



**HAL**  
open science

## Development of new genetic resources for faba bean (*Vicia faba* L.) breeding through the discovery of gene-based SNP markers and the construction of a high-density consensus map

Estefania Carrillo Perdomo, A. Vidal, J. Kreplak, H. Duborjal, M. Leveugle,  
J. Duarte, C. Desmetz, C. Deulvot, B. Raffiot, P. Marget, et al.

### ► To cite this version:

Estefania Carrillo Perdomo, A. Vidal, J. Kreplak, H. Duborjal, M. Leveugle, et al.. Development of new genetic resources for faba bean (*Vicia faba* L.) breeding through the discovery of gene-based SNP markers and the construction of a high-density consensus map. *Scientific Reports*, 2020, 10 (1), pp.6790. 10.1038/s41598-020-63664-7 . hal-03672081

**HAL Id: hal-03672081**

**<https://hal.inrae.fr/hal-03672081>**

Submitted on 19 May 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



OPEN

# Development of new genetic resources for faba bean (*Vicia faba* L.) breeding through the discovery of gene-based SNP markers and the construction of a high-density consensus map

E. Carrillo-Perdomo<sup>1</sup>✉, A. Vidal<sup>2</sup>, J. Kreplak<sup>1</sup>, H. Duborjal<sup>3</sup>, M. Leveugle<sup>3</sup>, J. Duarte<sup>3</sup>, C. Desmetz<sup>1</sup>, C. Deulvot<sup>1</sup>, B. Raffiot<sup>4</sup>, P. Marget<sup>1</sup>, N. Tayeh<sup>1</sup>, J. P. Pichon<sup>3</sup>, M. Falque<sup>2</sup>, O. C. Martin<sup>2</sup>, J. Burstin<sup>1</sup> & G. Aubert<sup>1</sup>

Faba bean (*Vicia faba* L.) is a pulse crop of high nutritional value and high importance for sustainable agriculture and soil protection. With the objective of identifying gene-based SNPs, transcriptome sequencing was performed in order to reduce faba bean genome complexity. A set of 1,819 gene-based SNP markers polymorphic in three recombinant line populations was selected to enable the construction of a high-density consensus genetic map encompassing 1,728 markers well distributed in six linkage groups and spanning 1,547.71 cM with an average inter-marker distance of 0.89 cM. Orthology-based comparison of the faba bean consensus map with legume genome assemblies highlighted synteny patterns that partly reflected the phylogenetic relationships among species. Solid blocks of macrosynteny were observed between faba bean and the most closely-related sequenced legume species such as pea, barrel medic or chickpea. Numerous blocks could also be identified in more divergent species such as common bean or cowpea. The genetic tools developed in this work can be used in association mapping, genetic diversity, linkage disequilibrium or comparative genomics and provide a backbone for map-based cloning. This will make the identification of candidate genes of interest more efficient and will accelerate marker-assisted selection (MAS) and genomic-assisted breeding (GAB) in faba bean.

Legume crops serve as a source of food and feed. They also play an important role in sustainable agriculture because of their ability to improve soil fertility by fixing atmospheric nitrogen and increasing crop yield when used in crop rotation with cereals or intercropping<sup>1</sup>. In particular, faba bean (*Vicia faba* L.; Vf) is a primary ingredient of daily meals in both developing and industrialized countries due to its high content in proteins, carbohydrates, dietary fibers and micronutrients<sup>2,3</sup>. It is the most yielding pulse crop after field pea. However, its yield is still about half that of wheat, indicating that great breeding efforts are still needed<sup>4</sup>. Faba bean yield is greatly affected by environmental conditions, especially extreme temperatures, drought and acidity<sup>5,6</sup>. In addition, diseases such as chocolate spot (*Botrytis fabae* S. or *B. cinerea* P.) or ascochyta blight (*Ascochyta fabae* S.), viruses such as faba bean necrotic viruses, parasitic weeds of *Orobanchaceae* genus and pests such as leaf weevil (*Sitona lineatus* L.), aphids (*Aphis fabae* S., *A. craccivora* K., *Acyrtosiphon pisum* H., *Myzus persicae* S.) or seed weevils (*Bruchus rufimanus* B.) considerably reduce its yield and affect the commercialization of the grains<sup>5,7</sup>. Other factors limiting the production of faba bean include the overproduction of flowers resulting in a variable fertilization

<sup>1</sup>Agroécologie, AgroSup Dijon, INRAE, Univ. Bourgogne, Univ. Bourgogne Franche-Comté, F-21000, Dijon, France.

<sup>2</sup>Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE - Le Moulon, 91190, Gif-sur-Yvette, France. <sup>3</sup>Biogemma, Chappes, France. <sup>4</sup>Terres Inovia, Thiverval-Grignon, France. ✉e-mail: [estefania.carrillo-perdomo@inrae.fr](mailto:estefania.carrillo-perdomo@inrae.fr)

rate and abortion of ovules<sup>2</sup>, the need for pollinators for outcrossing and fertilization of ovules<sup>8</sup> and the strong influence of symbiosis for optimal concentration of nitrogen (N) in the grain and for N soil fertility<sup>9</sup>.

Faba bean is a diploid outcrossing species ( $2n = 12$ ) with a “giant genome”<sup>10</sup> of approximately 13 Gb distributed on six chromosomes. The high content in transposable elements<sup>11</sup> complexifies the faba bean genome assembly and map-based cloning. Most of the linkage maps generated so far have a low to medium saturation and are based on morphological, isozyme, restriction fragment length polymorphism (RFLP), random amplified polymorphic DNA (RAPD), sequence characterized amplified region (SCAR), intron targeted amplified polymorphism (ITAP), simple sequence repeat (SSR) and low-density single-nucleotide polymorphism (SNP) markers<sup>12–21</sup>. SNP-based genetic maps have been recently developed<sup>21</sup>. To date, the most saturated map is the one reported by Webb *et al.*<sup>21</sup> consisting of 687 SNPs. This map provided a first glimpse of synteny of faba bean with other legumes such as barrel medic (*Medicago truncatula* L.; Mt), lupine (*Lupinus albus* L.), soybean (*Glycine max* (L.) M.; Gm) or lentil (*Lens culinaris* M.)<sup>16,21</sup>. In fact, the SNPs used to construct the consensus map of Webb *et al.*<sup>21</sup> were designed based on orthologous sequences in *M. truncatula* with the objective of physically anchoring the faba bean consensus map to the Medicago genome. Webb *et al.*<sup>21</sup> also took advantage of the macrosynteny between these two species to discern the level of conservation of the genetic organization between faba bean and lentil, one of its most closely related crop species. A high conservation of genomic blocks between *V. faba* and these legume species was reported.

The development of dense and robust genetic maps that involve multiple populations and gene-based markers is a prerequisite for marker-assisted selection (MAS) and paves the way to *V. faba* genome assembly. Now that pulse genomes are becoming available, it is important to implement more accurate comparative genomic approaches that will reinforce faba bean breeding programs through a faster and more efficient identification of candidate genes. Transcriptome sequencing has been intensively used in the development of SNP markers for genetic mapping and diversity panel structuration of model species and crops with large genomes<sup>21–28</sup>. In addition, Illumina MiSeq<sup>29</sup> has made possible the identification, location and functional characterization of genes that control traits of interest and has been used to provide a more comprehensive view of diversity and gene function in plants<sup>30–32</sup>. The design of SNP markers is strongly recommended for the construction of genetic linkage maps, as they stand out for their uniform distribution throughout the genome, for being numerous and for their tendency to be biallelic and codominant<sup>33,34</sup>. Therefore, we chose to exploit next-generation sequencing (NGS) technologies and transcriptome sequencing to specifically address the expressed gene fraction (exome) for the discovery of gene-based SNPs. The objective of our work was to develop high-resolution genetic linkage maps in three interconnected faba bean recombinant line populations and build a high-density consensus map. Our work also took advantage of macrosynteny among legume relatives to locate faba bean genomic regions conserved in different sequenced species. A comparative alignment and mapping between the faba bean consensus map developed herein and the genomes of barrel medic, birdsfoot trefoil (*Lotus japonicus* L.; Lj), chickpea (*Cicer arietinum* L.; Ca), common bean (*Phaseolus vulgaris* L.; Pv), cowpea (*Vigna unguiculata* (L.) W.; Vu), pea (*Pisum sativum* L.; Ps) and soybean has been performed. The identification of highly syntenic and collinear areas between faba bean and already sequenced species will facilitate candidate gene discovery. In this way, SNP markers developed in genes resulting from the present work will be very useful in MAS and in map-based isolation of candidate genes.

## Materials and Methods

**Plant material.** Three connected bi-parental mapping populations have been built using cultivar (cv.) HIVERNA as a common female parent, and the accessions NOVA GRADISKA, SILIAN, and QUASAR as male parents for Pop1, Pop2 and Pop3, respectively. The winter type cv. HIVERNA of German origin was selected as a common parent between the three populations. The parents NOVA GRADISKA (originated in Croatia) and SILIAN (originated in Northern Sudan) are both minor type landraces sown in late and early spring, respectively; while QUASAR (originated in United Kingdom) is a winter type cultivar adapted to oceanic climate (cool winters with abundant rainfall). NOVA GRADISKA and QUASAR have previously been reported as resistant to the faba bean weevil (*Bruchus* spp.) attack<sup>35</sup>. All segregating recombinant populations used for mapping were made of F<sub>3</sub> plants produced by single seed descent (SSD) of F<sub>2</sub> plants. Pop1 includes 102 F<sub>3</sub> individuals, Pop2 147 F<sub>3</sub> plants derived from two F<sub>1</sub>s and Pop3 96 F<sub>3</sub> plants. Samples were collected from the leaves of an F<sub>3</sub> plant for each recombinant line. Tissues were flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until DNA extraction for genotyping.

