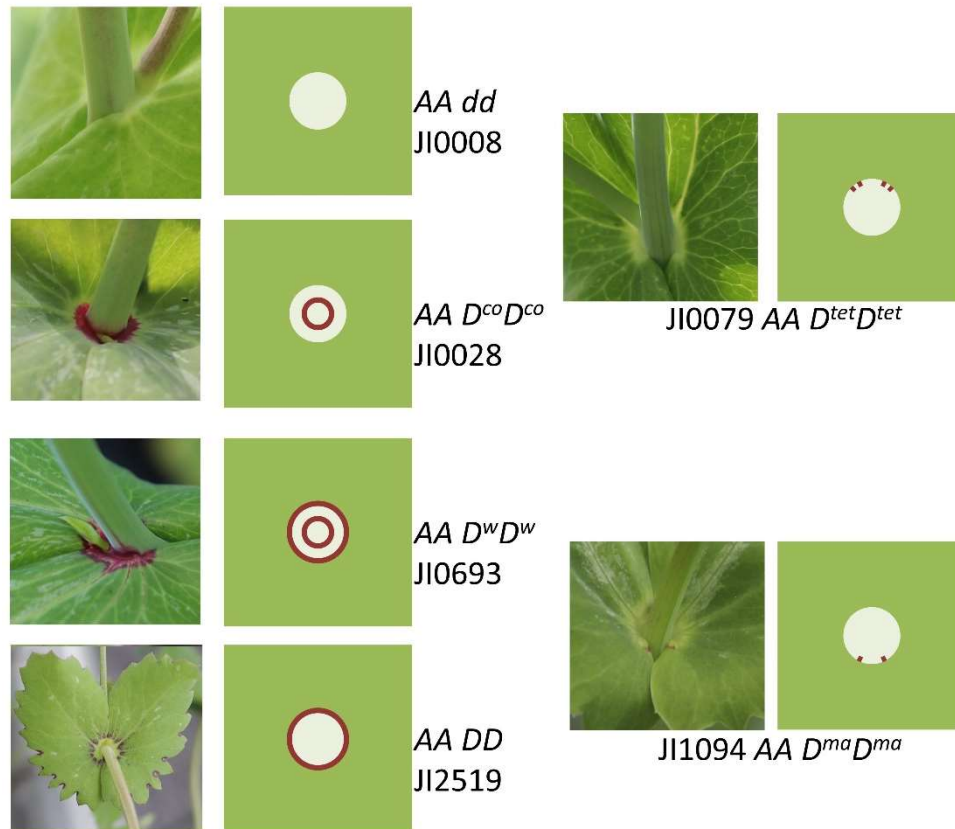
a

	Y (<i>fa/fa</i>)	H (<i>Fa/fa</i>)	G (<i>Fa/Fa</i>)	
	26	4 (d)	2 (c)	Y (<i>Mfa/Mfa</i>)
fasciated	54	0	1 (b)	H (<i>Mfa/maf</i>)
	7 (e)	0	0	G (<i>mfa/mfa</i>)
	0	52	24	Y (<i>Mfa/Mfa</i>)
wild type	1 (a)	90	50	H (<i>Mfa/maf</i>)
	30	31	25	G (<i>mfa/mfa</i>)
genotypes	<i>fa/fa Mfa/Mfa</i>	<i>Fa/fa Mfa/Mfa</i>	<i>Fa/Fa Mfa/Mfa</i>	
	<i>fa/fa Mfa/mfa</i>	<i>Fa/fa Mfa/mfa</i>	<i>Fa/Fa Mfa/mfa</i>	
	<i>fa/fa mfa/mfa</i>	<i>Fa/fa mfa/mfa</i>	<i>Fa/Fa mfa/mfa</i>	

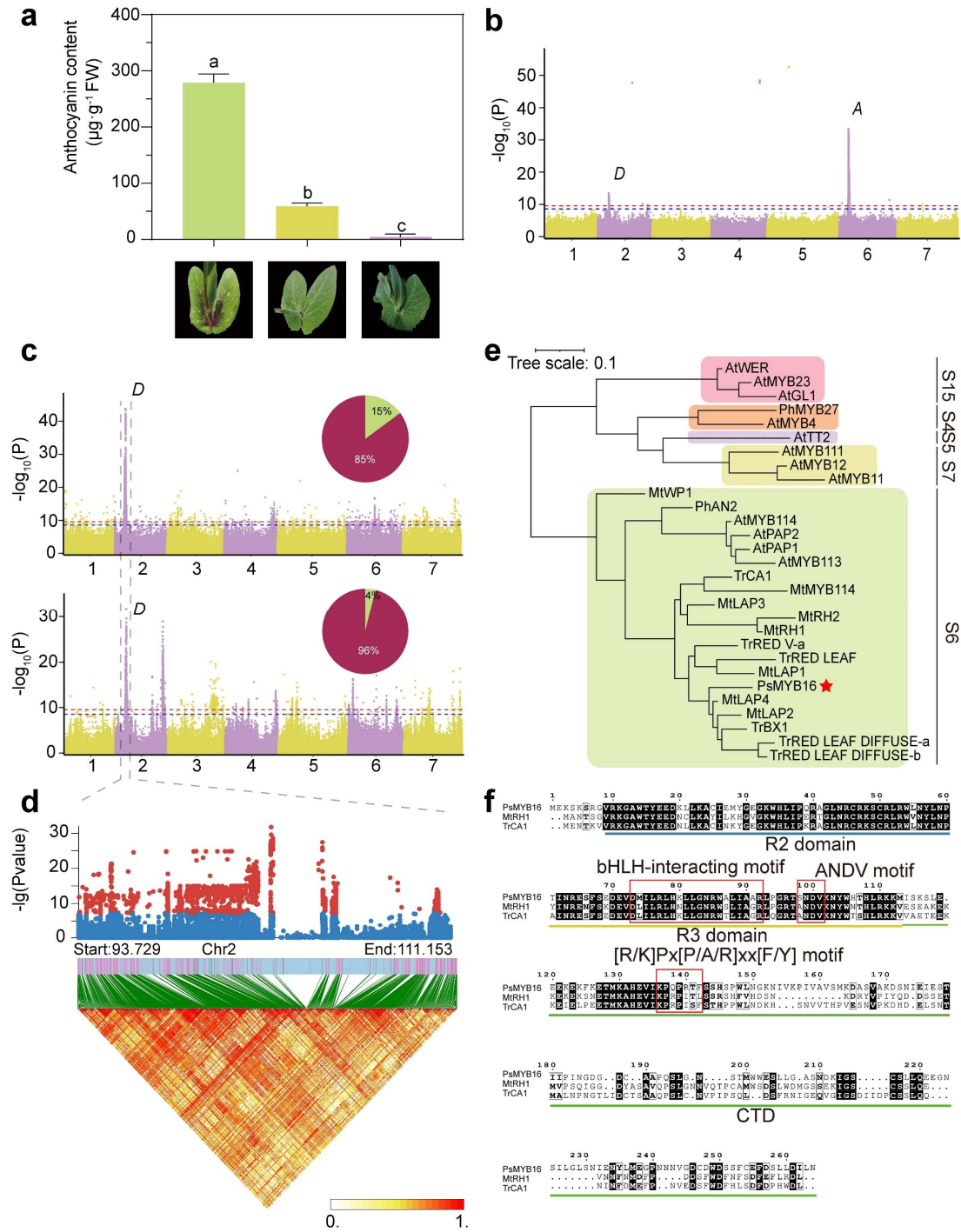
b

Ref in table a	F2 phenotype	F2 genotype		F3 phenotype		conclusion
(#)		<i>Fa</i>	<i>Mfa</i>	fasciated	WT	
a (1)	WT	Y (<i>fa/fa</i>)	H (<i>Mfa/maf</i>)	8	2	<i>fa/fa</i> , but segregating
b (1)	fasciated	G (<i>Fa/Fa</i>)	H (<i>Mfa/maf</i>)		10	F2 mis-score
c (2)	fasciated	G (<i>FaFa</i>)	Y (<i>MfaMfa</i>)		20	F2 mis-score
d (3)	fasciated	H (<i>Fafa</i>)	Y (<i>MfaMfa</i>)	7	21	F2 mis-score
e (7)	fasciated	Y (<i>fa/fa</i>)	G (<i>mfamfa</i>)	51	14	<i>fa/fa</i> , <i>mfa/mfa</i> can be WT
10	WT	Y (<i>fa/fa</i>)	G (<i>mfamfa</i>)	70	18	<i>fa/fa</i> , <i>mfa/mfa</i> can be WT

Supplementary Fig. 28 | (a) Segregation of fasciation in the F₂ of the cross JI2822 x JI0816 (Supplementary Table 18) based on the segregation of *Fa/fa* and *Mfa/mfa*. This figure corresponds to Extended Data Fig. 7. Discrepancies to be tested in F₃ progeny are highlighted in red. The number of F₂ individuals in each genetic class is indicated. Unexpected values are in red (see Supplementary Table 18 and Extended Data Fig. 7 and panel b below). **(b)** F₃ progeny tests of selected F₂ genotypes and phenotypes. #, the number in brackets is the number of F₂ plants of this class which contributed F₃ seed.

