University of Reading

Flanking SNP markers for vicine– convicine concentration in faba bean (Vicia faba L).

Article

Accepted Version

www.reading.ac.uk/centaur

Flanking SNP markers for vicine-convicine content in faba bean (*Vicia faba L.*)

Hamid Khazaei[1*], Donal M. O'Sullivan[2], Huw Jones[3], Nicholas Pitts[3,4], Mikko J. Sillanpää[5], Pertti Pärssinen[6], Outi Manninen[6] and Frederick L. Stoddard[1]

[1] Department of Agricultural Sciences, P O Box 27 (Latokartanonkaari 5), FIN-00014 University of Helsinki, Finland

[2] School of Agriculture, Policy and Development, University of Reading, Whiteknights, P O Box 237, Reading, UK

[3] National Institute of Agricultural Botany, Huntingdon Rd, Cambridge CB3 0LE, UK

[4] School of Life Sciences, University of Manchester, Stopford Building, Oxford Rd, Manchester M13 9PT, UK

[5] Department of Mathematical Sciences, Department of Biology and Biocenter Oulu, P O Box 3000, FIN-90014 University of Oulu, Finland

[6] Boreal Plant Breeding Ltd, Myllytie 10, FIN-31600 Jokioinen, Finland

[*]E-mail:hamid.khazaei@helsinki.fi

Abstract

The pyrimidine glycosides, vicine and convicine, limit the use of faba bean (*Vicia faba*) as food and feed. A single recessive gene, *vc-*, is responsible for a lowered vicine-convicine content. The biosynthetic pathway of these closely related compounds is not known, and the nearest available markers are several cM away from *vc-*. Improved markers would assist breeding and help to identify candidate genes. A segregating population of 210 F5 recombinant inbred lines was developed from the cross of Mélodie/2 (low vicine-convicine) x ILB938/2 (normal vicine-convicine), and vicine-convicine contents were determined twice on each line. The population was genotyped with a set of 188 SNPs. A single QTL for vicine-convicine content was identified on Chromosome I, flanked by markers 1.0 cM away on one side and 2.6 cM on the other. The interval defined by these markers in the model species *Medicago truncatula* includes about 340 genes, but no candidate genes were identified. Further fine-mapping should lead to the

identification of tightly-linked markers as well as narrowing down the search for candidate regulatory or biosynthetic genes which could underlie the vc^- locus.

Introduction

The pyrimidine glycosides, vicine and convicine, are the first factors limiting the use of faba bean (*Vicia faba* L.) in food and feed (Duc et al. 1989; Crépon et al. 2010). Their aglycones, divicine and isouramil, are powerful oxidants and in humans with glucose-6-phosphate dehydrogenase deficiency, they cause an acute haemolytic anemia called favism, while in chickens, they cause a similar rupture of erythrocytes, especially in broiler breeds, while in egg-laying hens productivity is decreased (reviewed by Crépon et al. 2010). The biosynthetic pathway of vicine and convicine is not known, so there are no identified enzymes or enzyme genes that can be sought and used in a candidate gene approach. Wild-type faba bean cultivars contain 4 to 16 g/kg of vicine-convicine (Khamassi et al. 2013), but *vc-* genotypes contain 5-10% of this (Duc et al. 1999) and *vc-* cultivars have subsequently been released (Duc et al. 2004). The *vc-* gene has been localized to 10.1 cM from colourless hilum (Duc et al. 2004), but not all colourless-hilum accessions are *vc-*. The identification of closer markers has had to await the development of new technologies that can be applied efficiently to the very large (13 Gbp) genome of this species. Gutierrez et al. (2006) reported the development of two RAPD (random amplified polymorphic DNA) markers converted to CAP (cleavage amplified polymorphism) linked to low vicine-convicine content, but these markers were further from *vc-* than hilum colour.

Large-scale transcriptome data together with genomic DNA-derived markers based on single nucleotide polymorphism (SNPs) are now facilitating the development of cost-effective and highly saturated genetic maps (Saxena et al. 2012; Kaur et al. 2014). SNPs provide low

genotyping cost per data point, high genomic abundance (highly polymorphic), locus specificity (accurate and reproducible, Yan et al. 2010), bi-allelism, co-dominance, simple documentation, potential for high throughput analysis and common occurrence among elite germplasm (Cottage et al. 2012) so they have emerged as the ideal powerful tools for genomics analyses and molecular breeding for crop improvement (Semagn et al. 2013). A set of SNP markers has been developed for faba bean (Cottage et al. 2013, Khazaei et al. 2014), offering the potential to more closely identify chromosome regions containing valuable genes.

For these reasons, we applied the SNP set to a population of recombinant inbred lines segregating for vicine-convicine content. The main objective was to identify putative QTLs (quantitative trait loci) as well as SNP markers linked to low vicine-convicine content. Our secondary aim was exploration of haplotype diversity in the region linked to low VC in diverse genetic backgrounds.