For SNP discovery, the parental lines were grown in a growth chamber (photoperiod of 16 h light/day, 15 °C night, 20 °C day, hygrometry 60% min) for 15 days. Samples of whole plant per parental line were collected. Tissues were flash-frozen in liquid nitrogen and stored at  $-80^{\circ}\text{C}$  until RNA extraction.

**Transcript assembly and SNP Discovery.** SNP discovery was carried out as described by Duarte *et al.*<sup>28</sup> with slight modifications. Total RNAs were extracted from the parental lines and checked for quality and integrity. RNAs were then converted to full-length double-stranded cDNA and normalized with the Mint-2 and Trimmer-2 kits (Evrogen, Moscow, Russia), respectively. Later, q-RT-PCR assays were developed on a set of genes of different abundances to assess the efficiency of normalization. Normalized double-stranded cDNAs were sheared into 450 bp fragment size using the Covaris E220 system (Covaris Inc., Massachusetts, USA). Individual indexed NGS libraries were then produced with the SPRIworks HT reagent kit (Beckman Coulter, Indianapolis, USA). An equimolar pool of the four libraries was sequenced on the Illumina MiSeq platform (V2 chemistry, PE 2 × 250 nt, 12 million of clusters) (Illumina, California, USA). FastQC was used to check raw data quality. Then, sequencing adaptor removal and quality trimming were performed using trim\_galore and SMART oligos (normalization primers) that were masked in the raw sequence readings using an in house script. MIRA 4.0 (<http://mira-assembler.sourceforge.net>) software was used to perform a *de novo* assembly of all samples at once. Reads of parental genotypes were remapped on the assembly using BWA<sup>36</sup> and only contigs with more than

10x coverage were kept for further analysis. Transdecoder<sup>37</sup> was used to predict open reading frames in contigs. BUSCO v3.0.1<sup>38</sup> was used on viridiplantae odb10 database to assess completeness of the dataset. Functional annotation was performed using eggNOG-mapper<sup>39</sup> on eggNOG database 5.0<sup>40</sup>.

The SNP discovery was then performed with SAMtools mpileup<sup>36</sup> and BCFtools call<sup>41</sup>. Only homozygous SNPs were retained.

**Genotyping F<sub>3</sub> populations.** DNA samples were extracted from leaf tissues of the parents and the individuals of the three populations using the NucleoSpin Plant II Mini kit (<https://www.mn-net.com/>, Hoerd, France) following the manufacturer's protocol. DNAs were normalized before being fragmented with Adaptive Focused Acoustics Technology (Covaris Inc., Massachusetts, USA). A 250 bp target size value was obtained by using a Covaris E220 system, according to the manufacturer's instructions. Then DNA fragments underwent a NGS library preparation procedure consisting in end repair and Illumina adaptor ligation using the KAPA HTP kit (Roche, Basel, Switzerland). Individual index sequences were added to each library for identifying reads and sorting them according to their initial origin. Two thousand SNPs were targeted to design capture probes. The Sequence Capture was done by using SeqCap EZ Developer kit from Roche according to the manufacturer's instructions. The sequence capture reaction efficiency was evaluated by measuring, with a qPCR assessment, a relative fold enrichment and loss of respectively targeted and non-targeted regions before and after the sequence capture reaction.

The captured samples were sequenced on HiSeq sequencing platform (Illumina, California, USA) with a Paired End sequencing strategy of 2 reads of 100 bases. The objective was to produce around 3 Million sequencing clusters per sample.

Raw reads were trimmed for adaptor sequence using cutadapt 1.8.3<sup>42</sup> and then aligned on the targeted regions with Novoalign (<http://www.novocraft.com>, Selangor, Malaysia). Genotype at each position of interest was determined using SAMtools mpileup<sup>36</sup> and in house Perl scripts to filter out low quality positions, call the SNPs, and to produce a genotyping matrix for all 2000 selected markers.

**Genetic maps construction for each population.** Markers were first filtered for deviation from Mendelian segregation (37.5:25:37.5) using the following index  $Distortion = \frac{1}{2} \cdot \frac{\sum_{i=1}^{N_{genotypicclasses}} |f_{i,observed} - f_{i,expected}|}{1 - \min(f_{expected})}$

with threshold of 0.8. Then, markers were assigned to linkage groups by estimating the pairwise recombination frequencies within each population using the maximum likelihood procedure with the forward-backward algorithm and a LOD score of 5.0 as the threshold for significant linkage in the software JoinMap V5.0<sup>43</sup>. In a second stage, markers were ordered and assigned to their positions in each linkage group by likelihood maximization. Candidate orders' likelihoods were computed as for F<sub>2</sub> populations using Spell-QTL Bayesian inference<sup>44</sup> and heuristic local maximization was performed using custom R scripts implementing the algorithm described earlier in Ganai *et al.*<sup>45</sup>. This heuristic was based on the serial inference of marker orders, proceeding according to the following steps. First, 10 "seed" markers were randomly chosen in each chromosome, and one statistically robust scaffold map replicate was constructed from each seed marker, by iteratively choosing the most strongly linked neighbour with at least 10 cM between adjacent markers. Then given these scaffold maps, marker density was increased to produce framework maps containing as many markers as possible while keeping a LOD score >3.0 for the robustness of marker orders. Finally, the complete maps were obtained by placement of additional markers using bin-mapping<sup>45</sup>. As a post-processing step, it was necessary to calculate the distance between pairs of markers taking into account that the recombination rates estimated by the algorithm assumed that the populations were F<sub>2</sub> whereas they were in fact F<sub>3</sub>. For that, we first determined the correspondence between rF<sub>2</sub>, the recombination rate estimated assuming an F<sub>2</sub> population, and the true recombination rate, *r*. That correspondence was obtained in two steps. In the first step we used the explicit formula relating rF<sub>2</sub> to the two-locus genotype frequencies  $f_{AaBb}$ ,  $f_{AaBB} = f_{AABb}$ ,  $f_{AABB} = faabb$ , and  $f_{AAbb} = faaBB$  where the two parental types are denoted A (respectively a) and B (respectively b) and the labels refer to unphased genotypes. Then in the second step we inferred *r* from those 7 frequencies by maximum likelihood using the formulas for those frequencies in F<sub>3</sub> as a function of *r*. Finally, centiMorgan (cM) distances were calculated using Haldane's mapping function<sup>46</sup>.

**Consensus genetic map construction.** SNP sequences anchored to the sequences of the SNPs previously mapped by Webb *et al.*<sup>21</sup> were used to assign each linkage group (defined by the initial seed used in the map construction) to one chromosome number. For each chromosome, we produce a consensus map that summarizes the recombination information within the three populations. Our method is quite general and in particular it does not assume collinearity of the individual maps. Its key feature is the minimization of an index *I<sub>D</sub>* which measures the differences between the (unknown) consensus map *M\** and all of the individual maps *M1*, *M2*. We forced our consensus map to include all the markers present in the individual maps. *M\** is to be specified by assigning a genetic position for each marker and of course that will also define the marker ordering in *M\**. Without loss of generality, the genetic position of the first marker of *M\** can be considered as the origin of genetic coordinates. The task is then to estimate the genetic positions of the rest of the markers.

In our framework based on the minimization of the index *I<sub>D</sub>*, we first define a "distance" between two maps via the formula:

$$D_{(M,M')} = \sum_i \sum_{j(i)} \left( d_{(M(i,j))} - d_{(M'(i,j))} \right)^2$$

where *i* labels the markers that are common in the two maps and *j(i)* labels the markers that are not only in common in the two maps but also meet a criterion for their distance to marker *i*. Specifically, if marker *j(i)* is quite far

Map	LG	Markers	Distance (cM)	Unique positions	Average marker density (marker/cM)	Average gap size (cM)	Gaps >10 cM	Biggest gap size (cM)
<b>Pop1<sup>a</sup></b>								
1		401	444.99	360	0.90	1.24	0	9.51
2		228	262.44	191	0.87	1.38	2	18.94
3		226	194.59	199	1.16	0.98	0	7.90
4		180	192.88	150	0.93	1.29	0	6.92
5		164	176.20	136	0.93	1.31	1	13.05
6		239	154.81	211	1.54	0.74	0	7.26
<i>Total</i>		1438	1425.91	1247	1.01	1.16	3	18.94
<b>Pop2<sup>b</sup></b>								
1		376	555.35	367	0.68	1.52	2	13.66
2		219	278.87	204	0.79	1.37	0	7.80
3		231	315.56	221	0.73	1.43	1	19.27
4		168	247.25	153	0.68	1.62	2	12.98
5		152	222.56	147	0.68	1.52	2	11.20
6		166	212.34	160	0.78	1.34	1	20.78
<i>Total</i>		1312	1831.93	1252	0.72	1.47	8	20.78
<b>Pop3<sup>c</sup></b>								
1		384	493.68	350	0.78	1.41	2	12.08
2		275	312.25	257	0.88	1.22	2	20.82
3		223	242.73	201	0.92	1.21	2	12.59
4		168	223.30	147	0.75	1.53	0	8.50
5		141	208.45	126	0.68	1.67	2	12.45
6		215	216.13	190	0.99	1.14	0	8.51
<i>Total</i>		1406	1696.53	1271	0.83	1.36	8	20.82
<b>Consensus</b>								
1		467	481.74	375	0.97	1.29	3	18.44
2		334	252.46	219	1.32	1.16	2	17.05
3		271	216.51	231	1.20	0.88	3	11.03
4		220	204.06	168	1.08	1.22	0	13.65
5		183	196.69	141	0.93	1.4	0	6.98
6		253	168.94	156	1.35	1.09	1	20.51
<i>Total</i>		1728	1520.40	1290	1.12	1.17	9	20.51