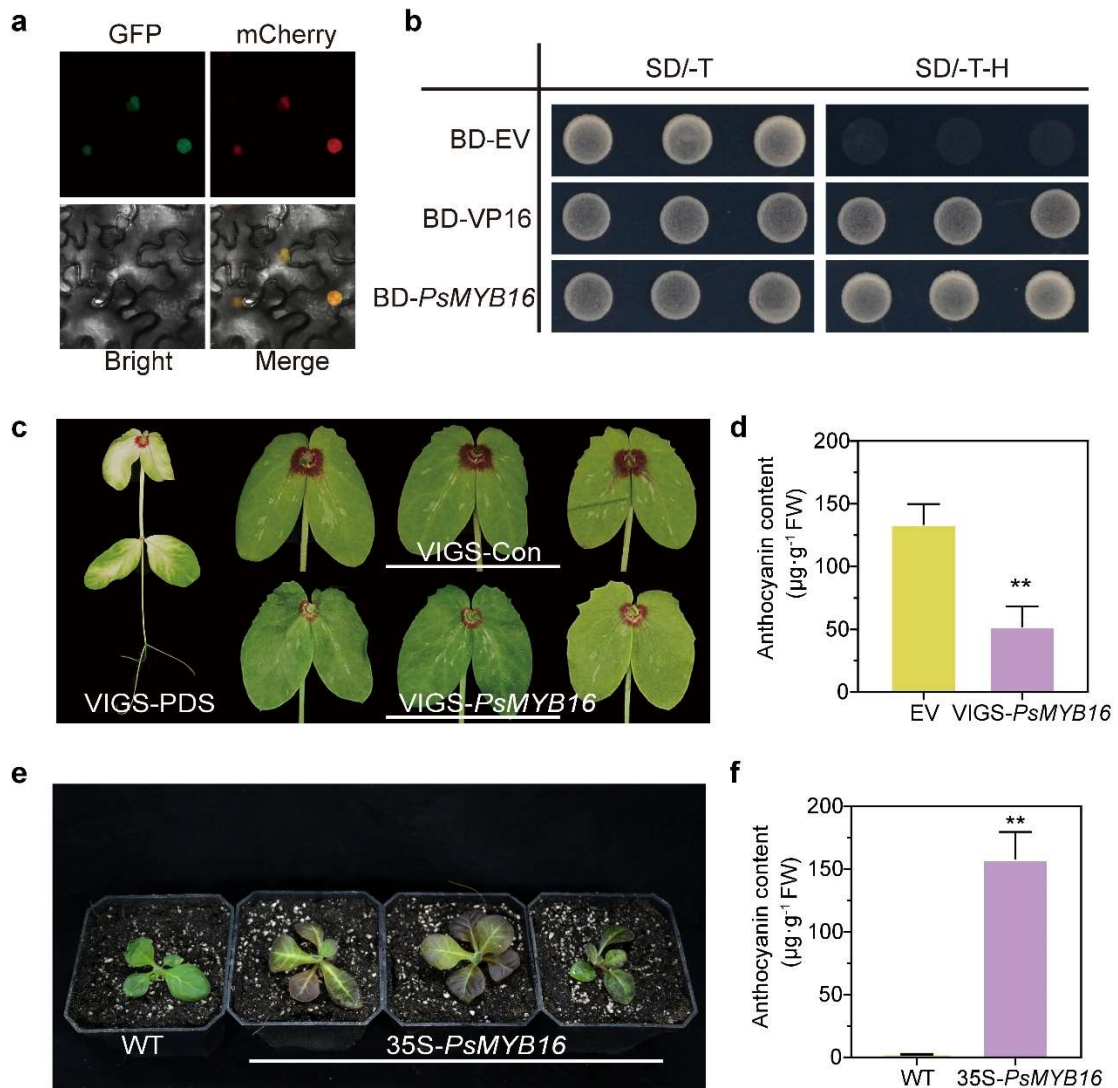


Supplementary Fig. 29 | The different *D* allele axil ring pigmentation phenotypes. The alleles *D^w* *D^{co}* *D^{tet}* *D^{ma}* and *d* are illustrated, as displayed in the JI accessions indicated. Note that JI0008, the type line for *d*, shows stem anthocyanin pigmentation, indicating the presence of the dominant *A* allele. JI2519 is an example of an accession where the inner axil ring is absent. These accessions often have additional anthocyanin pigmentation at the leaf margin, along the veins or as spots on the stipule lamina. These extensive anthocyanin pigmentation patterns on the lamina and on the veins have been assigned separate gene symbols *An* and *And* by Świącicki (1990)⁸⁸. *And* is linked to, but distinct from *D*, while the genetics of *An* have yet to be described.

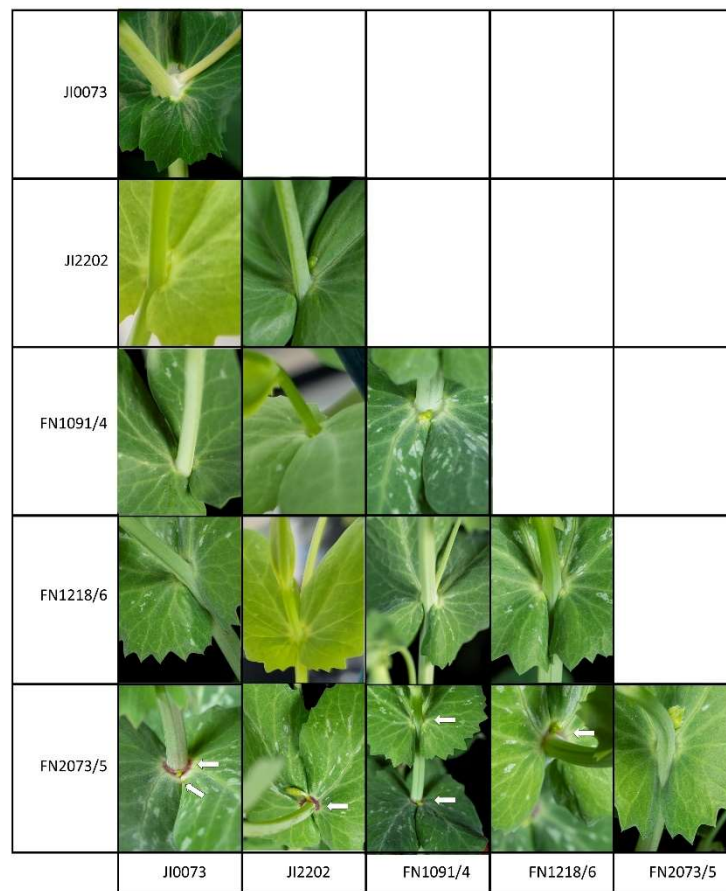


Supplementary Fig. 30 | Characterisation of one of the *MYB* genes within the gene cluster related to axillary anthocyanin. (a) Determination of total anthocyanin content in leaf axillary zones within the population. **(b)** Genome-wide association analysis of leaf axillary anthocyanins. **(c)** Genome-wide association analysis of leaf axillary anthocyanins after excluding white-flowered samples. The top row represents data from Shenzhen in 2021, and the bottom row represents data from Harbin in 2022. **(d)** Locus LD heatmap for the genomic interval under the peak. **(e)** Phylogenetic tree construction using *PsMYB16* and the *MYB*

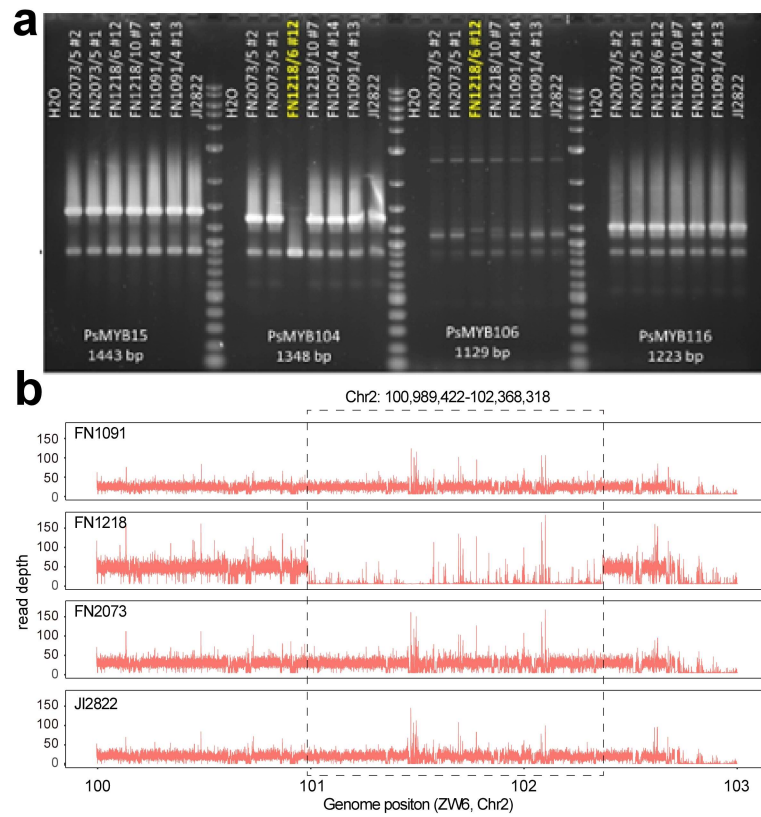
transcription factor family. (f) Multiple sequence alignment of *PsMYB16* with other genes controlling leaf anthocyanin.



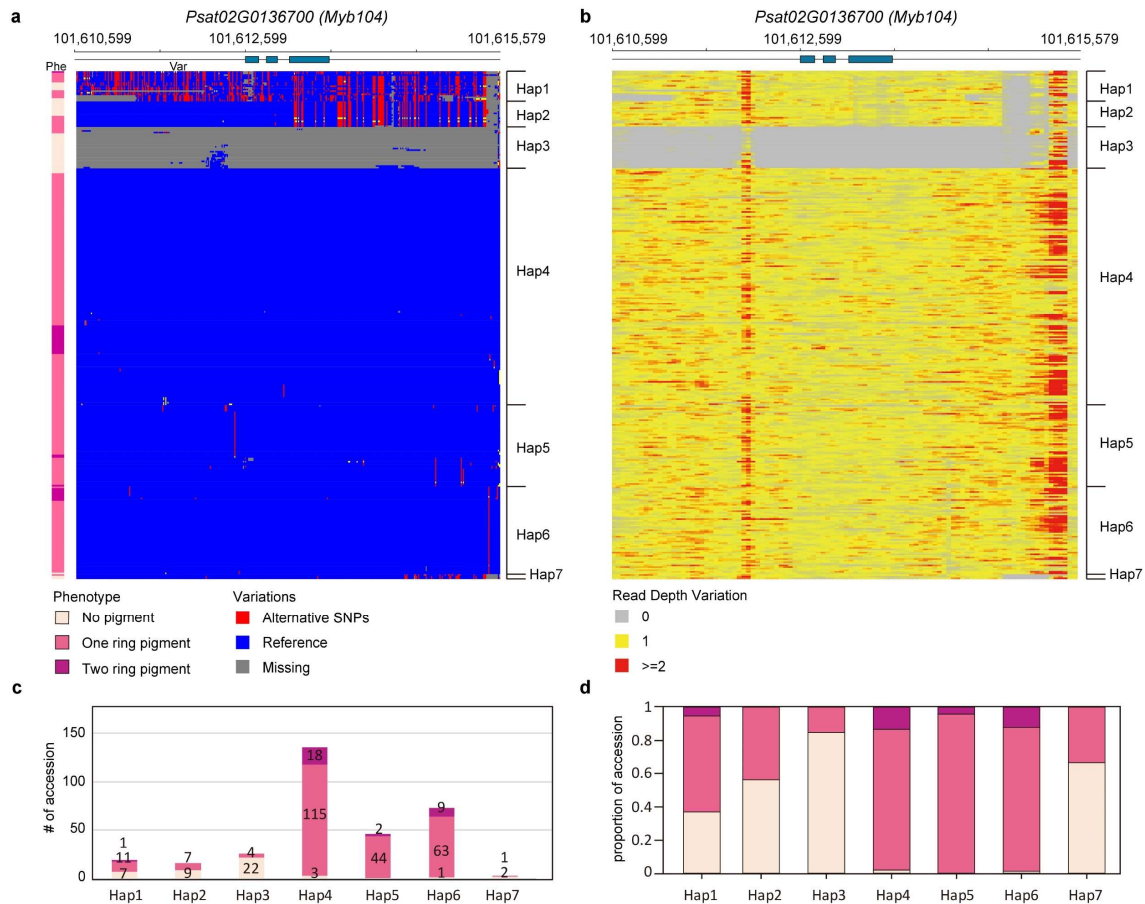
Supplementary Fig. 31 | D VIGS experiment. (a) Subcellular localization of *PsMYB16* (*Psat02G0138300*). (b) Transcriptional activation experiment of *PsMYB16*. (c) Phenotypes after gene silencing of *PsMYB16* using VIGS. The left panel shows a PDS control (VIGS-PDS), the top row of three pairs of leaflets is an empty vector control (VIGS-Con) and the lower three pairs shows the knock-down of *Psat02G0138300* (VIGS-*PsMYB16*). The plants in which these experiments were carried out have the *D^w* allele and so have both an inner and outer ring of axil pigmentation, suggesting that *Psat02G0138300* has a role in patterning axil ring pigmentation. (d) Anthocyanin measurement after gene silencing of *PsMYB16*. (e) Phenotypes in a transient overexpression assay of *PsMYB16* in *N. benthamiana*. (f) Anthocyanin measurement after overexpression of *PsMYB16*. ** represent a significant level at $P < 0.01$ using a student's t-test in panel d and f.