Materials and Methods

Mapping population

The parents for the mapping population were Mélodie/2, an inbred line selected from the French colourless-hilum and low vicine-convicine content cultivar Mélodie, and ILB 938/2, an inbred line selected from the Ecuadorian landrace ILB938 with black hilum and high vicine-convicine content (Khamassi et al. 2013). A population of 210 F_5 recombinant inbred lines (RILs) was generated by single-seed descent from the cross of Mélodie/2 (female) and ILB 938/2 (male) in the insect-proof greenhouse of Department of Agricultural Sciences, University of Helsinki, Finland during 2009-2012. The hilum colour was recorded during population development (F_1–F_5) and followed normal Mendelian segregation (Supplementary Table 1), as expected.

Growth conditions

The parental lines (15 replicates from each) and the mapping population were evaluated in a climate-controlled glasshouse of the Department of Agricultural Sciences, University of Helsinki, Finland using a randomized complete block design in three replicates in 2013. Seeds of all plants were inoculated with *Rhizobium leguminosarum* biovar. *viciae* (faba bean strain, Elomestari Oy, Tornio, Finland) before sowing. Seeds were sown in 4 L plastic pots containing a mixture of sand (size: 0.5–1.2 mm, Saint-Gobain Weber Oy Ab, Helsinki, Finland) and peat (White 420 W, Kekkilä Oy, Vantaa, Finland) (2:1 *v/v*) containing all essential nutrients. Photoperiod was adjusted to 14 h light and 10 h dark, and the temperature was 20°C day/15°C night ± 2°C. Photosynthetic photon flux density (PPFD) was about 300 µmol m^{-2} s^{-1} at the canopy level. Soil moisture level was maintained at field capacity with automatic irrigation during the experiments.

Vicine - convicine analysis

Two methods were employed to measure the vicine-convicine content in the parental lines and the mapping population, one based on HPLC (high-performance liquid chromatography) and the other on spectrophotometry. The partners at NIAB used the method described by Khamassi et al. (2013), using both peak height and peak area for each of vicine and convicine on all 210 RILs, while those at Boreal followed Sixdenier et al. (1996) for total vicine-convicine on a subset of 161 RILs, owing to low seed quantity of the 49 other RILs. Modifications included using only one weighed and crushed seed per sample, incubating only 1 h 15 min at 90°C, and diluting the sample 1:10 before spectroscopy. Absorbances were converted to concentrations based on earlier calibration experiments, where HPLC was used as the reference method (Gutierrez et al. 2006).

SNP selection

A set of 188 polymorphic SNPs, listed in Khazaei et al. (2014), was used in this work. SNP

genotyping was carried out using the KASPar™ (Kompetitive Allele Specific PCR) assay

(KBioscience, UK) platform. Details on the KASPar principle, amplification of targeted region,

fluorescence detection and allele calling are available at

http://www.kbioscience.co.uk/reagents/KASP_manual.pdf.

Map construction and QTL mapping

The genetic map was originally constructed by Khazaei et al. (2014). The linkage map was

constructed by using MapDisto v. 1.7.7.0.1 (Lorieux 2012) with logarithm of odds (LOD) score

of 3.0 and recombination fraction of 0.3. The Kosambi function was used to calculate the map

distance in centimorgans (cM) (Kosambi 1943). To map the hilum colour locus, we included

binary phenotype status in both cases as an extra morphological marker to the molecular linkage

map construction process, scoring the recessive colourless-hilum phenotype as homozygous and

the dominant black HC phenotype using the relevant genotypic ambiguity code to denote

possible heterozygosity.

For the F_5 generation progeny, we calculated observed segregation ratio at each locus and

averaged their values over loci to obtain mean segregation ratios.

Single marker analysis, interval mapping (IM; Lander and Botstein 1989) [FLS1]and composite

interval mapping (CIM; Jansen 1993; Zeng 1993) were used to detect the relationship between

each linkage group and putative QTL locations of studied traits by Windows QTL Cartographer

v 2.5_011 (Wang et al. 2012). Significant QTLs were analysed with CIM. Cofactors were

determined using the forward and backward method in the standard CIM model with the

probability in and out of 0.1. The number of control markers and window size were set to 5 and 10, respectively. The genome was scanned at a walk speed of 1 cM intervals. The 95% significance threshold for QTL detection (at each trait) was determined by the phenotype permutation using 1,000 permutations at an experiment-wise $P < 0.05$ (Churchill and Doerge 1994). LOD profiles over each of the linkage groups and for each trait were drawn. LOD score value greater than the threshold was considered as evidence of a QTL in these profiles. Uncertainty of the QTL position around maximum LOD-score was indicated by 1-LOD and 2-LOD support intervals (Conneally et al. 1985; Van Ooijen 1992). These intervals were calculated for the CIM run with 5 co-factors. Support interval threshold values were determined on the LOD-score scale by subtracting one (two) LOD from the maximum LOD-score value. Linkage map and QTL positions and profile were drawn by MapChart v. 2.2 (Voorrips 2002).

VC region haplotype diversity

A set of 37 diverse inbred accessions (Supplementary Table 2) was genotyped with 29 SNP markers from the faba bean consensus map Webb et al., 2015 covering the most telomeric-proximal portion (0-27cM) of the short arm of faba bean chr I and spanning the entire qV-C confidence interval.