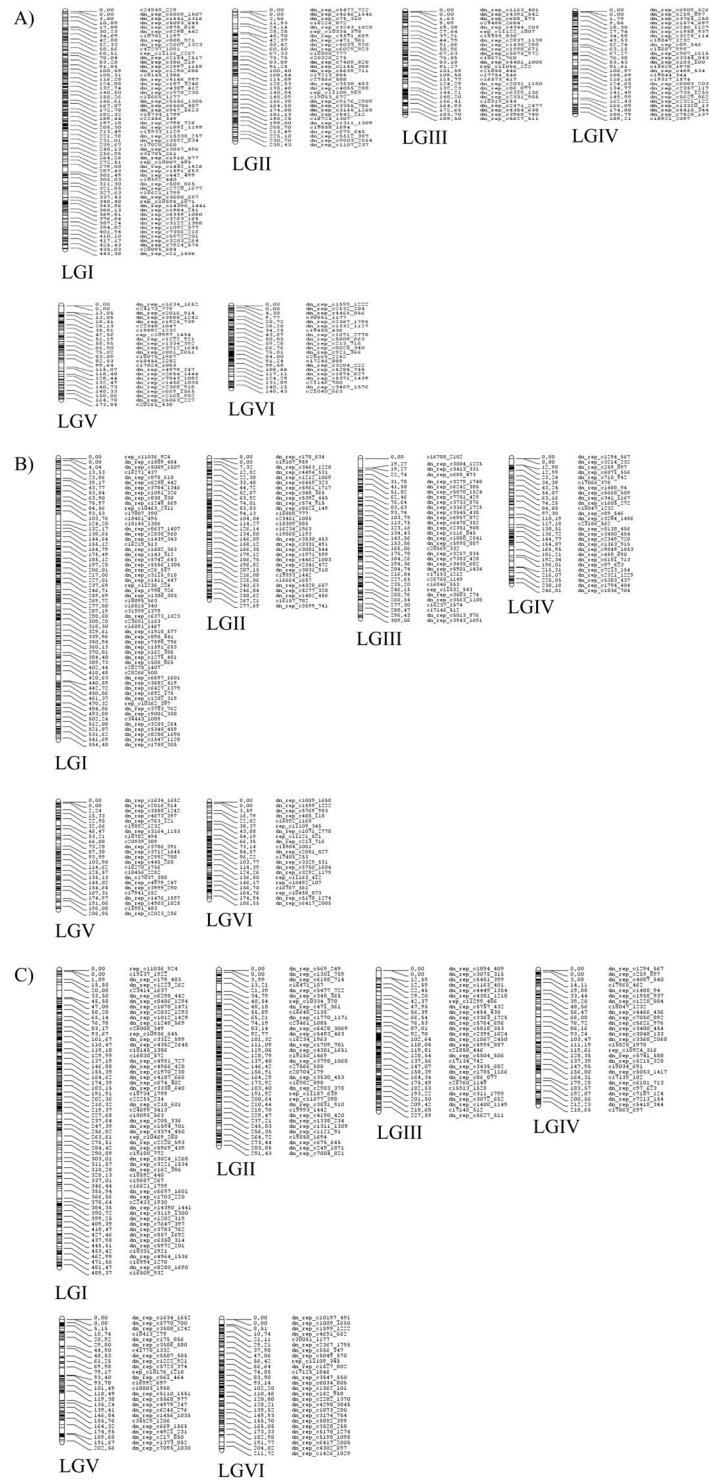
**Table 1.** Map features of the three individual genetic linkage maps and the consensus map of faba bean presented in this study. <sup>a</sup>Recombinant population derived from the cross Nova Gradiska × Hiverna. <sup>b</sup>Recombinant population derived from the cross Silian × Hiverna. <sup>c</sup>Recombinant population derived from the cross Quasar × Hiverna.

from marker  $i$ , that corresponding interval does not provide much information, so it is better to exclude it from the sum. In general, there are many of these markers and thus it is also computationally efficient to exclude them. Inversely, if a marker  $j(i)$  is very close to  $i$ , the corresponding distance is often not very well determined and so again it is better to exclude it. Therefore, our criterion imposes both a minimum and maximum distance between  $j(i)$  and  $i$ . In addition, to avoid having many  $j(i)$  markers for some  $i$  and only a few for others, we keep only a subset of the possible markers  $j(i)$  for a given  $i$ . As a result, all the  $i$ s are treated on an equal footing and have the same importance in  $D(M, M')$ . Given this definition for the distance  $D(M, M')$  between two arbitrary maps, we numerically search for the positions of the markers in  $M^*$  to minimize the index

$$I_D = \sum_n D_{(M^*, M_n)}$$

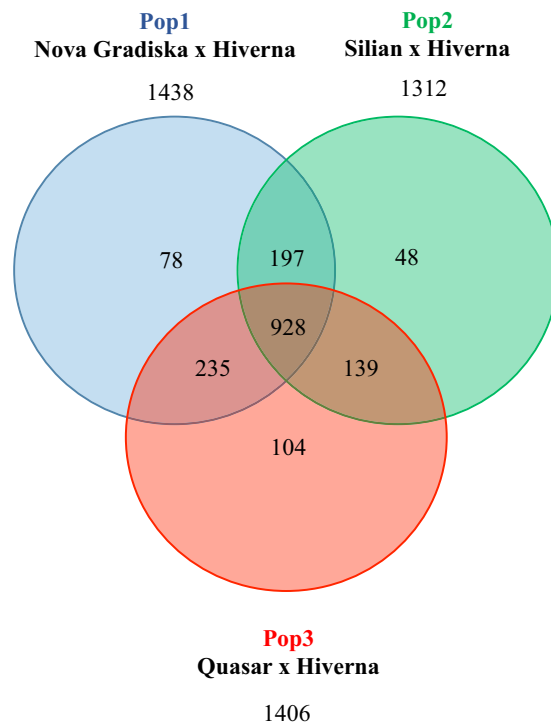
where  $n$  runs through all the individual maps from which the consensus is being built. If the maps have been determined using very different population sizes, we weight each term of this sum by the corresponding population size. In this way, if a map has a much smaller population than the other maps, it will have little influence on the construction of  $M^*$  which is justified since its marker positions are not very accurate. On the contrary, if a map has a much larger population than the other maps, it will have a strong influence on the construction of  $M^*$ , which is again justified since its marker positions are rather precisely determined.

**Synteny with other legume crops.** The flanking sequences of the SNP markers placed in the consensus map were searched against *C. arietinum* (v1)<sup>47</sup> ([https://www.ncbi.nlm.nih.gov/assembly/GCF\\_000331145.1/](https://www.ncbi.nlm.nih.gov/assembly/GCF_000331145.1/)),



**Figure 1.** Faba bean individual genetic linkage maps constructed from three populations of (A) 102 F<sub>3</sub> recombinant lines derived from the cross between *Vicia faba* cv. Nova Gradiska and Hiverna; (B) 147 F<sub>3</sub> derived from the cross between cv. Silian and Hiverna and (C) 96 F<sub>3</sub> derived from the cross between cv. Quasar and Hiverna.

*P. sativum* (v.1)<sup>48</sup> (<https://urgi.versailles.inra.fr/Species/Pisum/Pea-Genome-project>), *M. truncatula* (Mt5.0)<sup>49</sup> (<https://medicago.toulouse.inra.fr/MtrunA17r5.0-ANR/>), *L. japonicus* (v 2.5)<sup>50</sup> (<http://www.kazusa.or.jp/lotus/>), *G. max* (v2.0)<sup>51</sup> (<http://www.plantgdb.org/GmGDB/>), *Phaseolus vulgaris* (v2.1)<sup>52</sup> ([https://phytozome.jgi.doe.gov/pz/portal.html#linfo?alias=Org\\_Pvulgaris](https://phytozome.jgi.doe.gov/pz/portal.html#linfo?alias=Org_Pvulgaris)) and *Vigna unguiculata* (v1.1)<sup>53</sup> ([https://phytozome.jgi.doe.gov/pz/portal.html#linfo?alias=Org\\_Vunguiculata\\_er](https://phytozome.jgi.doe.gov/pz/portal.html#linfo?alias=Org_Vunguiculata_er)) genomes assemblies using the BLASTn function in order to identify the corresponding orthologous genes and their position in those genomes.