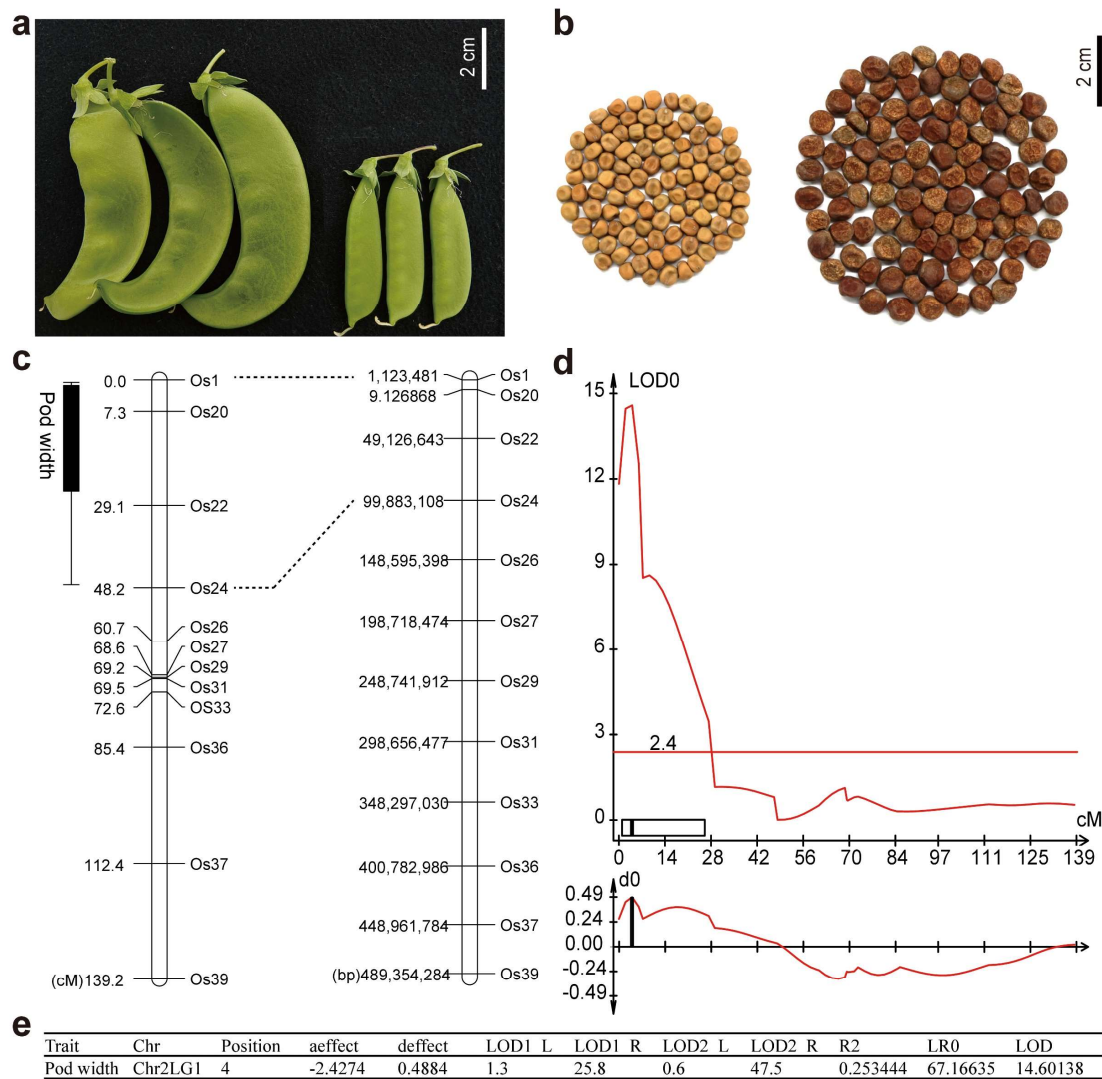
Supplementary Fig. 32 | Allelism tests of *d* mutants. The axils of self-seeded plants and F₁ hybrids of crosses between the lines indicated by the rows and columns are shown. There are two complementation groups, one of which corresponds to *d* from Jl0073 and Jl2202 and the other to FN2073/5.



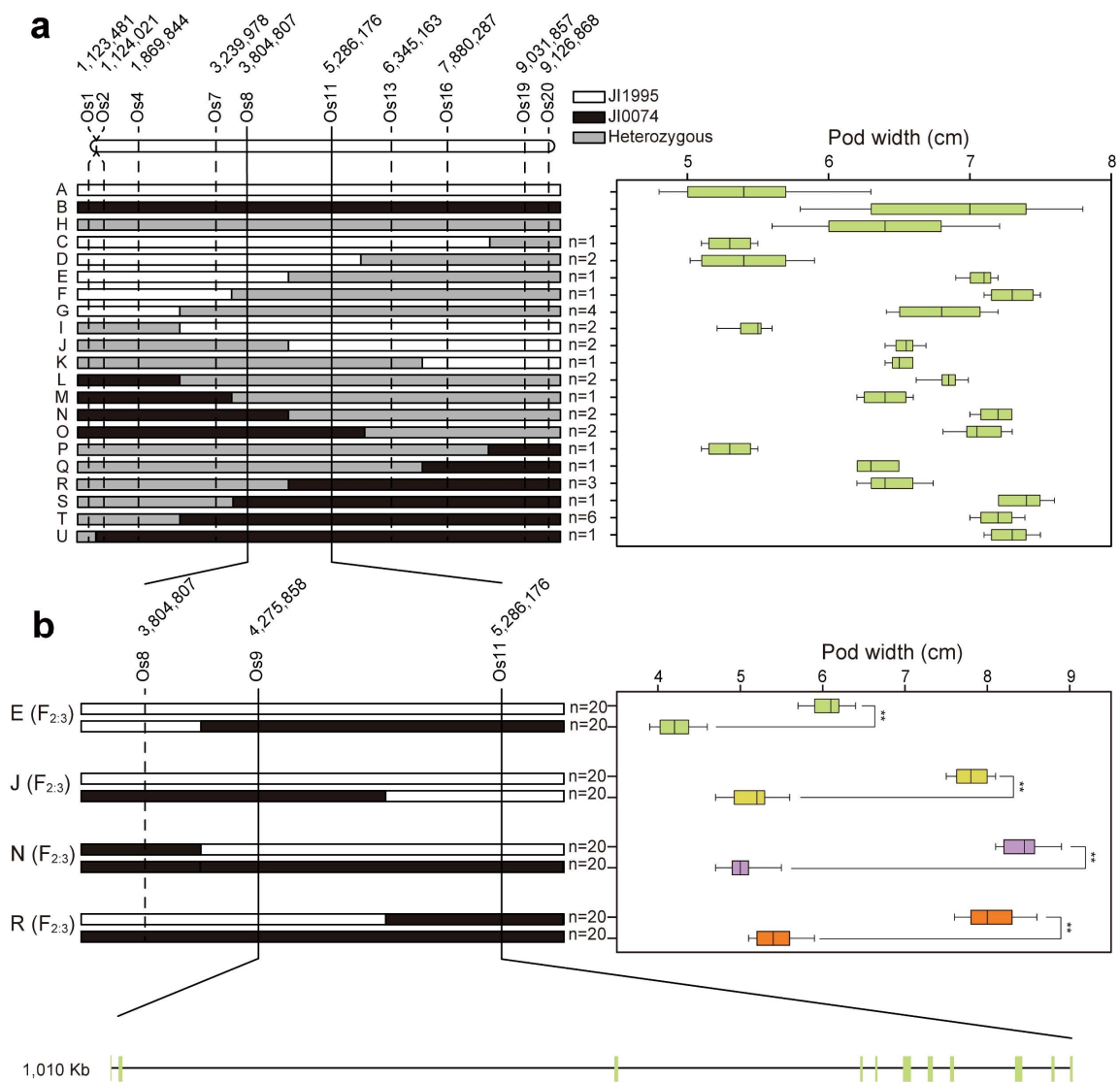
Supplementary Fig. 33 | *MYB* gene deletions. Confirmation of the deletion of *MYB104* and *MYB106* in FN1218 mutant. **(a)** PCR amplification of *MYB* genes, confirming the deletion of *PsMYB104* and *PsMYB106*, but not *PsMYB15* or *PsMYB116* in the FN1218/6 plant #12 (plant #10 was a wild type M4 segregant). **(b)** The read depth of FN lines and JI2822 aligned against the ZW6 reference genome.