Results and Discussion

In all generations, the segregation ratio of colourless to coloured hilum followed the expected Mendelian ratio (Supplementary Table 1), indicating that selection of the RIL population was unbiased in the region of hilum colour and vicine-convicine content.

Vicine content, convicine content, and total vicine-convicine content of the RIL population, as determined by the two partners on separate seed samples, were very highly correlated with each

other (Table 1). This result confirms that the methods used were consistent, and that peak height and peak area delivered essentially the same information. It also shows that there is little segregation for the ratio of vicine to convicine in the progeny of a single cross, in contrast to the wide range in ratios found in a germplasm collection (Khamassi et al. 2013).

The phenotype of vicine-convicine content in the F_5 generation was clearly bimodal, with the first mode around zero and the second mode at around 6.5 g/kg (dry weight basis) (Figure 1).

The mean segregation ratios for each SNP locus were close to 1:1, confirming that there were few heterozygotes at the marker loci. A bimodal phenotypic distribution is expected to be seen in case of a single strong quantitative trait locus (QTL) in backcross population where only two genotypes are segregating (Lander and Botstein 1989). In this case, the assumption of normality is still considered valid, because QTL mapping methods often assume that the trait, conditionally on the effects of the QTL, follows a normal distribution. In other words this means that normality is assumed only after the QTL is first pre-corrected away. It is also essential to realize that the phenotypic distribution seen in the table is not spike-at-zero type of distribution where part of the trait values are zero (censored) (Broman 2003). One important difference is that there are many values around zero, not a single value.

QTL analysis using composite interval mapping showed a single, very strong peak for the individual and combined vicine and convicine contents, 1.81 cM from the end of the linkage group, with two tightly linked flanking SNPs on each side of the QTL and a further marker 0.8 cM away (Figure 2, Table 2). The markers enclose a zone of 3.6 cM in length, and are thus much closer to the key vicine-convicine gene than the hilum colour marker, in this case 5.3 cM distant. All of the markers, except one, are derived from *Medicago truncatula* chromosome II and

enclose a zone of about 340 genes in that model species

(http://www.medicagohapmap.org/?genome). None of those 340 genes is clearly associated with pyrimidine metabolism. The remaining marker is derived from *M. truncatula* chromosome I, again unassociated with known genes for pyrimidine metabolism. The interval is associated with the distal end of the short arm of Chromosome I of faba bean Webb et al., 2015.

The peak markers under the qV-C QTL may be of immediate use in selecting LVC progeny from crosses involving Mélodie/2 as the LVC source, but we were interested in determining the LVC-linked allele frequencies in wider germplasm to assess the broad applicability of these markers in wider germaplsm. To this end, 29 SNPs from the vc-hc region of chr I of the faba bean consensus map Webb et al., 2015 were genotyped across 37 diverse inbred lines, seven of which had known (mostly wild type) VC status as determined by Khamassi et al., 2013. This made it clear that all 8 SNPs which distinguish Mélodie/2 from ILB938/2 are relatively frequent in diverse germplasm, ranging from 18% to 81% for the Mélodie/2 allele (Supp Table 3).The co-segregating peak markers within the qV-C interval - Vf_Mt2g007780_001 and Vf_Mt2g008330_001 – which are the best candidates for marker-assisted selection of the LVC trait show Mélodie/2 allele frequencies of 76 and 65% respectively and thus are probably informative in a minority of crosses and are certainly not diagnostic for the trait. This is not greatly surprising as the trait originated from a mutagenesis programme (ref Duc) and the causative mutation is probably therefore not found in nature and will simply be associated with a common haplotype represented by the original mutagenized background in those breeding lines and varieties into which the trait has been deliberately or accidentially introgressed from the single mutant source. Selection for low vicine-convicine content in a breeding program is also easy to perform on phenotypic level using the crude and cheap spectroscopic method. However,

when using this method the seed is destroyed which means that measuring the vicine-convicine content in segregating breeding material is not straightforward. When developing pure lines by single seed descent, it would be useful to fix low vicine-convicine content in the early generations by testing the F_2 individuals with SNP markers. Selection of homozygous plants for further selfing would fix the useful trait in next generations.

Conclusion

A single QTL determining vicine-convicine content was identified on chromosome I of faba bean. Flanking markers were identified 1.0 cM upstream and 2.6 cM downstream of the QTL, but the available alleles were not diagnostic of low vicine-convicine content.

Acknowledgments

Literature cited

Arese, P., Bosia, A., Naitana, A., Gaetani, S., D'Aquino, M. and Gaetani, G.F. (1981) Effect of divicine and isouramil on red cell metabolism in normal and G6PD-deficient (Mediterranean variant) subjects. Possible role in the genesis of favism. Prog Clin Biol Res 55, 725-746.