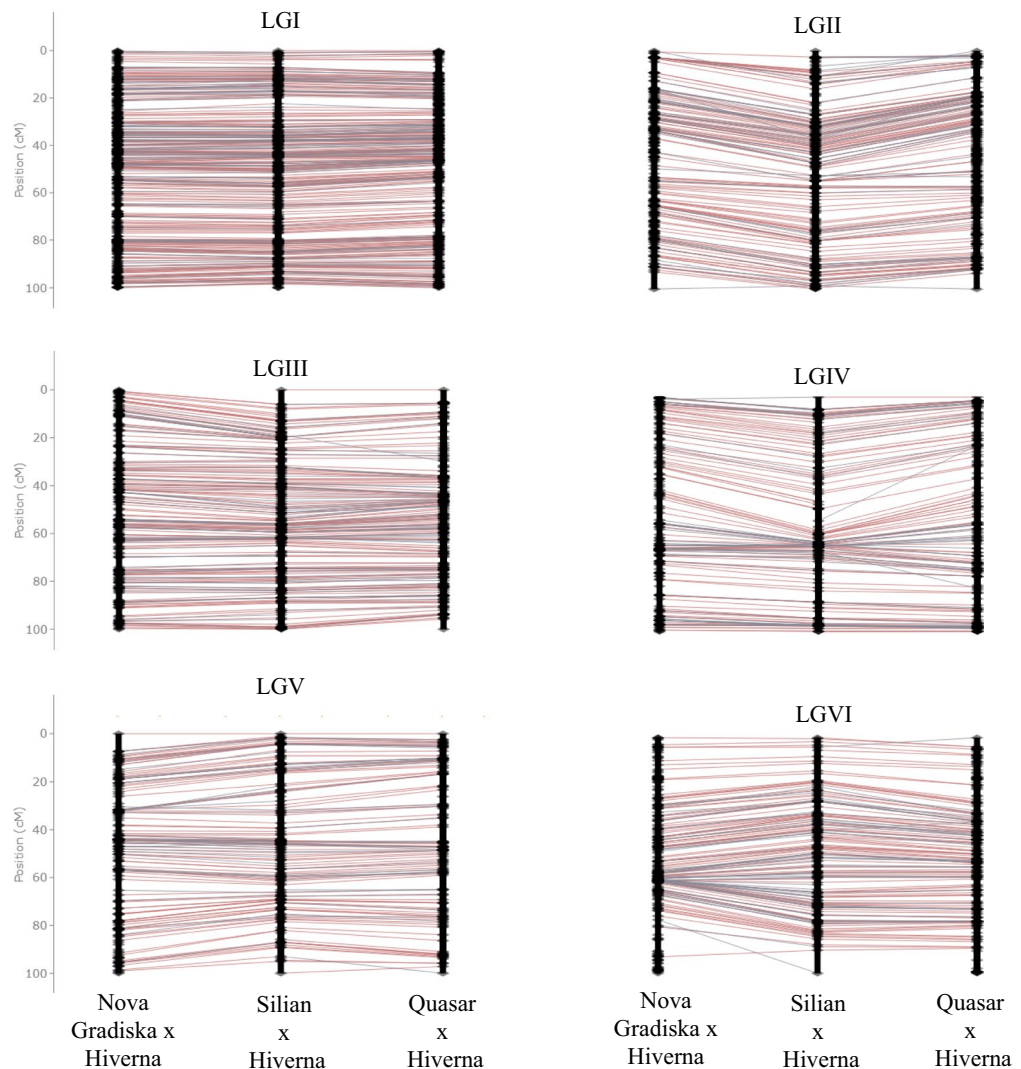
**Figure 2.** Venn diagram showing the SNP markers shared between the individual genetic maps of the three populations presented in this study.

## Results

**Transcriptome assembly and gene annotation.** A total of 1.1 million RNA-seq reads from 4 faba bean accessions, i.e. HIVERNA, NOVA GRADISKA, SILIAN and QUASAR, were assembled into 164,529 contigs (Table S1). After filtering on coverage (see Materials and Methods section), a transcriptome resource of 39,423 high-quality contigs with a N50 of 1460 bp (Table S2) and a BUSCO completeness score of 84.8% was built (Table S3). Functional annotation was obtained for 24,507 contigs (Fig. S1). Each contig of the unigene set was assigned, if possible, to the categories of biological processes, molecular functions and cellular components (Fig. S1). Among the biological processes, the metabolism of nucleobase, nucleoside, nucleotide and nucleic acid (10.26%), biosynthesis (9.67%) and cell organization and biogenesis (8.02%) were the main classes represented (Fig. S1A). Catalytic (36.49%) and transferase (13.17%) activities contributed in greater proportion to the category of molecular function (Fig. S1B). The cytoplasmic (31.68%) and nucleus (15.91%) cellular components were the most represented classes within the annotation (Fig. S1C).

**SNP discovery, selection, genotyping and individual genetic linkage of the three F<sub>3</sub> populations.** A total of 105,828 homozygous SNPs (Fig. S2) were detected on 19,190 contigs (5.5 SNP/contig) with an average coverage of 47.7 $\times$ . In total, 64.77% of the SNPs were transitions while 35.23% were transversions. Out of these robust SNPs, 2,000 were selected based on the following parameters: SNP quality score, polymorphism in more than one population (see Materials and Methods section), potential synteny with pea and *Medicago truncatula* and 1 SNP maximum per contig. Capture probes were then designed to allow large-scale targeted genotyping. Of the 2,000 gene-based SNP markers selected above, 1,911 markers were successfully scored on the progenies of three F<sub>3</sub> inbred populations, i.e., Pop1-3 (Table S4). Of them, 95.2% (1,819 SNPs) were polymorphic in at least one population. Altogether, 1,446 SNPs were polymorphic in Pop1, 1,499 in Pop2 and 1,409 in Pop3 (Table S4).

**Linkage mapping of the three F<sub>3</sub> populations.** Individual genetic maps were constructed for the three F<sub>3</sub> populations after filtering markers for distortion and missing data (see Materials and Methods section). Two hundred thirty-three (Pop1), 189 (Pop2) and 175 (Pop3) SNP markers were placed on the individual scaffold maps (Fig. S3). The framework maps included 350 (Pop1), 209 (Pop2) and 326 (Pop3) markers (Fig. S4). The full maps had 1,438 markers for Pop1, 1,312 markers for Pop2 and 1,406 markers for Pop3 that covered all six faba bean LGs (Table 1; Figs. 1, 2 and S5, Tables S5–S7). The number of markers per LG ranged between 141 (Pop3, LGV) and 401 (Pop1, LGI) SNP markers (Table 1). Total map lengths were: 1,426 (Pop1), 1,832 (Pop2) and 1,697 (Pop3) cM. The density of markers was high for all LGs in the three populations. Pop1 had an average marker density of 1.01 markers per cM, while in Pop2 this density was 0.72 markers/cM and in Pop3 it was 0.83 markers/cM. The average gap size between pairs of non-colocalized markers was 1.16 cM in Pop1, 1.47 cM in Pop2 and 1.36 cM in Pop3 (Table 1). Only a few large gaps (>10 cM) were observed: three gaps on the linkage map for Pop1 (LGII and LGV), nine gaps on the map from Pop2 (LGI, LGIII, LGIV, LGV and LGVI) and eight gaps on the map from Pop3 (LGI, LGII, LGIII and LGV) (Table 1). Nine hundred twenty-eight markers were common to the



**Figure 3.** Order comparison of the marker assignments by linkage group between the individual linkage maps of the three recombinant populations presented in this study. Marker positions are normalized.

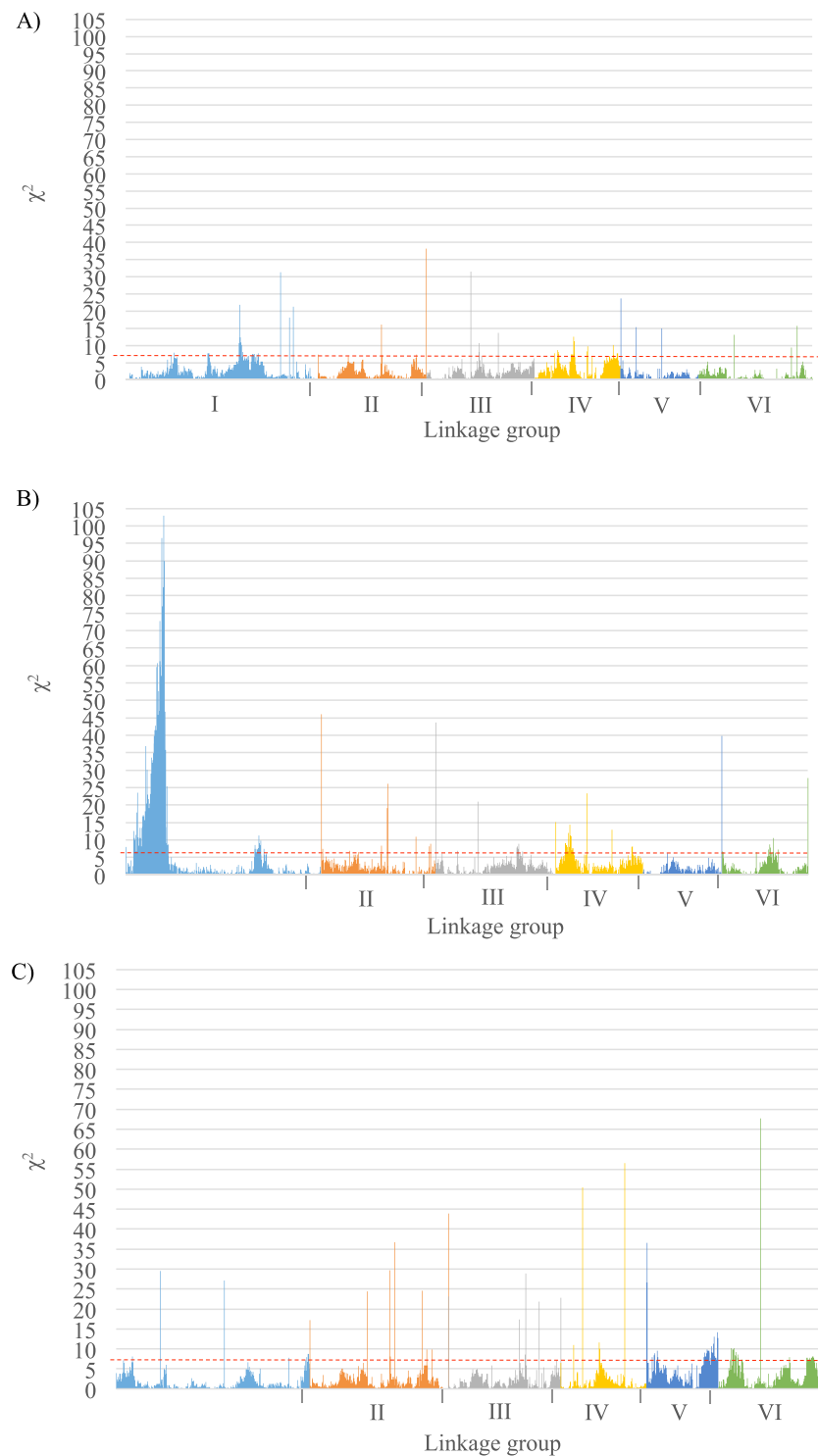
three populations (Fig. 2). One thousand five hundred and four SNPs were mapped onto at least two of the three linkage maps. Pairwise comparisons of the positions and the orders of common markers were performed among the three populations to assess synteny and collinearity (Fig. 3). The positions of the marker were consistent in the LGs of the three populations. High positive correlations (Spearman test) between map orders were obtained: Pop1-Pop2,  $r = 0.98$ ; Pop1-Pop3,  $r = 0.99$  and Pop2-Pop3,  $r = 0.98$ ;  $P < 0.001$ .