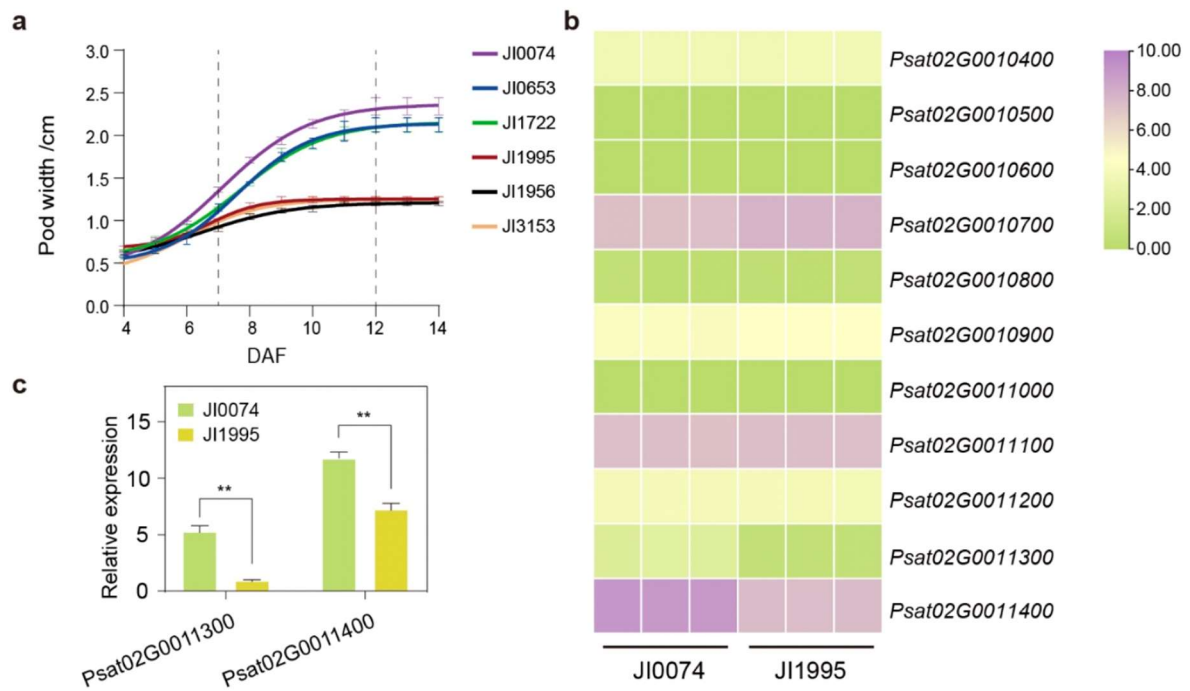
Supplementary Fig. 34 | The various *d* alleles are different haplotypes of several *MYB* genes, and *MYB104* is shown to be associated with the phenotypic effect. (a) Haplotypes of *Psat02G0136700 (MYB104)*. The phenotypes and variants are colour coded as shown at the bottom of the haplotype plot. **(b)** Variation of read depths of SNPs around *Psat02G0136700 (MYB104)*. The read depth is colour coded, with grey, yellow and red showing read depth of 0, 1 and a value greater than 2, respectively. The haplotypes are shown on the right of the plot. **(c)** The number of lines with different phenotypes categorized in each haplotype. **(d)** The proportion of lines with different phenotypes in each haplotype. The phenotype colour code in **c** and **d** is the same as the code shown in panel **a**.



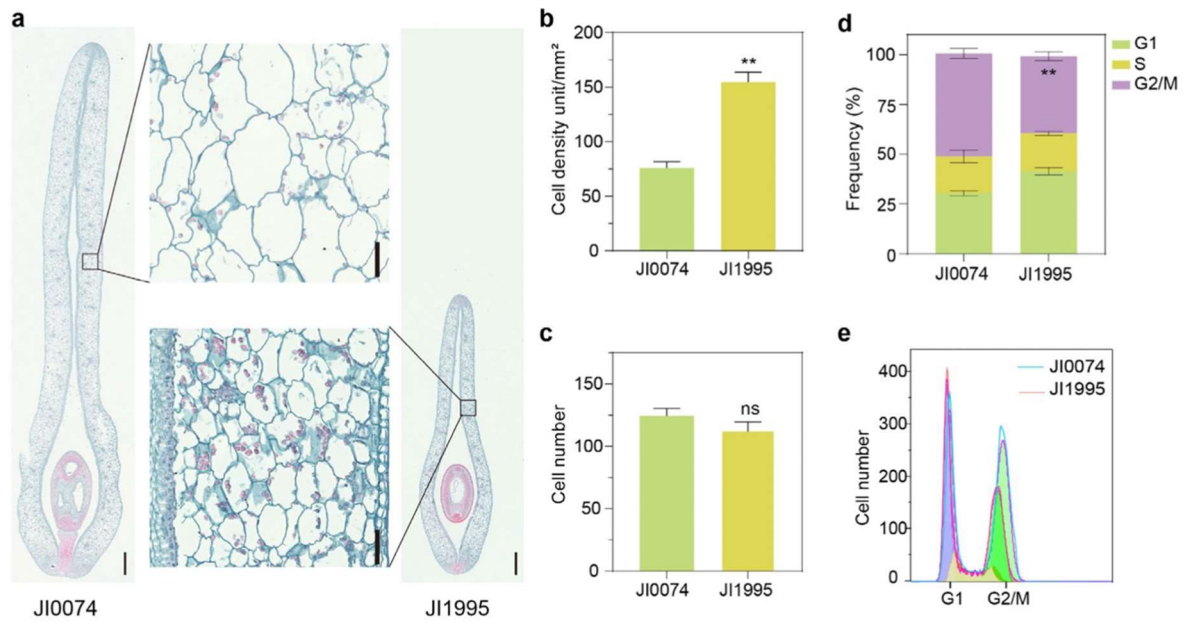
Supplementary Fig. 35 | QTL mapping of *Os1* locus. Comparison of the phenotypes of **(a)** Wide (JI0074) and narrow pods (JI1995), **(b)** Large and small grains, in the two parental lines of JI0074 and JI1995, respectively, which were used to build the F₂ mapping population. The scale bars in **a** and **b** represent 2 cm. **(c)** The linkage group of KASP markers (left) mapped using a F₂ population derived from the cross between JI0074 and JI1995. The physical positions of these markers in this candidate region are shown in the right-hand map. **(d)** Interval Mapping scores (logarithm of odds, LOD) of markers for pod width. Horizontal red lines indicated the threshold of LODs. **(e)** Location and statistics of the target *Os1* locus.



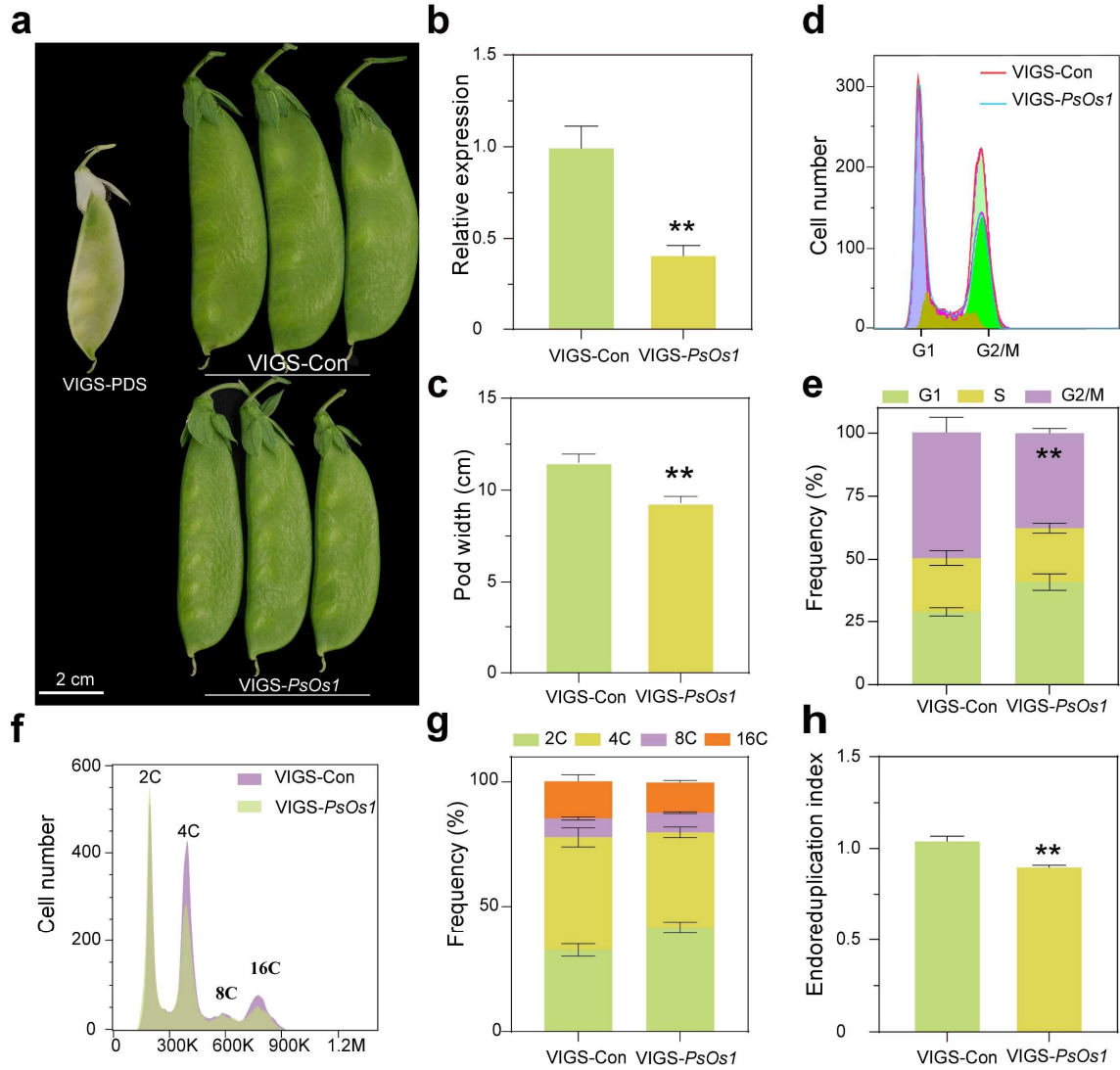
Supplementary Fig. 36 | Fine mapping of the *OsI* locus. (a) The graphical distribution of the genotypes in the candidate regions (left) and distribution of phenotypes of pod width (right) of the lines in a F₂ population derived from the cross of JI0074 and JI1995. The white, black, and grey bars in the box represent genotypes of JI1995, JI0074 and heterozygotes, respectively. The number of individual plants (n) of each genotype is shown. (b) The graphical genotypes in the candidate regions (left) and phenotypes of pod width (right) of the lines in F_{2:3} populations derived from the cross of JI0074 and JI1995, which enabled the narrowing down of the candidate regions into a locus containing only 11 genes, as shown beneath. The statistics of pod width are shown on the right and the significant differences are indicated by * (P < 0.05), ** (P < 0.01) (Student's t-test).