Broman KW (2003) Mapping quantitative trait loci in the case of a spike in the phenotype distribution. Genetics 163:1169-1175

Churchill G, Doerge R (1994) Empirical threshold values for quantitative trait mapping. Genetics 138: 963–971

Conneally PM, Edwarks JH, Kidd KK, Lalouel J-M, Morton NE, Ott J, White R (1985) Report of the committee on methods of linkage analysis and reporting. Cytogenet Cell Genet 40:356–359Cottage A, Gostkiewicz K, Thomas JE, Borrows R, Torres AM, O'Sullivan DM (2012) Heterozygosity and diversity analysis using mapped SNPs in a faba bean inbreeding programme. Mol Breed 30:1799–1809

Crépon K, Marget P, Peyronnet C, Carrouée B, Arese P, Duc G (2010) Nutritional value of faba bean (*Vicia faba* L.) seeds for feed and food. Field Crops Res 115:329–339

Duc, G., Sixdenier, G., Lila, M. and Furstoss, V. (eds) (1989) Search of genetic variability for vicine and convicine content in Vicia faba L. A first report of a gene which codes for nearly zero-vicine and zero-convicine contents.: Wageningen Pers.

Duc, G., Marget, P., Esnault, R., Le Guen, J., Bastianelli, D. (1999). Genetic variability for feeding value of faba bean seeds (*Vicia faba* L.): comparative chemical composition of isogenics involving zero-tannin and zero-vicine genes. J. Agric. Sci., Camb. 133, 185-196.

Duc, G., Marget, P., Page, D., Domoney, C. (2004). Facile breeding markers to lower contents of vicine and convicine in faba bean seeds and trypsin inhibitors in pea seeds. 4. International Workshop on Antinutritional Factors in Legume Seeds and Oilseeds, 2004/03/08-10, Toledo (ESP). EEAP, European association for animal production (FRA), in: Musquiz, M., Hill, G.D., Cuadrado, M.M. & Burbano, C. (Eds.), Recent Advances of Research in Antinutritional Factors in Legume Seeds and Oilseeds. Wageningen Academic Publishers, Wageningen (Netherlands), pp. 281–285.

Gutierrez, N., Avila, C.M., Duc, G., Marget, P., Suso, M.J., Moreno, M.T. and Torres, A.M. (2006) CAPs markers to assist selection for low vicine and convicine contents in faba bean (Vicia faba L.). Theor Appl Genet 114, 59-66.

Jansen, R.C., 1993 Interval mapping of multiple quantitative trait loci. Genetics 135: 205-211.

Kaur, S., Kimber RBE, Cogan NOI, Materne M, Forster JW, Paull J (2014) SNP discovery and high-density genetic mapping in faba bean (*Vicia faba* L.) permits identification of QTLs for ascochyta blight resistance. Plant Sci 217-218: 47–55

Khamassi K, Ben Jeddi F, Hobbs D, Irigoyen J, Stoddard F, O'Sullivan DM, and Jones H (2013) A baseline study of vicine–convicine levels in faba bean (*Vicia faba* L.) germplasm. Plant Genetic Resources: Characterization and Utilization 11: 250–257

Khazaei, H., O'Sullivan, D.M., Sillanpää, M.J. & Stoddard, F.L. 2014. Using Medicago truncatula-derived SNPs and synteny to identify QTLs and candidate genes for stomatal traits on chromosome II in faba bean (*Vicia faba* L.). Theoretical and Applied Genetics.

Kosambi DD (1943) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Lander, E.S. and Botstein, D. (1989) Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121: 185-194.

Lorieux M (2012) MapDisto: fast and efficient computation of genetic linkage map. Mol Breed 30:1231–1235

Marquardt, R.R., D.S., M. and A.A., F. (1983) Purification and some properties of Vicine and Convicine isolated from faba bean (Vicia faba L.) protein concentrate. Journal of Agricultural and Food Chemistry 31, 839-844.

Saxena RK, Varma Penmetsa RV, Upadhyaya HD, Kumar A, Carrasquilla-Garcia N, Schlueter JA, Farmer A, Whaley AM, Sarma BK, May GD, Cook DR, Varshney RK (2012) Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeon pea and comparative mapping in legumes. DNA Res 19:449–461

Semagn K, Babu R, Hearne S, Olsen M (2014) Single nucleotide polymorphism genotyping using Kompetitive Allele Specific PCR (KASP): overview of the technology and its application in crop improvement. Mol Breed 33:1–14

Sixdenier G, Cassecuelle F, Guillaumin L, Duc G (1996) Rapid spectrophotometric method for reduction of vicine and convicine in faba bean seed. FABIS Newsletter 38/39: 42–44

van Ooijen JW (1992) Accuracy of mapping quantitative trait loci in autogamous species. Theor Appl Genet 84:803–811

Voorrips RE (2002) MapChart: Software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wang S, Basten CJ, Zeng Z-B (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Webb, A., Cottage, A., Wood, T., Khamassi, K., Hobbs, D., Gostkiewicz K., White M., Febrer, M., Rogers, J., Duc, G., Stoddard, F., Maalouf, F., Ogbonnaya, F., Link, W., Smith, P., Thomas, J.E. and O'Sullivan D, M. (2015) A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (Vicia faba L.) Plant Biotechnology Journal in press.