**Segregation distortion of the individual maps.** There was a significant distortion of segregation ( $\chi^2$  test,  $P < 0.05$ ) with respect to the expected Mendelian segregation ratio (37.5:25:37.5) for a minority of markers in the three populations (5.57% in Pop1, 9.14% in Pop2, 8.52% in Pop3). Only a few markers exceeded the index distortion threshold of 0.8 used for mapping (see Materials and Methods section).

Pop2 presented a region with segregation distortion towards SILIAN alleles at the bottom of LGI (Figs. 4 and S6). Although with less intensity, LGI of Pop1 also displayed a region in which segregation favoured the alleles of the male parent NOVA GRADISKA (Figs. 4 and S6). Outside of this region, the segregation distortion slightly favoured NOVA GRADISKA alleles in Pop1 (except in LGIII where HIVERNA alleles were over-represented) and SILIAN alleles in Pop2 (except in LGVI where HIVERNA alleles were favoured) while in Pop3 HIVERNA alleles were favoured (except in LGIII and LGIV where QUASAR alleles were more frequent) (Fig. S6).

**Integration of the individual maps into a consensus map of faba bean.** General collinearity among the individual genetic linkage maps (Fig. 3) helped to construct a dense consensus map of 1,547.71 cM that included 1,728 markers (95% of the polymorphic SNPs) (Table 1, Fig. 5, Table S8). The number of markers per LG varied from a minimum of 153 (LGV) to a maximum of 375 (LGI) SNPs (Table 1). The density of markers in this consensus map was high in all LGs, with an average density of 1.12 markers/cM. The average distance between two markers was 1.17 cM (Table 1). Nonetheless, 9 intervals were found with a distance between two markers

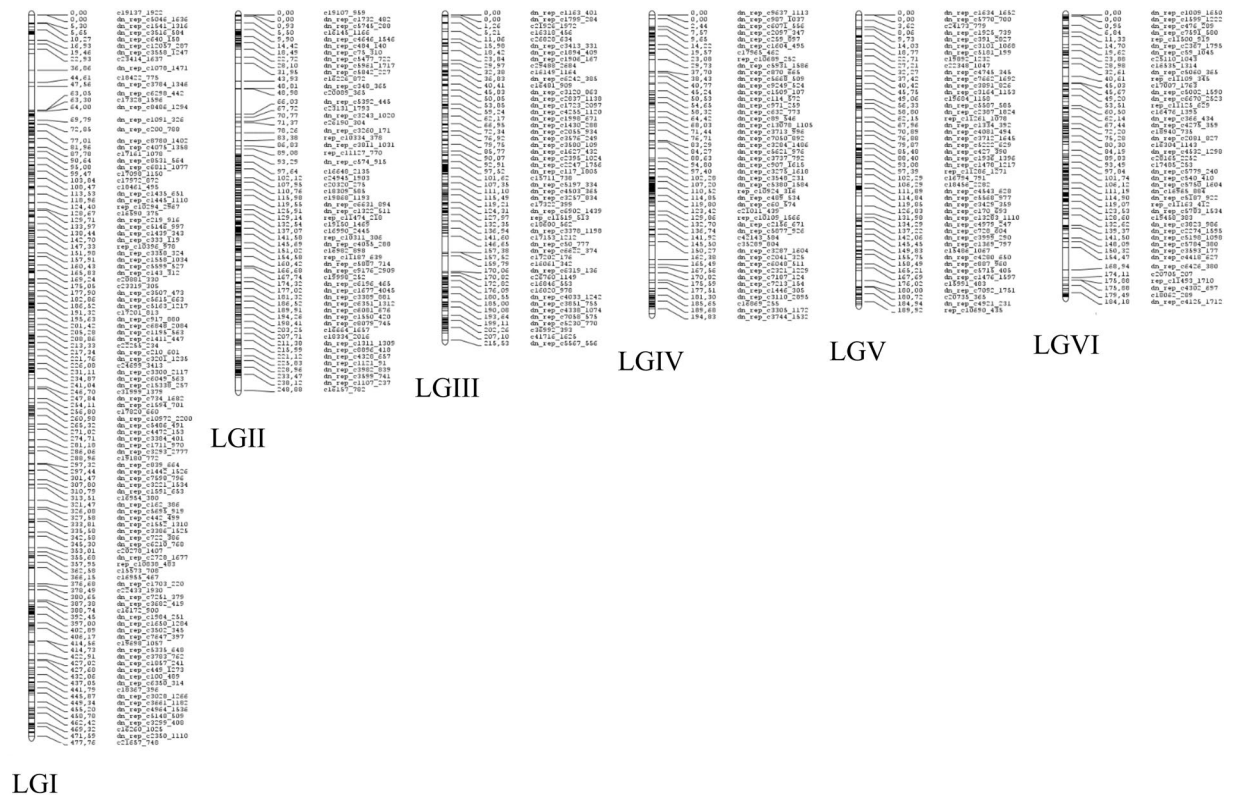




**Figure 4.** Distribution of the segregation distortion ( $\chi^2$  test,  $P = 0,05$ ) of the SNPs used to construct the three faba bean individual maps in the recombinant populations studied throughout the six linkage groups. (A) Pop1; (B) Pop2 and; (C) Pop3. Red dashed line indicates the threshold above which segregation distortion is significant.

greater than 10 cM, the largest gap being 21.51 cM (Table 1). The position and order of markers on the individual and consensus maps was overall conserved ( $r = 0.99$  for Pop1-consensus, Pop2-consensus Pop3-consensus,  $P < 0.001$ ) (Figs. 6 and S7). A few local inversions of the order of markers were observed (Fig. S7).

Comparison between the faba bean consensus map SNP marker sequences reported here and those presented in the Webb *et al.*<sup>21</sup> consensus map using BLASTn search highlighted markers located in the same genes. Eighty-eight common markers (e.g. markers corresponding to the same *Medicago truncatula* gene sequence) were found. The distribution and position of these common markers between both maps showed highly conserved