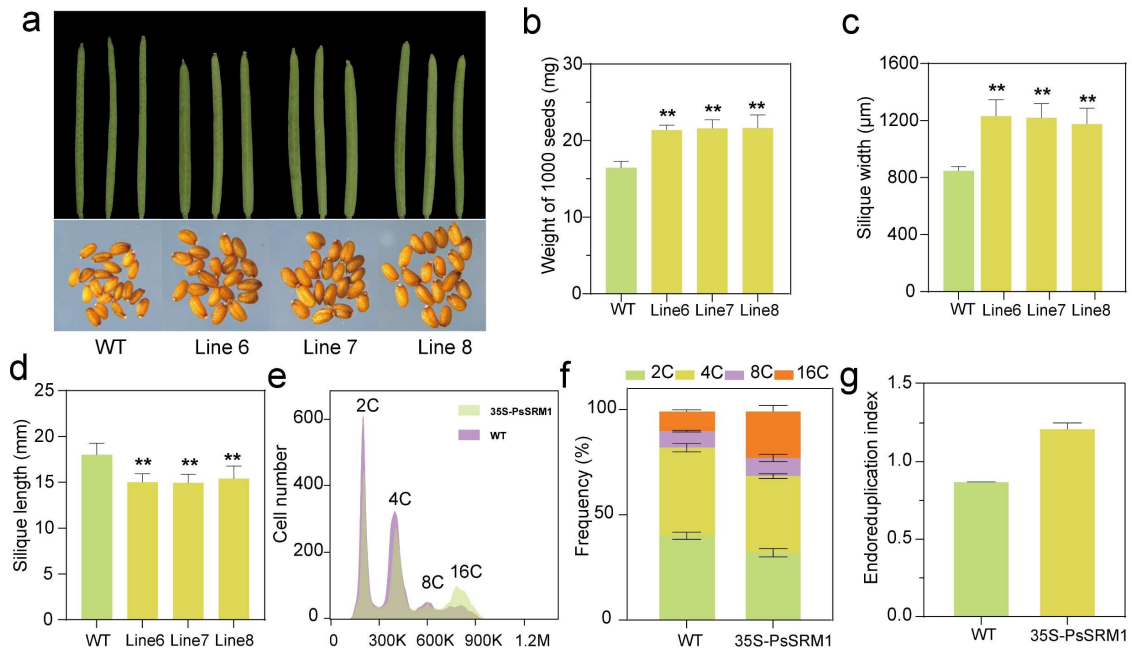
Supplementary Fig. 37 | Expression analysis of candidate genes at the *PsOs1* locus. (a) Growth curves of wide-pod and narrow-pod varieties. **(b)** Heatmap of relative gene expression within candidate intervals for the tissues in the pods. **(c)** Fluorescence quantitative detection of differentially expressed genes. Three biological replicates were used for each sample.



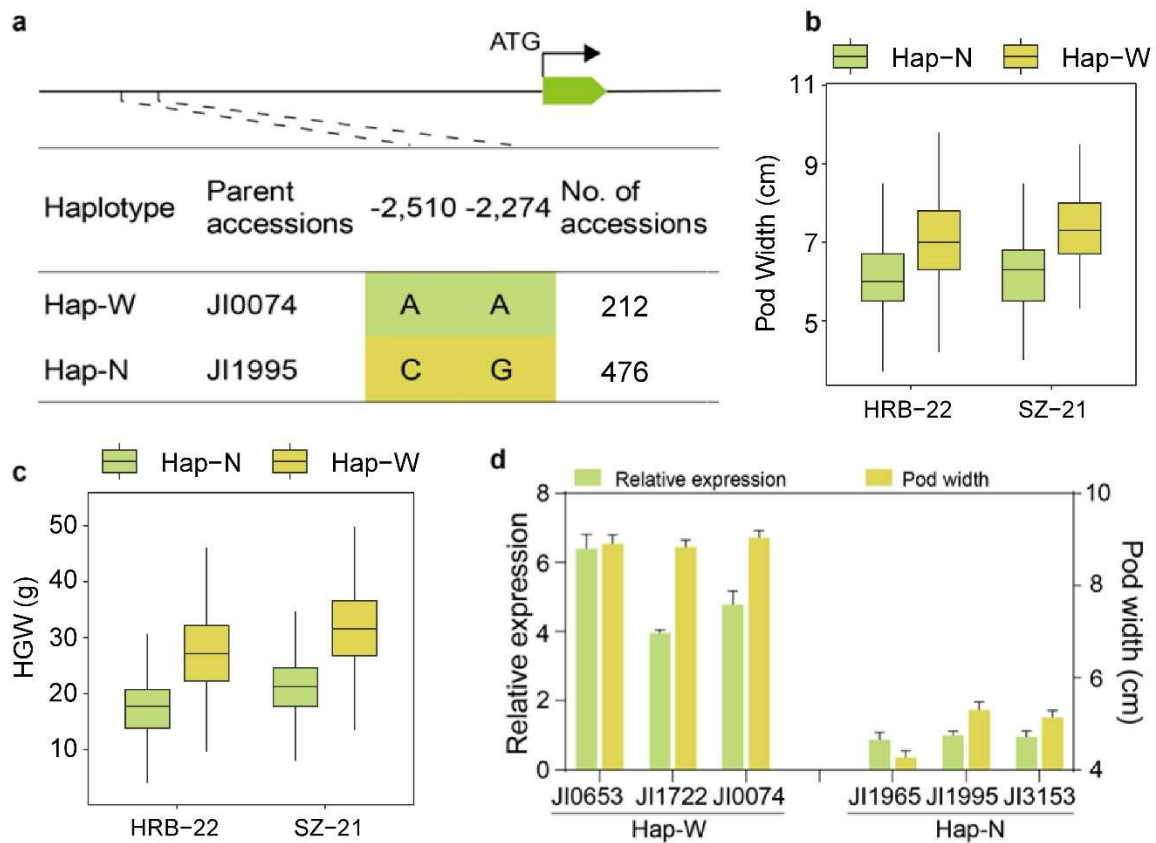
Supplementary Fig. 38 | Cellular observations and comparison in the two parental lines differing in pod width. (a) Cross-section of pods from JI0074 and JI1995 at 12 days after flowering, stained with Fuchsin Green dye. **(b)** Cell density per unit area within the pod cross-sections of JI0074 and JI1995. **(c)** Number of cells in transverse arrangement within the pod cross-sections of JI0074 and JI1995. **(d)** Cell cycle statistics for pods of the two lines. **(e)** Flow cytometry display of cells at different cell cycles in pods of the two lines.



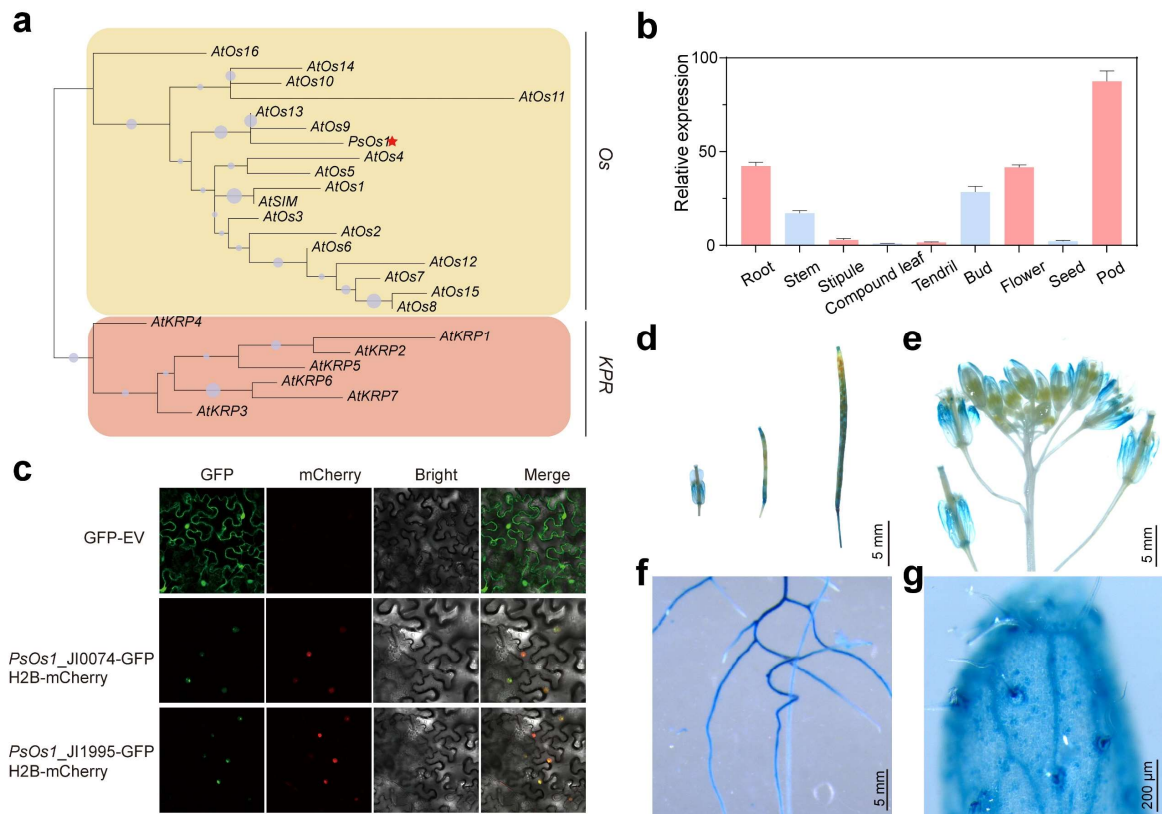
Supplementary Fig. 39 | VIGS experiment for *PsOs1*. (a) Phenotypic changes in *PsOs1*-silenced lines of pea (VIGS-*PsOs1*) in comparison with the negative control, a line inoculated with VIGS-Con. (b) Relative expression level of *PsOs1* gene in the lines inoculated with VIGS-Con (control) and VIGS-*PsOs1*. (c) Pod width of lines of VIGS-Con and VIGS-*PsOs1*. (d) Quantification of cells at different phases of cell cycle in lines of VIGS-Con and VIGS-*PsOs1*. (e) Proportion of cells at G1, S and G2/M phases in lines of VIGS-Con and VIGS-*PsOs1*. (f) Quantification of cells at different ploidy levels (2C-16C) in lines of VIGS-Con and VIGS-*PsOs1*. (g) Proportion of cells with different ploidy levels (2C, 4C, 8C and 16C) in lines of VIGS-Con and VIGS-*PsOs1*. (h) Endoreduplication index of lines of VIGS-Con, VIGS-*PsOs1*. *** indicates statistical significance at $P < 0.01$ (Student's t-test).