Yan J, Yang X, Shah T, Sanchez-Villeda H, Li J, Warburton ML, Zhou Y, Crouch JH, Xu Y (2010) High-throughput SNP genotyping with the GoldenGate assay in maize. Mol Breed 25:441–451

Zeng, Z.-B., 1993 Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. Proc. Natl. Acad. Sci. USA. 90: 10972-10976.

Figures and tables

Table 1. Correlations among vicine (V), convicine (C) and vicine-convicine (V-C) contents measured according to Khamassi et al (2013) and Sixdenier et al (1996) on the F_5 RIL population (n = 210) of Mélodie/2 × ILB 938/2.

	C_a	$V\text{-}C_a$	V_h	C_h	$V\text{-}C_h$	$V\text{-}C_s^{\dagger}$
V_a	0.984	0.944	0.988	0.862	0.945	0.868
C_a		0.869	0.862	0.988	0.945	0.894
$V\text{-}C_a$			0.932	0.960	0.983	0.909
V_h				0.859	0.947	0.855
C_h					0.978	0.889
$V\text{-}C_h$						0.907

a, calculated by area; *h*, calculated by height (Khamassi et al. 2013).

S, Sixdenier et al. (1996) method.

All values are significant at *P* < 0.001.

[†] n = 161

Table 2 Putative QTLs by composite interval mapping (CIM) for vicine-convicine content of the Mélodie/2 × ILB 938/2 RIL population.

Trait	QTL name[a]	Position (cM)	LOD score	LOD at P= 0.05[b]	R^2	Additive effect[c]
Vicine	qV	1.81	51.16	13.08	0.719	−0.193
Convicine	qC	1.81	57.62	13.16	0.706	−0.198
Vicine-Convicine	qV-C	1.81	60.20	13.28	0.765	−0.308

SNPs that are underlined show the nearest marker to the QTL.

[a] QTLs are abbreviated by the trait name.

[b] Permutation-based LOD threshold to determine significant QTL at P=0.05 with 1,000 times permutation.

R^2 represents the proportion of phenotypic variance explained by each QTL.

[c] Negative value shows the contribution of ILB 938/2 allele.

Table 3 Sequences of SNP markers linked to vicine-convicine content.

SNP Name	K BIO. ID	Sequence
Vf_Mt2g005900_001	1770605960	GCTATAAAGGCTGTTGCCACTCCAAATCCAGCTGTGGAATTGCCATTAACCG CTGAAAATGTAGAAACTGTGCTTGATGAAATTCG[W]CCGTATCTYATTTCTG ATGGTGGAAATGTTGCTCTACATGAAATCGATGGCAATGTTGTACGGTTGAA GTTACAAGGTGCTTGTGGATCGTGTCCGAGTT
Vf_Mt1g083330_001	1919331600	AGGAGGAGTTGCAGAGAACGAATCTGGAATTGGAATCTTACAAAGAGAAGA T[Y]AAAGAGAATCAGGAGAAAATTAAGCTCAATAAACAACTACCCTACCTCG TTGGTAATATTGTTGAG
Vf_Mt2g007220_001	1242748400	GAATTCACTGGGGACTATGATGCTCTTCAACCCTTTCCAGAGATTGATTTTGT TGACTTTTTCAACAGCCATCCCAAGCTGCTCAAGTTTGATAT[Y]CATGGAGCC ATGTTTGCAGCACTATGCCAGAAGAACAGTCTAAAACATGT
Vf_Mt2g010740_001	1182404129	AGGTACAATGTCAGAACAGAAATTGTCCATCAAATAAGCTTGTTTTTCCATTG ATGGCACTACTAGCAGGAATATTAGGCGGTGTTTTCGGANNNNAT[Y]GGAGG TGGAATGCTTATAAGTCCACTTCTTCTTCAAGTCGGAATAGCTCCTGAGGTAA
Vf_Mt2g010880_001	1873029962	GGATGGACTTGGTAGCTGGTCAACACTTTTCCAAAGGCACCGCTTCCTCTTAT CTGCTCTCGTGCTCCTTACTATTCTATGTACTGTTTATCTTTACTT[Y]GCTATC ACCATAGGAAGTGGCTCCTGTTCTGGATTGAGTGGAGCACAGAAAGCTTCAT GCCATATGGAGCTTGTCAAGGATTCTATTTCCAACGGCA