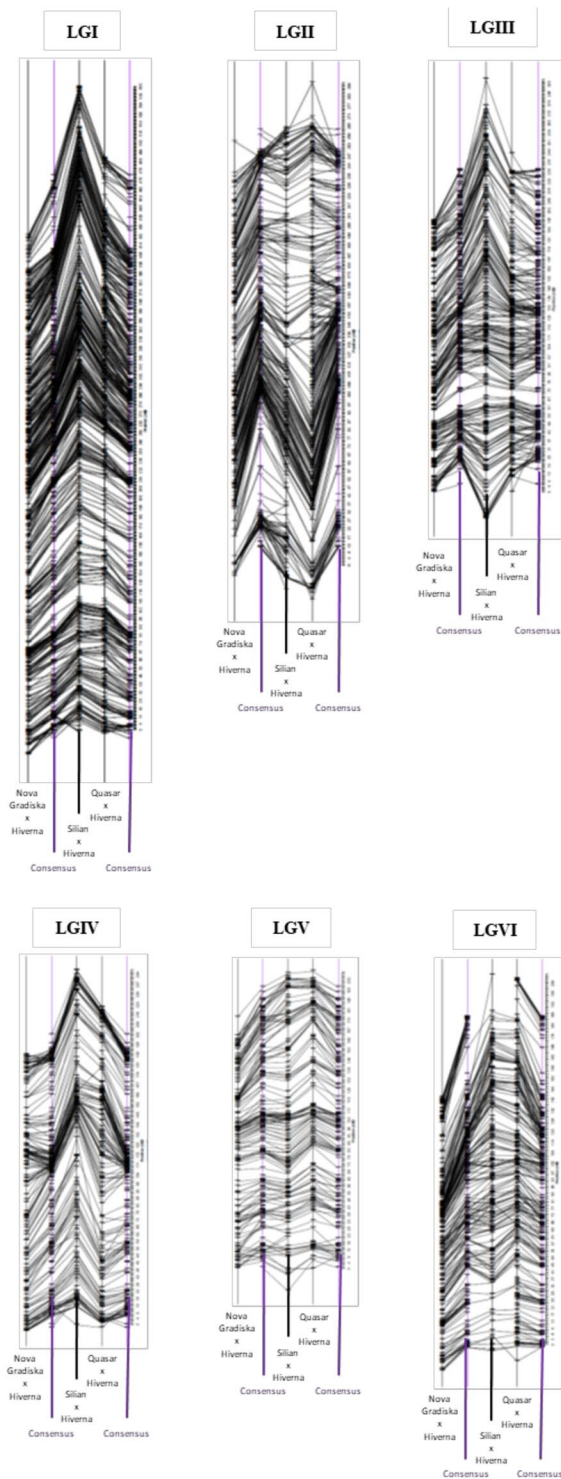


**Figure 5.** Faba bean consensus genetic linkage map generated after the integration of three individual maps originated from three populations consisting of (A) 102 F<sub>3</sub> recombinant lines derived from the cross between *Vicia faba* cv. Nova Gradiska and Hiverna; (B) 147 F<sub>3</sub> derived from the cross between cv. Silian and Hiverna and (C) 96 F<sub>3</sub> derived from the cross between cv. Quasar and Hiverna.

collinearity between them in the six LGs, although a few marker inversions were also observed (Fig. S8), which is another evidence of the reliability of the data presented here.

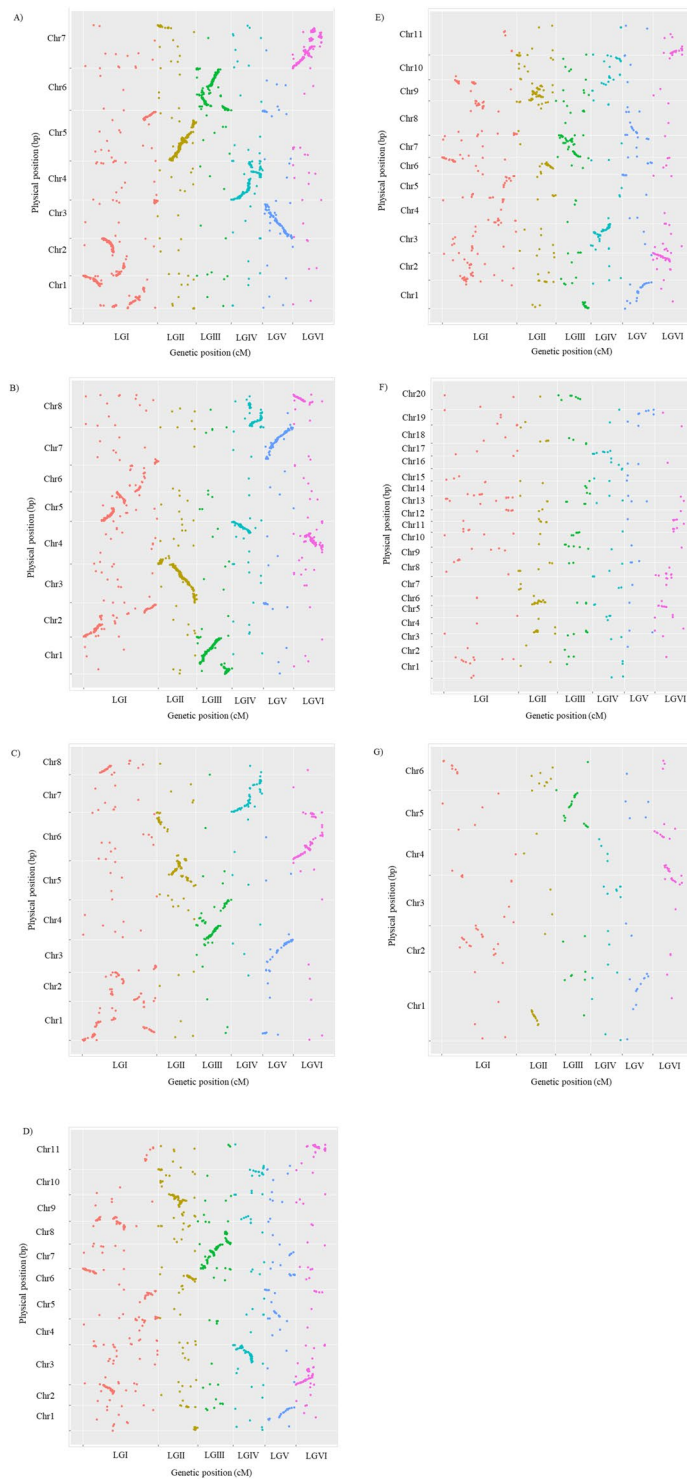
**Macrosyteny between the faba bean consensus map and the genomes of related legume species.** Syntenic and collinear relationships between the faba bean consensus map presented in this study and the genomes of related legume species are summarized in Fig. 7. As expected, the degree of synteny and collinearity between faba bean and the legume species compared here increased when the phylogenetic distance decreased and vice versa. The genomes of *P. sativum*, *C. arietinum* and *M. truncatula* showed high levels of macrosyteny with our consensus map (Fig. 7). The best blast hits on the *M. truncatula* and *P. sativum* genomes for the faba bean markers' flanking sequences are described in Table S4, including their annotations and positions. There was a high conservation of synteny and collinearity between LGs II, IV, V and VI of faba bean and PsChr5, PsChr4, PsChr3 and PsChr7 of *P. sativum* and LGs II, III, IV and V of faba bean and MtChr3, MtChr1, MtChr4 and MtChr7 of *M. truncatula*, respectively (Fig. 7). LGs III, IV, V and VI of faba bean turned out to be almost completely collinear to the chromosomes CaChr4, CaChr7, CaChr3 and CaChr6 of *C. arietinum*, respectively (Fig. 7). However, syntenic regions were associated with several chromosomes of the sequenced legumes in the rest of the LGs of faba bean. The most striking case was that of LGI, which was the longest LG. Syntenic blocks corresponding to *V. faba* LGI were found in CaChr1, CaChr2 and CaChr8, in PsChr1, PsChr2 and PsChr5, and also in MtChr2 and MtChr5 (Fig. 7). Substantial collinear blocks were also found between *V. faba* and *P. vulgaris*, *V. unguiculata* and *L. japonicus*, despite their greater phylogenetic distance from faba bean (Fig. 7). The same was observed between faba bean and *L. japonicus*: vFLGI-LjChr2 and -LjChr6, vFLGII-LjChr1, vFLGIII-LjChr5, vFLGV-LjChr1, vFLGVI-LjChr3 and -LjChr4; faba bean and *P. vulgaris*: vFLGI-PvChr6, vFLGII-PvChr6 and -PvChr9, vFLGIII-PvChr7, vFLGIV-PvChr3, vFLGV-PvChr1 and -PvChr8, vFLGVI-PvChr2 and -PvChr11; faba bean and *V. unguiculata*: vFLGI-VuChr2, -VuChr7 and -VuCVhr9, vFLGII-VuChr6 and -VuChr9, vFLGIII-VuChr7 and -VuChr8, vFLGIV-VuChr3, vFLGV-VuChr1 and vFLGVI-VuChr3 (Fig. 7). In the case of soybean, conservation patterns can be intuited between vFLGI-GmChr13, vFLGII-GmChr4 and -GmChr6, vFLGIII-GmChr10, -GmChr14 and -GmChr20, vFLGIV-GmChr17, vFLGV-GmChr19 and vFLGVI-GmChr5, -GmChr8, -Gm12 and -Gm13 (Fig. 7).

**Discussion**  
**New faba bean genetic resources: transcriptome, gene-based SNPs, gene-based SNP markers and a consensus map.** Europe suffers a significant deficit of plant proteins that makes it necessary to import up to 70% of the plant-based proteins consumed<sup>54</sup>. Grain legumes including faba bean are good candidates to boost EU plant protein production due to the high protein content of their seeds. Despite this potential and



**Figure 6.** Comparison of the marker assignments by position between the individual maps of the three populations and the consensus map.

the important environmental services related to grain legume production, these crops represent only 3–4% of the arable land. The low investment in breeding programs has limited the development of stable high-yielding varieties, resistant to biotic and abiotic stresses<sup>55</sup>. The future of the faba bean crop depends on efficient breeding programs including MAS and/or genomic-assisted breeding (GAB), in which the development of improved varieties is accelerated. For this, new faba bean genetic resources are needed. Fortunately, the progress and cheapening of NGS and assembly technologies in recent years are allowing the development of new genetic resources. The suitability and effectiveness of the transcriptome sequencing approach for the generation of gene-based SNP markers resulted in a transcriptomic resource of 39,423 faba bean transcripts obtained after *de novo* assembly



**Figure 7.** Dot-plots for synteny and collinearity comparisons between the faba bean consensus map (cM) and the genomes (bp) of (A) *Pisum sativum*; (B) *Medicago truncatula*; (C) *Cicer arietinum*; (D) *Vigna unguiculata* (E) *Phaseolus vulgaris*; (F) *Glycine max* and (G) *Lotus japonicas*.