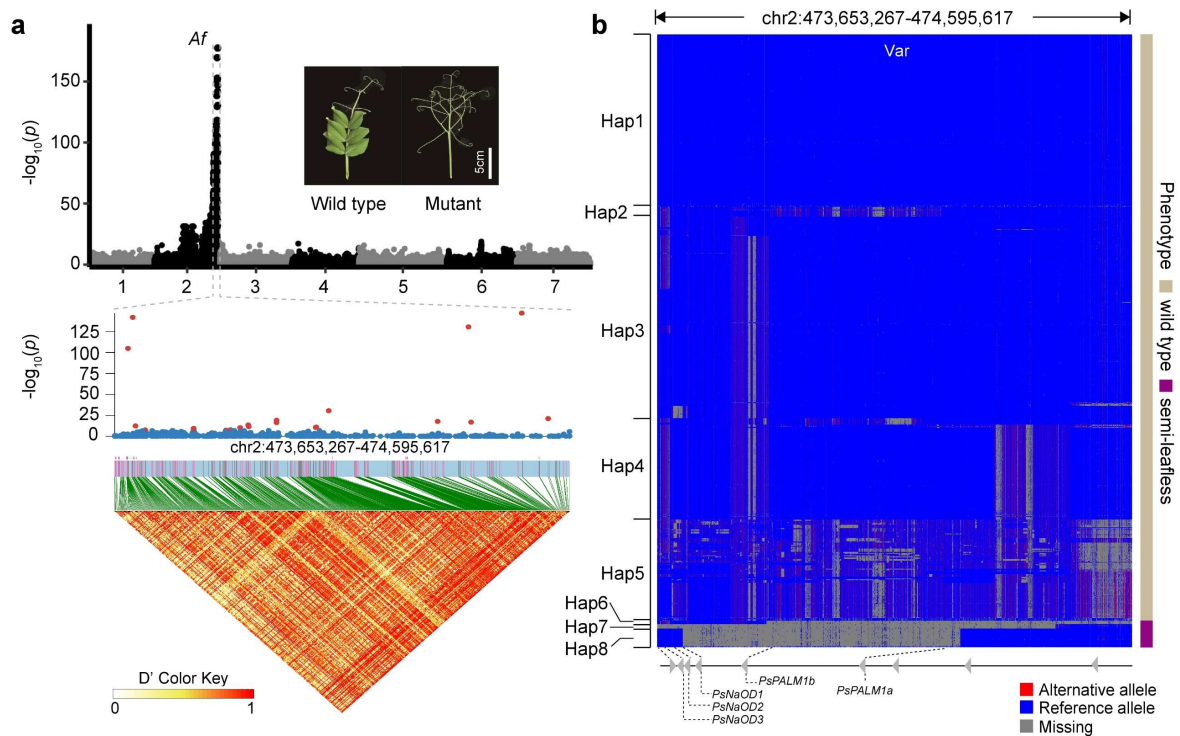
Supplementary Fig. 40 | Analysis of the phenotype of *PsOs1* transgenic Arabidopsis. (a) Overexpression of *PsOs1* in Arabidopsis resulted in an increase in silique width and seed weight. **(b)** Difference in weight of 1000 seeds between the wild-type (WT) and transgenic lines of Arabidopsis expressing *PsOs1*. **(c)** Difference in silique width between the wild-type and transgenic lines of Arabidopsis overexpressing *PsOs1*. **(d)** Difference in silique length between the WT and transgenic lines of Arabidopsis overexpressing *PsOs1*. **(e)** Quantification of cells at different ploidies (2C-16C) in Arabidopsis overexpressing *PsOs1* and wild type. **(f)** Proportion of cells with different ploidies (2C, 4C, 8C and 16C) in Arabidopsis overexpressing *PsOs1* compared to WT. **(g)** Endoreduplication index of Arabidopsis lines overexpressing *PsOs1* compared to WT.



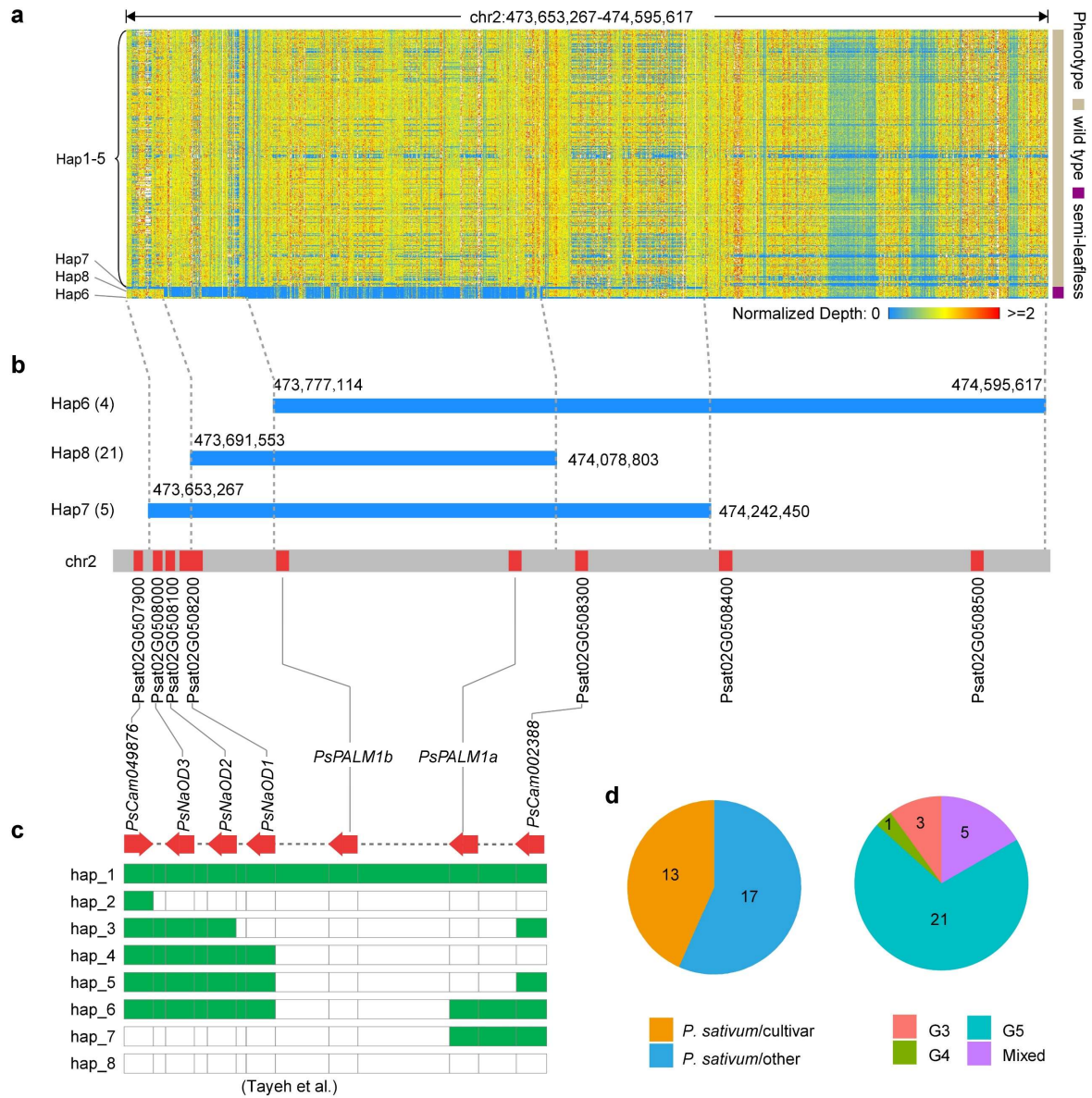
Supplementary Fig. 41 | Analysis of the haplotypes of the promoter region of *PsOs1*. (a) Haplotypes (Hap-W and Hap-N) of *PsOs1* separate the germplasms with wide pods from those with narrow pods. (b) The pod width of germplasms with Hap-W and Hap-N. (c) Hundred-seed weight (HSW) of the seeds of germplasms with Hap-W and Hap-N. (d) Relative expression levels and pod width of *PsOs1* in the pods of selected accessions of Hap-W and Hap-N.



Supplementary Fig. 42 | Expression analysis of *PsOs1*. (a) Clustering of *PsOs1* and homologous genes in Arabidopsis in the phylogenetic tree. (b) Relative expression levels of *PsOs1* in different tissues, including pod, of pea. (c) Subcellular location of *PsSMR1* indicated by the expression of *PsOs1*-GFP fusion protein in tobacco. (d) GUS staining of the silique of transgenic Arabidopsis (*PsOs1pro::GUS*) expressing GUS promoted by the promoter of *PsOs1*. (e) GUS staining of the inflorescence of transgenic Arabidopsis (*PsOs1pro::GUS*) expressing GUS promoted by the promoter of *PsOs1*. (f) GUS staining of the root of transgenic Arabidopsis (*PsOs1pro::GUS*) expressing GUS promoted by the promoter of *PsOs1*. (g) GUS staining of the trichomes on the sepals of transgenic Arabidopsis (*PsOs1pro::GUS*) expressing GUS promoted by the promoter of *PsOs1*.



Supplementary Fig. 43 | GWAS and haplotype analysis of the *Afila/afila* phenotype. (a) SNP-based GWAS (top left) and local Manhattan plot with the local LD heatmap for the genomic interval under the peak. The genomic interval that includes the three types of deletion events is about 940kb. **(b)** Haplotype-phenotype association study reveals eight different haplotypes, in which Hap1-5 correspond to the wild type, whereas Hap6-8 correspond to the so-called ‘semi-leafless’ phenotypes that have the deletion events. The five key genes (*PsPALM1a/b*, *PsNaOD1-3*) are given below the heatmap.



Supplementary Fig. 44 | Gene constitution and haplotype for the *afila* phenotypes. (a) Indels-based haplotype-phenotype association for the 940kb genomic interval. The insertion and deletion were calculated based on read depth followed by normalization. **(b)** All the 30 ‘semi-leafless’ accessions are clustered into the three types of deletion haplotypes (Hap6-8); the deletion end-points and the key deleted genes (such as *PsPALM1a/b*) are illustrated. **(c)** Correspondence of the ZW6 gene ID and the markers reported by Tayeh et al. (2024)⁹⁶. The haplotypes (hap_1-8) used in this panel were the names used by Tayeh et al. **(d)** The subgroup distribution of the 30 *afaf* genotypes. All the 30 *afaf* genotypes are from *Pisum sativum*.