Figure 1 Frequency distributions of vicine-convicine content (determined by peak area) in 210 RILs derived from a cross between Mélodie/2 and ILB 938/2. Each parental value is the mean of 15 replicates. Skewness = 0.267, Kurtosis = –1.217.
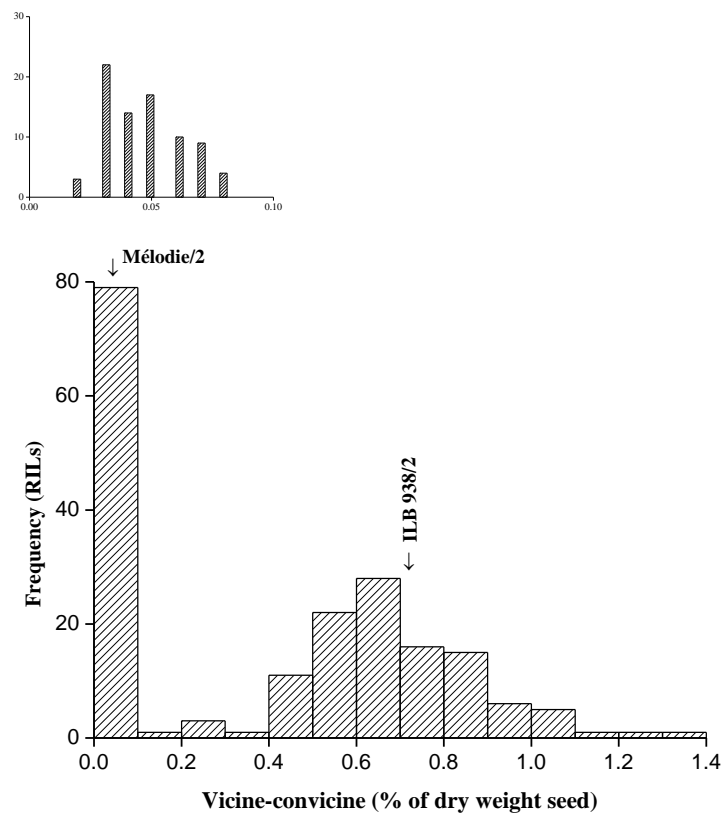
Figure 2 Comparison of LOD profiles of QTLs of vicine (v), convicine (c) and vicine-convicine (V-C) at Mélodie/2 × ILB 938/2 F_5 RILs obtained by composite interval mapping. HC, hilum colour. Recombinant fractions were converted to cM using the Kosambi mapping function. QTLs are represented by boxes extended by lines representing the LOD-1 and LOD-2 confidence intervals. HC, hilum colour
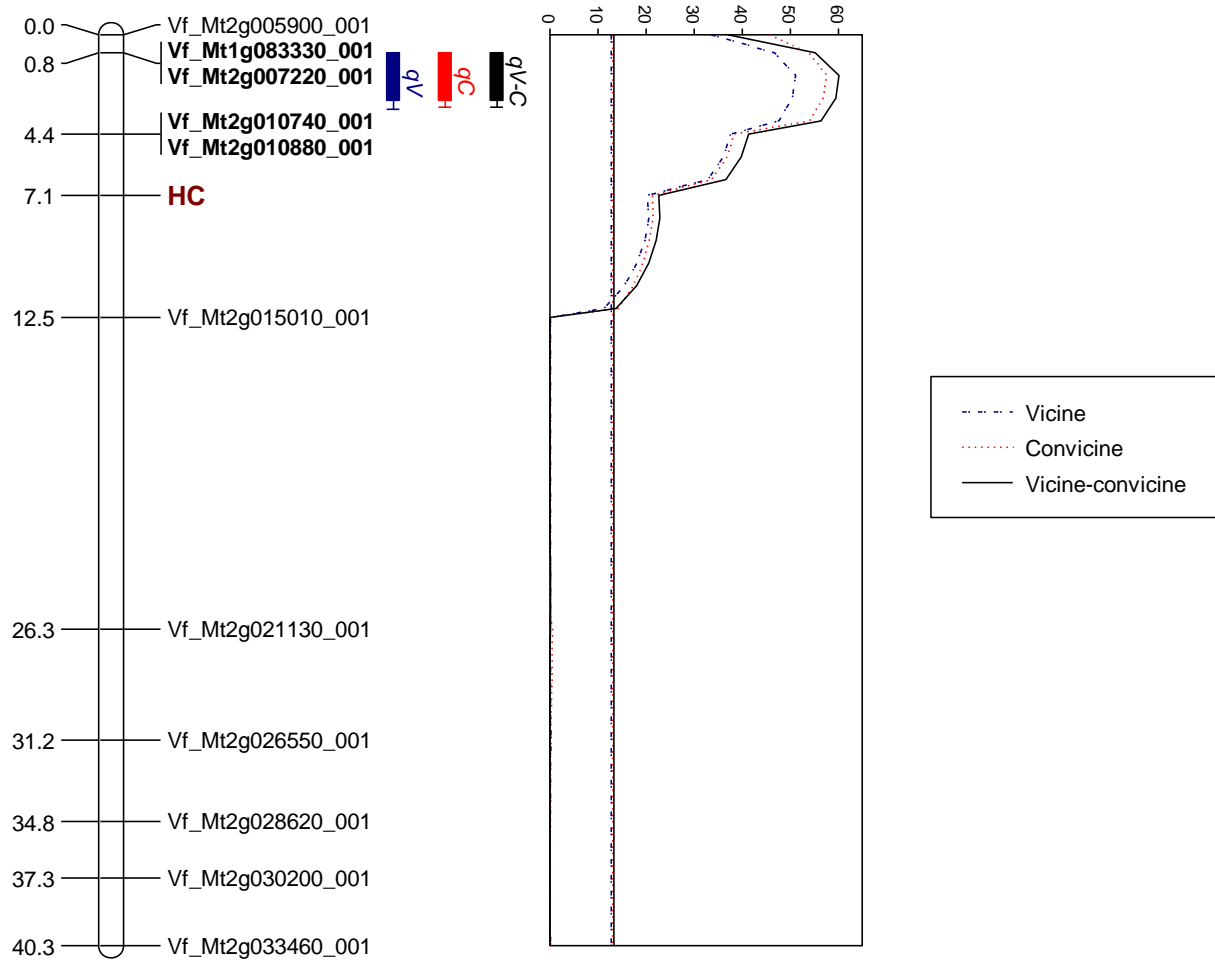
Supplementary Table 1 Segregation and χ^2 test for hilum colour amongst F_1, F_2, F_3, F_4 and F_5 generations in cross Mélodie/2 (colourless) × ILB 938/2 (black).