(File S1), of which 24,507 contigs were annotated. Although this amount is lower than that of other studies<sup>56,57</sup>, it is a sufficiently high number for the discovery of robust SNP markers well distributed throughout the faba bean genome. These data are also available for use in transcriptome comparisons with other faba bean genotypes and between faba bean and other species. In addition, the faba bean transcriptome was used to identify 105,828 gene-based SNPs (File S2). The present work makes this new source of SNPs available to the faba bean community, who can develop additional SNP markers useful in other genetic backgrounds. Two thousand non-redundant *loci* (Table S4) from the 105,828 SNPs were selected to develop the molecular markers for genotyping 245

recombinant lines from three populations and the four parental lines that originated them. One thousand nine hundred eleven SNP markers (95.5%) were validated after a successful genotyping (Table S4). Our set of 1,819 polymorphic gene-based SNPs is a valuable tool for the faba bean community and particularly for breeders. Since the markers were designed in genes, they are highly informative and allow establishing syntenic relationships with other species<sup>32,58,59</sup>. The high quality of this set of markers was confirmed after the construction of three individual genetic maps derived from three different populations.

High collinearity between individual maps led to the construction of the densest faba bean consensus map known to date. The map includes 1,728 well-distributed markers along six LGs that correspond to the six faba bean chromosomes and covers 1,547.71 cM with a dense marker placement (0.89 cM between adjacent pairs of markers on average) (Fig. 5). In accordance with previous cytogenetic studies LG1, corresponding to chromosome 1, was the largest linkage group<sup>60,61</sup>. The total map size is consistent with the 1,403.8 cM size of the consensus map previously published by Webb *et al.*<sup>21</sup>. Despite having different genetic backgrounds, both consensus maps showed a good collinearity (Fig. S8), which confirming the good quality of both maps. Consensus gene-based genetic linkage maps are useful in meta-QTL analysis, phylogenetic and comparative genomic studies, map-based cloning and GAB, especially in the absence of a genome sequence. The phenotypic characteristics of the parents of the recombinant lines make these populations and maps useful resources. The next steps in our research will be to perform Quantitative Trait Locus (QTL) analyses to identify potential candidate genes for resistance to faba bean seed weevils in Pop1 and Pop3. In a previous work<sup>35</sup>, we reported that the male parents NOVA GRADISKA and QUASAR present partial resistance to the attack of bruchids. Differences in parental responses to bruchids attacks suggested distinct resistance mechanisms in the two accessions. This would be of great advantage in breeding since different genes could be pyramided and introgressed simultaneously in cultivars, making the resistance to faba bean weevils more durable and contributing to agriculture with less need for pesticides. In the case of Pop2, a distorted region was located on the top of LG1 (Figs. 4 and S6). Knowledge about this area is of great importance for MAS because the genes located in this part of LG1 of Pop2 seem to segregate together in favour of the genetic background of SILLAN. If favourable alleles of a gene of interest were located in this area of the LG1 but at the same time there were nearby genes that carried unfavourable alleles, most likely, they would all segregate together. Thus, the introgression of favourable agronomic traits will be quite difficult in such a situation. Other previously published faba bean maps have also noted the presence of distorted regions throughout the different linkage groups<sup>62–64</sup>.

### Syntenic regions shared between faba bean and other legumes will facilitate future comparative genomic studies.

Exploitation of the syntenic relationships between the faba bean consensus map developed in this work and available legume genomes will make the identification of candidate genes of important traits easier in the future and will enable forthcoming syntenic-based gene cloning approaches. As expected, robust macrosyntenic blocks that sometimes nearly cover a complete chromosome were found between faba bean LGs and pea, barrel medic and chickpea chromosomes since their phylogenetic proximity is greater than that of the rest of the compared sequenced legumes (Fig. 7). In accordance with our results, Webb *et al.*<sup>21</sup> also reported good levels of synteny between their consensus map, the genome of *M. truncatula* and the genetic map of lentil developed by Sharpe *et al.*<sup>27</sup>. In addition, we have been able to locate abundant blocks of macrosynteny between faba bean and common bean, cowpea or birdsfoot trefoil despite their greater evolutionary distances (Fig. 7), providing further evidence of the mapping accuracy. By contrast, the number of markers in our consensus map may not be enough to clarify the macrosynteny between faba bean and soybean due to the extensive chromosomal rearrangements and polyploidization of the soybean genome. Hopefully, gene conservation with soybean will be more evident once the faba bean genome is available, as has happened in the case of the pea<sup>48</sup>. Although duplication of the soybean genome and chromosomal rearrangement are a limitation for translational genomics with faba bean, synteny in duplicate regions would be a good resource to exploit. Despite the large size of the faba bean genome, synteny data reflects a globally conserved organization with respect to the genome of the legumes studied here. There is of course a certain amount of reorganization that can be easily observed, for example, in the condensed LG1 of faba bean that gathers the genes located on chromosomes 1, 2 and 5 of pea. These results include *V. faba* as an additional syntenic species in the paleogenomic scheme described in Kreplak *et al.*<sup>48</sup>.

In conclusion, this work provides to faba bean researchers and breeders a new faba bean exome assembly originated from transcriptome data of four accessions (HIVERNA, NOVA GRADISKA, SILLAN and QUASAR), a set of 105,828 gene-based SNPs and 1,819 mapped SNP markers on three individual linkage maps and one consensus map. The high quality of the assembly resulted in the identification of a large number of SNPs of the most informative type due to their location in genes. The molecular markers designed from this set of SNPs were validated in three recombinant populations, resulting in the densest faba bean consensus map to date. The SNP markers designed here are available for genotyping other inbred populations that could be integrated later into our consensus map. These robust resources will be useful for trait mapping, genetic diversity and linkage disequilibrium studies or map-based cloning, and will enable faba bean MAS and GAB as well as the identification of candidate genes of agronomic interest through syntenic-based approaches.

Received: 2 March 2020; Accepted: 2 April 2020;

Published online: 22 April 2020

### References

1. Stagnari, F., Maggio, A., Galieni, A. & Pisante, M. Multiple benefits of legumes for agriculture sustainability: an overview. *Chemical and Biological Technologies in Agriculture* **4** (2017).
2. O'Sullivan, D. M. & Angra, D. Advances in Faba Bean Genetics and Genomics. *Front. Genet.* **7**, 150 (2016).
3. Mulualem, T., Dessalegn, T. & Dessalegn, Y. Participatory varietal selection of faba bean (*Vicia faba* L.) for yield and yield components in Dabat district, Ethiopia. *Wudpecker. J. Agric. Res* **7**, 270–274 (2012).