3. References:

- 1 Trněný, O. *et al.* Molecular evidence for two domestication events in the pea crop. *Genes* **9**, 535 (2018).
- 2 Dahl, W. J., Foster, L. M. & Tyler, R. T. Review of the health benefits of peas (*Pisum sativum* L.). *Br. J. Nutr.* **108**, S3-S10 (2012).
<https://doi.org/10.1017/S0007114512000852>
- 3 Faligowska, A. *et al.* The nitrogen fixation and yielding of pea in different soil tillage systems. *Agronomy* **12**, 352 (2022).
- 4 Tsyganov, V. E. & Tsyganova, A. V. Symbiotic regulatory genes controlling nodule development in *Pisum sativum* L. *Plants* **9**, 1741 (2020).
- 5 Ellis, T. N. *et al.* The taxonomic status of genera within the Fabeae (*Vicieae*, with a special focus on *Pisum*. *Diversity* **16**, 365 (2024).
- 6 Cheng, S. Gregor Mendel: The father of genetics who opened a biological world full of wonders. *Mol. Plant* **15**, 1641-1645 (2022).
<https://doi.org/10.1016/j.molp.2022.10.013>
- 7 Mendel, G. Versuche über Pflanzenghybriden. *Verhandlungen des naturforschenden Vereines in Brünn* **4**, 3-47 (1866).
- 8 Auffray, C. & Noble, D. Gregor Mendel at the source of genetics and systems biology: Celebrating the relevance of Gregor Mendel's experiments on the development of hybrid plants on the occasion of his bicentenary. *Biol. J. Linn. Soc.* **137**, 720-736 (2022).
- 9 Schacherer, J. Beyond the simplicity of Mendelian inheritance. *C.R. Biol.* **339**, 284-288 (2016).
- 10 Stenseth, N. C., Andersson, L. & Hoekstra, H. E. Vol. 119 e2201327119 (National Academy of Sciences, 2022).
- 11 Radick, G. The professor and the pea: Lives and afterlives of William Bateson's campaign for the utility of Mendelism. *Stud. Hist. Philos. Sci. Part A* **44**, 280-291 (2013).
- 12 Visscher, P. M. & Goddard, M. E. From RA Fisher's 1918 paper to GWAS a century later. *Genetics* **211**, 1125-1130 (2019).
- 13 Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796-815 (2000).
- 14 Zhang, X., Henriques, R., Lin, S.-S., Niu, Q.-W. & Chua, N.-H. Agrobacterium-mediated transformation of *Arabidopsis thaliana* using the floral dip method. *Nat. Protoc.* **1**, 641-646 (2006).
- 15 Koornneef, M. & Meinke, D. The development of *Arabidopsis* as a model plant. *The Plant Journal* **61**, 909-921 (2010).
- 16 Shoemaker, R. *et al.* Genome duplication in soybean (*Glycine* subgenus *soja*). *Genetics* **144**, 329-338 (1996).
- 17 Kreplak, J. *et al.* A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* **51**, 1411-1422 (2019). <https://doi.org/10.1038/s41588-019-0480-1>
- 18 Yang, T. *et al.* Improved pea reference genome and pan-genome highlight genomic features and evolutionary characteristics. *Nat. Genet.* **54**, 1553-1563 (2022).
<https://doi.org/10.1038/s41588-022-01172-2>
- 19 Liu, N. *et al.* Reference genome sequence and population genomic analysis of peas provide insights into the genetic basis of Mendelian and other agronomic traits. *Nat. Genet.* **56**, 1964-1974 (2024). <https://doi.org/10.1038/s41588-024-01867-8>
- 20 Poore, J. & Nemecek, T. Reducing food's environmental impacts through producers and consumers. *Science* **360**, 987-992 (2018). <https://doi.org/10.1126/science.aag0216>

- 21 Armstead, I. *et al.* Cross-Species Identification of Mendel's *i* Locus. *Science* **315**, 73-73 (2007). <https://doi.org/10.1126/science.1132912>
- 22 Shirasawa, K., Sasaki, K., Hirakawa, H. & Isobe, S. Genomic region associated with pod color variation in pea (*Pisum sativum*). *G3 Genes/Genomes/Genetics* **11**, jkab081 (2021). <https://doi.org/10.1093/g3journal/jkab081>
- 23 Tishkoff, S. A. & Verrelli, B. C. Patterns of Human Genetic Diversity: Implications for Human Evolutionary History and Disease. *Annu. Rev. Genomics Hum. Genet.* **4**, 293-340 (2003). <https://doi.org/10.1146/annurev.genom.4.070802.110226>
- 24 De Beukelaer, H., Davenport, G. F. & Fack, V. Core Hunter 3: flexible core subset selection. *BMC Bioinf.* **19**, 203-203 (2018). <https://doi.org/10.1186/s12859-018-2209-z>
- 25 Jing, R. *et al.* The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evol. Biol.* **10**, 44-44 (2010). <https://doi.org/10.1186/1471-2148-10-44>
- 26 Bhattacharyya, M. K., Smith, A. M., Ellis, T. H. N., Hedley, C. & Martin, C. The wrinkled-seed character of pea described by Mendel is caused by a transposon-like insertion in a gene encoding starch-branching enzyme. *Cell* **60**, 115-122 (1990). [https://doi.org/10.1016/0092-8674\(90\)90721-P](https://doi.org/10.1016/0092-8674(90)90721-P)
- 27 Burstin, J. *et al.* Genetic diversity and trait genomic prediction in a pea diversity panel. *BMC Genomics* **16**, 105-105 (2015). <https://doi.org/10.1186/s12864-015-1266-1>
- 28 Rayner, T. *et al.* Genetic Variation Controlling Wrinkled Seed Phenotypes in *Pisum*: How Lucky Was Mendel? *Int. J. Mol. Sci.* **18**, 1205-1205 (2017). <https://doi.org/10.3390/ijms18061205>
- 29 Moreau, C. *et al.* An allelic series of starch-branching enzyme mutants in pea (*Pisum sativum* L.) reveals complex relationships with seed starch phenotypes. *Carbohydr. Polym.* **288**, 119386-119386 (2022). <https://doi.org/10.1016/j.carbpol.2022.119386>
- 30 Wang, T. L., Bogracheva, T. Y. & Hedley, C. L. Starch: as simple as A, B, C? *J. Exp. Bot.* **49**, 481-502 (1998). <https://doi.org/10.1093/jxb/49.320.481>
- 31 Casey, R. *et al.* The effect of modifying carbohydrate metabolism on seed protein gene expression in peas. *J. Plant Physiol.* **152**, 636-640 (1998). [https://doi.org/10.1016/S0176-1617\(98\)80023-0](https://doi.org/10.1016/S0176-1617(98)80023-0)
- 32 Kooistra, E. On the differences between smooth and three types of wrinkled peas. *Euphytica* **11**, 357-373 (1962). <https://doi.org/10.1007/BF00031985>
- 33 Singh, N., Jain, P., Ujainwal, M. & Langyan, S. Escalate protein plates from legumes for sustainable human nutrition. *Front. Nutr.* **9** (2022). <https://doi.org/doi:10.3389/fnut.2022.977986>
- 34 Turner, S. R., Barratt, D. H. P. & Casey, R. The effect of different alleles at the *r* locus on the synthesis of seed storage proteins in *Pisum sativum*. *Plant Mol. Biol.* **14**, 793-803 (1990). <https://doi.org/https://doi.org/10.1007/BF00016512>
- 35 Sato, Y., Morita, R., Nishimura, M., Yamaguchi, H. & Kusaba, M. Mendel's green cotyledon gene encodes a positive regulator of the chlorophyll-degrading pathway. *Proc. Natl. Acad. Sci.* **104**, 14169-14174 (2007). <https://doi.org/10.1073/pnas.0705521104>
- 36 Shimoda, Y., Ito, H. & Tanaka, A. Arabidopsis STAY-GREEN, Mendel's Green Cotyledon Gene, Encodes Magnesium-Dechelataase. *The Plant Cell* **28**, 2147-2160 (2016). <https://doi.org/10.1105/tpc.16.00428>
- 37 SanMiguel, P., Gaut, B. S., Tikhonov, A., Nakajima, Y. & Bennetzen, J. L. The paleontology of intergene retrotransposons of maize. *Nat. Genet.* **20**, 43-45 (1998). <https://doi.org/10.1038/1695>
- 38 Jing, R. *et al.* Insertional Polymorphism and Antiquity of *PDR1* Retrotransposon Insertions in *Pisum* Species. *Genetics* **171**, 741-752 (2005). <https://doi.org/10.1534/genetics.105.045112>

- 39 Ubayasena, L., Bett, K., Tar'an, B., Vijayan, P. & Warkentin, T. Genetic control and QTL analysis of cotyledon bleaching resistance in green field pea (*Pisum sativum* L.). *Genome* **53**, 346-359 (2010). <https://doi.org/10.1139/G10-013>
- 40 Weldon, W. F. R. MENDEL'S LAWS OF ALTERNATIVE INHERITANCE IN PEAS. *Biometrika* **1**, 228-233 (1902). <https://doi.org/10.1093/biomet/1.2.228>
- 41 Bateson, W. *Mendel's Principles of Heredity, a Defence*. (Cambridge University Press, 1902).
- 42 Ellessawy, F. M., Bazghaleh, N., Vandenberg, A. & Purves, R. W. Polyphenol profile comparisons of seed coats of five pulse crops using a semi-quantitative liquid chromatography-mass spectrometric method. *Phytochem. Anal.* **31**, 458-471 (2020). <https://doi.org/10.1002/pca.2909>
- 43 Yoon, S. & Klein, B. P. Some properties of pea lipoxygenase isoenzymes. *J. Agric. Food. Chem.* **27**, 955-962 (1979). <https://doi.org/10.1021/jf60225a026>
- 44 Domoney, C., Casey, R., Turner, L. & Ellis, N. *Pisum* lipoxygenase genes. *Theor. Appl. Genet.* **81**, 800-805 (1991). <https://doi.org/10.1007/BF00224993>
- 45 Buckle, K. A. & Edwards, R. A. Chlorophyll degradation and lipid oxidation in frozen unblanched peas. *J. Sci. Food Agric.* **21**, 307-312 (1970). <https://doi.org/10.1002/jsfa.2740210609>
- 46 Delmas, F. *et al.* ABI3 controls embryo degreening through Mendel's *i>I</i>* locus. *Proc. Natl. Acad. Sci.* **110**, E3888-E3894 (2013). <https://doi.org/10.1073/pnas.1308114110>
- 47 Le Signor, C. *et al.* Genome-wide association studies with proteomics data reveal genes important for synthesis, transport and packaging of globulins in legume seeds. *New Phytol.* **214**, 1597-1613 (2017). <https://doi.org/10.1111/nph.14500>
- 48 Hellens, R. P. *et al.* Identification of Mendel's White Flower Character. *PLoS One* **5**, e13230-e13230 (2010). <https://doi.org/10.1371/journal.pone.0013230>
- 49 Harker, C. L., Ellis, T. H. & Coen, E. S. Identification and genetic regulation of the chalcone synthase multigene family in pea. *The Plant Cell* **2**, 185-194 (1990). <https://doi.org/10.1105/tpc.2.3.185>
- 50 Ellis, N. *et al.* Recombinant inbred lines derived from wide crosses in *Pisum*. *Sci. Rep.* **13**, 20408-20408 (2023). <https://doi.org/10.1038/s41598-023-47329-9>
- 51 Liu, Q., Wu, K., Harberd, N. P. & Fu, X. Green Revolution DELLAs: From translational reinitiation to future sustainable agriculture. *Mol. Plant* **14**, 547-549 (2021). <https://doi.org/10.1016/j.molp.2021.03.015>
- 52 Spielmeier, W., Ellis, M. H. & Chandler, P. M. Semidwarf sd-1, "green revolution" rice, contains a defective gibberellin 20-oxidase gene. *Proc. Natl. Acad. Sci.* **99**, 9043-9048 (2002). <https://doi.org/10.1073/pnas.132266399>
- 53 Weston, D. E. *et al.* The Pea DELLA Proteins LA and CRY Are Important Regulators of Gibberellin Synthesis and Root Growth. *Plant Physiol.* **147**, 199-205 (2008). <https://doi.org/10.1104/pp.108.115808>
- 54 Davidson, S. E., Smith, J. J., Helliwell, C. A., Poole, A. T. & Reid, J. B. The Pea Gene LH Encodes ent-Kaurene Oxidase. *Plant Physiol.* **134**, 1123-1134 (2004). <https://doi.org/10.1104/pp.103.032706>
- 55 Lester, D. R., Ross, J. J., Davies, P. J. & Reid, J. B. Mendel's stem length gene (*Le*) encodes a gibberellin 3 beta-hydroxylase. *The Plant Cell* **9**, 1435-1443 (1997). <https://doi.org/10.1105/tpc.9.8.1435>
- 56 Martin, D. N., Proebsting, W. M. & Hedden, P. Mendel's dwarfing gene: cDNAs from the *Le* alleles and function of the expressed proteins. *Proc. Natl. Acad. Sci.* **94**, 8907-8911 (1997). <https://doi.org/10.1073/pnas.94.16.8907>