Generation	hilum colour		Segregation ratio	X^2	P
	Black	colourless			
F_1	30	0	**1:0**	–	–
F_2	260	88	**3:1**	0.015	0.901
F_3	209	122	**5:3**	0.058	0.809
F_4	172	130	**9:7**	0.061	0.805
F_5	112	98	**17:15**	0.004	0.952

Supplementary Table 2 Descriptive statistics for vicine-convicine content (measured by peak area).

Population / trait	Mélodie/2 Mean (SD)	ILB 938/2 Mean (SD)	F_5 Mean (SD)	Min	Max	Skewness	Kurtosis
Vicine	**0.00** (0.00)	**0.33** (0.00)	0.13 (0.13)	0.00	0.58	0.576	−0.432
Convicine	**0.03** (0.00)	**0.38** (0.01)	0.28 (0.23)	0.02	0.87	0.345	−1.132
Vicine-convicine	**0.04** (0.01)	**0.71** (0.01)	0.41 (0.35)	0.02	1.31	0.267	−1.127

SD, standard deviation

Supplementary Table 3. Haplotypes of 37 accessions in the vicine – hilum colour region of chromosome I.

Marker Name / 2013 Consensus Map Position>>>, cM	Vf_Mt2g005930_001 (0)	Vf_Mt2g005590_001 (0)	Vf_Mt2g005900_001 (0.41)	Vf_Mt2g006070_001 (0.41)	Vf_Mt2g007220_001 (1.29)	Vf_Mt1g083330_001 (1.29)	Vf_Mt2g007780_001 (1.31)	Vf_Mt2g007930_001 (1.72)	Vf_Mt2g008440_001 (2.13)	Vf_Mt2g008150_001 (2.13)	Vf_Mt2g008610_001 (2.93)	Vf_Mt2g009320_001 (3.76)	Vf_Mt2g010540_001 (4.52)	Vf_Mt2g010740_001 (4.9)	Vf_Mt2g010880_001 (5.27)	Vf_Mt2g010980_001 (5.27)	Vf_Mt2g010970_001 (5.9)	Vf_Mt2g011080_001 (5.9)	Vf_Mt2g012470_001 (8.31)	Vf_Mt2g012810_001 (9.37)	Vf_Mt2g013220_001 (10.53)	Vf_Mt2g013690_001 (11.28)	Vf_Mt2g013900_001 (12.4)	Vf_Mt2g015010_001 (14.59)	Vf_Mt2g015370_001 (14.97)	Vf_Mt2g015390_001 (14.97)	Vf_Mt2g019790_001 (26.08)	Vf_Mt2g020160_001 (26.46)	Vf_Mt2g021130_001 (27.23)
NV734_1_ILB938	C	G	T	T	T	C	A	C	A	T	T	T	T	C	T	T	T	C	G	C	G	G	C	G	A	A	G	T	A
NV735_1_MELODIE	C	G	A	T	C	T	A	C	A	T	C	A	T	T	C	T	T	C	G	C	G	G	C	T	A	A	G	T	G
NV662_1_VF136	C	G	T	T	C	C	A	C	A	T	C	A	T	C	C	T	T	G	A	C	G	G	C	T	A	A	G	T	G
NV644_1_KASZTEL	C	A	T	G	C	T	G	T	G	C	C	A	T	C	H	T	T	G	G	T	G	A	C	H	H	A	G	T	G
NV729_1_VF6	C	G	T	T	C	C	A	C	G	T	C	T	T	T	C	T	T	C	G	A	C	A	G	T	C	A	G	T	A
NV728_1_TW	C	G	A	T	C	T	A	T	A	T	C	T	T	T	C	T	T	C	A	C	G	G	C	T	A	A	G	T	G
NV730_1_I4347_4	C	A	T	T	T	T	A	C	A	T	C	T	C	C	C	T	C	G	A	C	G	G	C	T	A	A	G	T	G
NV731_1_I4347_3	C	A	T	T	T	T	A	C	A	T	C	T	C	C	C	T	C	G	A	C	G	G	C	T	A	A	G	T	G
NV643_3_S2_ALBUS	C	A	T	G	C	H	H	H	G	?	H	A	T	C	T	T	T	G	G	T	G	A	C	G	C	A	A	T	A
NV739-1_CRB2702	C	G	A	T	C	H	A	C	A	T	T	A	C	T	T	T	T	G	A	C	?	G	C	H	A	A	G	T	G
NV656_3_S2A2_9_BPL183	C	G	T	T	C	T	A	C	A	T	T	A	T	T	?	T	C	G	G	C	H	G	C	T	A	A	G	T	G
NV722_1_BANNER	C	G	A	T	C	T	A	C	A	T	T	A	T	C	C	T	T	C	A	C	G	G	C	T	A	A	G	T	G
NV648_2_S3B2_7_BPL10	A	G	A	T	T	?	A	C	A	T	T	T	C	T	C	C	C	C	A	C	A	G	T	T	A	G	G	C	G
NV714_1_HIVERNA	C	G	T	T	C	C	A	T	A	T	T	T	T	T	T	C	T	?	C	A	C	G	C	G	A	A	G	T	A
NV718_1_L977_88	C	G	A	T	C	C	A	C	A	T	T	T	T	C	T	C	T	C	G	A	C	A	G	C	A	A	G	T	A
NV720_1_BOURDON	C	A	T	T	C	C	A	C	A	T	T	T	T	T	C	T	T	G	A	C	G	G	C	T	C	A	A	T	G
NV726_1_GIZA402	C	G	T	T	C	C	A	C	A	T	T	T	T	T	C	C	C	C	G	T	G	A	C	T	A	A	G	T	G
NV727_1_BPL4628	C	G	T	T	C	C	A	C	A	T	T	T	T	T	C	T	C	C	C	G	T	H	G	T	A	A	G	T	G
NV732_1_BPL710	C	G	?	T	C	C	A	C	A	T	T	T	T	C	C	T	C	C	G	C	G	G	C	T	A	A	G	T	G