4. Food and Agriculture Organization of the United Nations (FAO). FAOSTAT. Available at, <http://www.fao.org/faostat> (2017).
5. Kharrat, M., Le Guen, J. & Tivoli, B. Genetics of resistance to 3 isolates of *Ascochyta fabae* on Faba bean (*Vicia faba* L.) in controlled conditions. *Euphytica* **151**, 49–61 (2006).
6. Cernay, C., Ben-Ari, T., Pelzer, E., Meynard, J.-M. & Makowski, D. Estimating variability in grain legume yields across Europe and the Americas. *Sci. Rep.* **5**, 11171 (2015).
7. Maalouf, F. *et al.* Development of faba bean productivity and production in the Nile Valley, Red Sea and Sub-Saharan region. (2009).
8. Nayak, G. K. *et al.* Interactive effect of floral abundance and semi-natural habitats on pollinators in field beans (*Vicia faba*). *Agric. Ecosyst. Environ.* **199**, 58–66 (2015).
9. Denton, M. D., Pearce, D. J. & Peoples, M. B. Nitrogen contributions from faba bean (*Vicia faba* L.) reliant on soil rhizobia or inoculation. *Plant Soil* **365**, 363–374 (2013).
10. Cooper, J. W. *et al.* Enhancing faba bean (*Vicia faba* L.) genome resources. *J. Exp. Bot.* **68**, 1941–1953 (2017).
11. Negruk, V. Mitochondrial Genome Sequence of the Legume *Vicia faba*. *Front. Plant Sci.* **4**, 128 (2013).
12. Patto, M. C. V., Torres, A. M., Kobližkova, A., Macas, J. & Cubero, J. I. Development of a genetic composite map of *Vicia faba* using F 2 populations derived from trisomic plants. *TAG. Theor. Appl. Genet.* **98**, 736–743 (1999).
13. Román, B., Torres, A. M., Rubiales, D., Cubero, J. I. & Satovic, Z. Mapping of quantitative trait loci controlling broomrape (*Orobanche crenata* Forsk.) resistance in faba bean (*Vicia faba* L.). *Genome* **45**, 1057–1063 (2002).
14. Avila, C. M. *et al.* Isolate and organ-specific QTLs for ascochyta blight resistance in faba bean (*Vicia faba* L.). *TAG. Theor. Appl. Genet.* **108**, 1071–1078 (2004).
15. Gutierrez, M. V. *et al.* Cross-species amplification of *Medicago truncatula* microsatellites across three major pulse crops. *Theor. Appl. Genet.* **110**, 1210–1217 (2005).
16. Ellwood, S. R. *et al.* Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. *BMC Genomics* **9**, (2008).
17. Zeid, M. *et al.* Simple sequence repeats (SSRs) in faba bean: new loci from *Orobanche*-resistant cultivar ‘Giza 402’. *Plant Breed.* **128**, 149–155 (2009).
18. Diaz-Ruiz, R. *et al.* Confirmation of QTLs controlling *Ascochyta fabae* resistance in different generations of faba bean (*Vicia faba* L.). *Crop Pasture Sci.* **60**, 353 (2009).
19. Satovic, Z. *et al.* A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.). *BMC Genomics* **14** (2013).
20. Kaur, S. *et al.* 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 (2014).
21. Webb, A. *et al.* A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba* L.). *Plant Biotechnol. J.* **14**, 177–185 (2016).
22. Barbazuk, W. B., Emrich, S. J., Chen, H. D., Li, L. & Schnable, P. S. SNP discovery via 454 transcriptome sequencing. *Plant J.* **51**, 910–918 (2007).
23. Ma, Y. *et al.* Development and characterization of 21 EST-derived microsatellite markers in *Vicia faba* (fava bean). *Am. J. Bot.* **98**, e22–4 (2011).
24. Galeano, C. H. *et al.* Saturation of an Intra-Gene Pool Linkage Map: Towards a Unified Consensus Linkage Map for Fine Mapping and Synteny Analysis in Common Bean. *PLoS One* **6**, e28135 (2011).
25. Kaur, S. *et al.* Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics* **13**, 104 (2012).
26. Loridon, K. *et al.* Single-nucleotide polymorphism discovery and diversity in the model legume *Medicago truncatula*. *Mol. Ecol. Resour.* **13**, 84–95 (2013).
27. Sharpe, A. G. *et al.* Ancient orphan crop joins modern era: Gene-based SNP discovery and mapping in lentil. *BMC Genomics* **14**, 192 (2013).
28. Duarte, J. *et al.* Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *BMC Genomics* **15**, 126 (2014).
29. Kim, C. *et al.* Application of genotyping by sequencing technology to a variety of crop breeding programs. *Plant Sci.* **242**, 14–22 (2016).
30. Davey, J. W. *et al.* Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat. Rev. Genet.* **12**, 499–510 (2011).
31. Rothberg, J. M. *et al.* An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348–352 (2011).
32. Tayeh, N. *et al.* Development of two major resources for pea genomics: The GenoPea 13.2K SNP Array and a high-density, high-resolution consensus genetic map. *Plant J.* **84**, 1257–1273 (2015).
33. Schlötterer, C. Opinion: The evolution of molecular markers — just a matter of fashion? *Nat. Rev. Genet.* **5**, 63–69 (2004).
34. Gupta, P. K., Rustgi, S. & Mir, R. R. Array-based high-throughput DNA markers for crop improvement. *Heredity (Edinb.)* **101**, 5–18 (2008).
35. Carrillo-Perdomo, E. *et al.* Identification of Novel Sources of Resistance to Seed Weevils (*Bruchus* spp.) in a Faba Bean Germplasm Collection. *Front. Plant Sci.* **9**, 1914 (2019).
36. Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
37. Haas, B. J. *et al.* De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nat. Protoc.* **8**, 1494–1512 (2013).
38. Waterhouse, R. M. *et al.* BUSCO Applications from Quality Assessments to Gene Prediction and Phylogenomics. *Mol. Biol. Evol.* **35**, 543–548 (2018).
39. Huerta-Cepas, J. *et al.* Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Mol. Biol. Evol.* **34**, 2115–2122 (2017).
40. Huerta-Cepas, J. *et al.* eggNOG 5.0: a hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. *Nucleic Acids Res.* **47**, D309–D314 (2019).
41. Li, H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics* **27**, 2987–2993 (2011).
42. Martin, M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal* **17**, 10 (2011).
43. Van Ooijen, J. W. In *Kyazma BV* 33, 1371 (2006).
44. Leroux, D. & Jasson, S. Spell-QTL, a New Tool for QTL Analysis on Modern Datasets. in *PAG XXV - Plant and Animal Genome Conference* (2017).
45. Ganal, M. W. *et al.* A Large Maize (*Zea mays* L.) SNP Genotyping Array: Development and Germplasm Genotyping, and Genetic Mapping to Compare with the B73 Reference Genome. *PLoS One* **6**, e28334 (2011).
46. Haldane, J. The combination of linkage values, and the calculation of distances between the loci of linked factors. *J. Genet.* **8**, 299–309 (1919).
47. Varshney, R. K. *et al.* Genetic dissection of drought tolerance in chickpea (*Cicer arietinum* L.). *Theor. Appl. Genet.* **127**, 445–462 (2014).
48. Kreplak, J. *et al.* A reference genome for pea provides insight into legume genome evolution. *Nat. Genet.* **51**, 1411–1422 (2019).
49. Pecrix, Y. *et al.* Whole-genome landscape of *Medicago truncatula* symbiotic genes. *Nature Plants* **4**, 1017–1025 (2018).

50. Sato, S. *et al.* Genome structure of the legume, *Lotus japonicus*. *DNA Res.* **15**, 227–239 (2008).
51. Schmutz, J. *et al.* Genome sequence of the palaeopolyploid soybean. *Nature* **463**, 178–183 (2010).
52. Schmutz, J. *et al.* A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**, 707–713 (2014).
53. Lonardi, S. *et al.* The genome of cowpea (*Vigna unguiculata* [L.] Walp.). *Plant J.* **98**, 767–782 (2019).
54. Rubiales, D. & Mikic, A. Introduction: Legumes in Sustainable Agriculture. *CRC. Crit. Rev. Plant Sci.* **34**, 2–3 (2015).
55. Magrini, M.-B. *et al.* Pulses for Sustainability: Breaking Agriculture and Food Sectors Out of Lock-In. *Front. Sustain. Food Syst.* **2** (2018).
56. Khan, M. A. *et al.* Transcriptome profiling of faba bean (*Vicia faba* L.) drought-tolerant variety hassawi-2 under drought stress using RNA sequencing. *Electron. J. Biotechnol.* **39**, 15–29 (2019).
57. Gao, B. *et al.* Comprehensive transcriptome analysis of faba bean in response to vernalization. *Planta* **251** (2020).
58. Choi, H., Mun, J. & Kim, D. . . . H. Z.-P. of the & 2004, U. Estimating genome conservation between crop and model legume species. *Natl. Acad. Sci.* **101**, 15289–15294 (2004).
59. Aubert, G. *et al.* Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. *Theor. Appl. Genet.* **112**, 1024–1041 (2006).
60. Lucretti, S., Doležel, J., Schubert, I. & Fuchs, J. Flow karyotyping and sorting of *Vicia faba* chromosomes. *Theor. Appl. Genet.* **85**, 665–672 (1993).
61. Doležel, J. & Lucretti, S. High-resolution flow karyotyping and chromosome sorting in *Vicia faba* lines with standard and reconstructed karyotypes. *Theor. Appl. Genet.* **90**, 797–802 (1995).
62. Ellwood, S. R. *et al.* Construction of a comparative genetic map in faba bean (*Vicia faba* L.); conservation of genome structure with *Lens culinaris*. *BMC Genomics* **9**, 1–11 (2008).
63. Cruz-Izquierdo, S. *et al.* Comparative genomics to bridge *Vicia faba* with model and closely-related legume species: Stability of QTLs for flowering and yield-related traits. *Theor. Appl. Genet.* **125**, 1767–1782 (2012).
64. Khazaei, H., O'Sullivan, D. M., Sillanpää, M. J. & Stoddard, F. L. Use of synteny to identify candidate genes underlying QTL controlling stomatal traits in faba bean (*Vicia faba* L.). *Theor. Appl. Genet.* **127**, 2371–2385 (2014).

## Acknowledgements

Authors thank the field and laboratory staff for technical assistance, especially Philippe Declerck and Dominique Notteau (RAGT Semences), Gilles Furet and Corie Dekker (Limagrain) and Matthieu Floriot (Agri Obtentions) for their participation in the production of the recombinant populations. In addition, authors are grateful to Nathalie Rivière (Biogemma) for useful advice in the choice of the genotyping technology. This work was supported by the Project Investissements d'Avenir PeaMUST under the grant number ANR-11-BTBR-0002. GQE - Le Moulon benefited from the support of Saclay Plant Sciences-SPS (ANR-17-EUR-0007).

## Author contributions

Estefania Carrillo-Perdomo analysed the data, built the genetic maps and wrote the manuscript; Adrien Vidal developed the mapping codes, built the maps and critically reviewed the manuscript; Jonathan Kreplak filtered the contig sequences of the SNPs and contributed to transcriptome and synteny analyses; Hervé Duborjal, Magalie Leveugle, Jorge Duarte, Jean Philippe Pichon and Catherine Desmetz conducted experiments; Chrystel Deulvot, Blandine Raffiot and Pascal Marget developed the recombinant populations; Nadim Tayeh contributed to map comparisons and synteny analyses and critically reviewed the manuscript; Matthieu Falque and Olivier C. Martin designed the mapping algorithm, specified the software, followed the developments and critically reviewed the manuscript; Judith Burstin conceived the experiments and critically reviewed the manuscript; Grégoire Aubert conceived and supervised the experiments, edited and reviewed the manuscript. All authors read and approved the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-020-63664-7>.

**Correspondence** and requests for materials should be addressed to E.C.-P.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020