- 57 Reid, J. B., Murfet, I. C. & Potts, W. C. Internode Length in *Pisum*. II. Additional Information on the Relationship and Action of Loci *Le*, *La*, *Cry*, *Na* and *Lm*. *J. Exp. Bot.* **34**, 349-364 (1983). <https://doi.org/10.1093/jxb/34.3.349>
- 58 Price, D. N. & Hedley, C. L. The effect of the *gp* gene on fruit development in *Pisum sativum* L. *New Phytol.* **110**, 271-277 (1988). <https://doi.org/10.1111/j.1469-8137.1988.tb00262.x>
- 59 White, O. E. in *II. The Present State of Knowledge of Heredity and Variation in Peas*. Vol. 56 487-588 (Proceedings of the American Philosophical Society, 1917).
- 60 Neff, D. I. & White, O. E. Inheritance studies in *Pisum*. VI Multiple allelomorphism and the inheritance of green and yellow foliage and pod color. *Am. J. Bot.* **14**, 379-394 (1927). <https://doi.org/10.1002/j.1537-2197.1927.tb04852.x>
- 61 Dalmais, M. *et al.* UTILLdb, a *Pisum sativum* in silico forward and reverse genetics tool. *Genome Biol.* **9**, R43-R43 (2008). <https://doi.org/10.1186/gb-2008-9-2-r43>
- 62 Ellis, T. H. N., Hofer, J. M. I., Swain, M. T. & van Dijk, P. J. Mendel's pea crosses: varieties, traits and statistics. *Hereditas* **156**, 33-33 (2019). <https://doi.org/10.1186/s41065-019-0111-y>
- 63 Zhang, P. *et al.* Fine mapping *PsPS1*, a gene controlling pod softness that defines market type in pea (*Pisum sativum*). *Plant Breed.* **141**, 418-428 (2022). <https://doi.org/10.1111/pbr.13020>
- 64 Karaca, D. E. *Molecular mapping of the V locus in pea (Pisum sativum L.)*, Washington State University, (2019).
- 65 Lamprecht, H. & Mrkos, H. The inheritance of the bracteole in *Pisum* and linkage of the gene *Br*. (1950).
- 66 Blixt, S. Mutation genetics in *Pisum*. *Agric. Hort. Genet.* **30**, 1-293 (1972).
- 67 Zhong, R. & Ye, Z.-H. IFL1, a Gene Regulating Interfascicular Fiber Differentiation in Arabidopsis, Encodes a Homeodomain-Leucine Zipper Protein. *The Plant Cell* **11**, 2139-2152 (1999). <https://doi.org/10.1105/tpc.11.11.2139>
- 68 Sussmilch, F. C., Ross, J. J. & Reid, J. B. Mendel: From genes to genome. *Plant Physiol.* **190**, 2103-2114 (2022). <https://doi.org/10.1093/plphys/kiac424>
- 69 Etchells, J. P., Smit, M. E., Gaudinier, A., Williams, C. J. & Brady, S. M. A brief history of the TDIF-PXY signalling module: balancing meristem identity and differentiation during vascular development. *New Phytol.* **209**, 474-484 (2016). <https://doi.org/10.1111/nph.13642>
- 70 Zhang, Y. *et al.* Transcription factor PagMYB31 positively regulates cambium activity and negatively regulates xylem development in poplar. *The Plant Cell* **36**, 1806-1828 (2024). <https://doi.org/10.1093/plcell/koae040>
- 71 Smit, M. E. *et al.* A PXY-Mediated Transcriptional Network Integrates Signaling Mechanisms to Control Vascular Development in Arabidopsis. *The Plant Cell* **32**, 319-335 (2020). <https://doi.org/10.1105/tpc.19.00562>
- 72 Yang, C. *et al.* Transcription Factor MYB26 Is Key to Spatial Specificity in Anther Secondary Thickening Formation. *Plant Physiol.* **175**, 333-350 (2017). <https://doi.org/10.1104/pp.17.00719>
- 73 Miyamoto, T., Tobimatsu, Y. & Umezawa, T. MYB-mediated regulation of lignin biosynthesis in grasses. *Curr. Plant Biol.* **24**, 100174-100174 (2020). <https://doi.org/10.1016/j.cpb.2020.100174>
- 74 Yang, C. *et al.* Arabidopsis MYB26/MALE STERILE35 Regulates Secondary Thickening in the Endothecium and Is Essential for Anther Dehiscence. *The Plant Cell* **19**, 534-548 (2007). <https://doi.org/10.1105/tpc.106.046391>
- 75 Parker, T. A., Lo, S. & Gepts, P. Pod shattering in grain legumes: emerging genetic and environment-related patterns. *The Plant Cell* **33**, 179-199 (2021). <https://doi.org/10.1093/plcell/koaa025>

- 76 Gerard, J. *The Herball, or, Generall historie of plantes*. (1597).
- 77 Sinjushin, A. & Gostimskii, S. Relationship between different fasciated lines of pea. *Pisum Genetics* **39**, 16-18 (2007).
- 78 Würschum, T., Groß-Hardt, R. & Laux, T. APETALA2 Regulates the Stem Cell Niche in the Arabidopsis Shoot Meristem. *The Plant Cell* **18**, 295-307 (2006).
<https://doi.org/10.1105/tpc.105.038398>
- 79 Krusell, L. *et al.* The Clavata2 genes of pea and *Lotus japonicus* affect autoregulation of nodulation. *The Plant Journal* **65**, 861-871 (2011). <https://doi.org/10.1111/j.1365-313X.2010.04474.x>
- 80 Sinjushin, A. A. & Gostimskii, S. A. Genetic control of fasciation in pea (*Pisum sativum* L.). *Russ. J. Genet.* **44**, 702-708 (2008). <https://doi.org/10.1134/S1022795408060100>
- 81 Osipova, M. A. *et al.* WUSCHEL-RELATED HOMEODOMAIN5 Gene Expression and Interaction of CLE Peptides with Components of the Systemic Control Add Two Pieces to the Puzzle of Autoregulation of Nodulation *Plant Physiol.* **158**, 1329-1341 (2012).
<https://doi.org/10.1104/pp.111.188078>
- 82 Fletcher, J. C. CLAVATA3 Signaling Buffers Arabidopsis Shoot Apical Meristem Activity in Response to Photoperiod. *Int. J. Mol. Sci.* **25**, 9357-9357 (2024).
<https://doi.org/10.3390/ijms25179357>
- 83 Hu, C. *et al.* A group of receptor kinases are essential for CLAVATA signalling to maintain stem cell homeostasis. *Nat. Plants* **4**, 205-211 (2018).
<https://doi.org/10.1038/s41477-018-0123-z>
- 84 Brand, U., Fletcher, J. C., Hobe, M., Meyerowitz, E. M. & Simon, R. d. Dependence of Stem Cell Fate in Arabidopsis on a Feedback Loop Regulated by CLV3 Activity. *Science* **289**, 617-619 (2000). <https://doi.org/10.1126/science.289.5479.617>
- 85 Fletcher, J. C., Brand, U., Running, M. P., Simon, R. d. & Meyerowitz, E. M. Signaling of Cell Fate Decisions by CLAVATA3 in Arabidopsis Shoot Meristems. *Science* **283**, 1911-1914 (1999). <https://doi.org/10.1126/science.283.5409.1911>
- 86 Schoof, H. *et al.* The Stem Cell Population of Arabidopsis Shoot Meristems Is Maintained by a Regulatory Loop between the CLAVATA and WUSCHEL Genes. *Cell* **100**, 635-644 (2000). [https://doi.org/10.1016/S0092-8674\(00\)80700-X](https://doi.org/10.1016/S0092-8674(00)80700-X)
- 87 Domoney, C. *et al.* Exploiting a fast neutron mutant genetic resource in *Pisum sativum* (pea) for functional genomics. *Funct. Plant Biol.* **40**, 1261-1261 (2013).
<https://doi.org/10.1071/FP13147>
- 88 Świącicki, W. K. Two new anthocyanin genes in the *d* segment of chromosome 1. *Pisum Newsletter* **22**, 56-61 (1990).
- 89 Sawada, C. *et al.* An Integrated Linkage Map of Three Recombinant Inbred Populations of Pea (*Pisum sativum* L.). *Genes* **13**, 196-196 (2022).
<https://doi.org/10.3390/genes13020196>
- 90 Nomoto, Y. *et al.* A hierarchical transcriptional network activates specific CDK inhibitors that regulate G2 to control cell size and number in Arabidopsis. *Nat. Commun.* **13**, 1660-1660 (2022). <https://doi.org/10.1038/s41467-022-29316-2>
- 91 Tourdot, E., Mauxion, J.-P., Gonzalez, N. & Chevalier, C. Endoreduplication in plant organogenesis: a means to boost fruit growth. *J. Exp. Bot.* **74**, 6269-6284 (2023).
<https://doi.org/10.1093/jxb/erad235>
- 92 Dominguez, I. P. *et al.* An economic assessment of GHG mitigation policy options for EU agriculture. *JRC Science for Policy Report* **27973** (2020).
- 93 Commission., E. *Food, Farming, Fisheries*. (2018).
- 94 Hylton, C. & Smith, A. M. The *rb* Mutation of Peas Causes Structural and Regulatory Changes in ADP Glucose Pyrophosphorylase from Developing Embryos. *Plant Physiol.* **99**, 1626-1634 (1992). <https://doi.org/10.1104/pp.99.4.1626>

- 95 Zhuang, L.-L. *et al.* LATHYROIDES, Encoding a WUSCHEL-Related Homeobox1 Transcription Factor, Controls Organ Lateral Growth, and Regulates Tendril and Dorsal Petal Identities in Garden Pea (*Pisum sativum* L.). *Mol. Plant* **5**, 1333-1345 (2012).
<https://doi.org/10.1093/mp/sss067>
- 96 Tayeh, N. *et al.* *afila*, the origin and nature of a major innovation in the history of pea breeding. *New Phytol.* **243**, 15 (2024).
<https://doi.org/https://doi.org/10.1111/nph.19800>