NV865-1_WEBO	C	G	T	T	C	H	A	C	A	C	T	T	T	T	C	T	T	G	A	C	G	G	T	T	A	A	A	T	A
NV639_1_HEDIN	C	A	A	T	C	T	A	C	A	C	T	T	T	T	C	T	T	G	A	C	G	G	C	T	A	A	G	T	A
NV736_1_AURORA	C	G	T	T	C	T	?	C	A	C	T	T	T	T	C	T	C	G	A	C	G	G	C	T	A	A	G	T	?
NV153_1DW_S3A2	C	G	A	T	C	T	A	C	A	T	T	T	T	T	C	C	T	C	G	A	T	G	G	C	T	A	A	G	T	G
NV657_26_INRA29H	C	G	T	T	T	T	A	C	A	T	T	T	T	T	T	C	T	C	C	A	C	G	G	T	T	A	A	G	T	A
NV658_2_CGN0771	C	G	A	T	C	T	?	C	A	T	T	T	T	T	C	T	C	G	A	T	G	G	C	T	A	A	G	T	G
NV713_1_COTEDOR	C	A	A	T	C	T	G	C	A	T	T	T	T	T	C	C	T	T	C	A	C	H	G	C	T	A	A	G	T	G
NV716_1_WIBO	C	A	A	T	C	T	G	C	A	T	T	T	T	T	C	C	T	T	C	A	C	H	G	C	T	A	A	G	T	A
NV717_1_L79_79	C	G	T	T	C	T	G	C	A	T	T	T	T	T	C	T	T	C	A	C	G	G	C	T	A	A	G	C	G
NV719_1_L979_S1	C	G	A	T	T	T	A	C	A	T	T	T	T	T	C	T	T	G	G	C	G	G	C	G	A	A	G	T	A
NV721_1_ARRISOT	C	G	T	T	T	T	A	C	A	T	T	T	T	T	C	C	T	C	C	A	C	G	G	T	G	A	A	G	T	A
NV723_1_BULLDOG	C	A	T	T	C	T	A	C	A	T	T	T	T	C	C	T	C	C	A	C	G	G	C	T	A	A	G	T	A
NV724_1_PIETRAL	C	G	T	T	C	T	?	C	A	T	T	T	T	T	T	T	C	G	G	C	H	G	T	T	A	A	G	T	G
NV725_1_GIZA3_2	C	G	A	T	C	T	A	C	A	T	T	T	T	T	C	T	C	C	A	C	A	G	T	T	A	A	G	T	G
NV733_1_NA112	C	G	A	T	C	T	A	C	A	T	T	T	T	T	C	T	C	C	A	T	H	G	C	T	A	A	G	T	G
NV737-1_CRB285	C	A	T	T	C	T	A	C	A	T	T	T	T	T	H	T	C	G	A	T	A	G	C	T	A	A	G	T	?
NV738-1_CRB2516	C	G	T	T	T	T	A	C	A	T	T	T	T	T	C	T	C	C	A	C	G	G	C	G	A	A	G	T	A
NV740-1_C100107	C	G	T	T	T	T	?	C	A	T	T	T	?	T	C	T	C	G	A	C	G	G	C	T	?	A	G	T	G