

RESEARCH ARTICLE

Pea Marker Database (PMD) – A new online database combining known pea (*Pisum sativum* L.) gene-based markers

Olga A. Kulaeva¹, Aleksandr I. Zhernakov¹, Alexey M. Afonin¹, Sergei S. Boikov¹, Anton S. Sulima¹, Igor A. Tikhonovich^{1,2}, Vladimir A. Zhukov^{1*}

1 All-Russia Research Institute for Agricultural Microbiology, Podbelsky chausse, Saint-Petersburg, Russia, **2** Saint-Petersburg State University, Universitetskaya embankment, Saint-Petersburg, Russia

* VZhukov@arriam.ru



OPEN ACCESS

Citation: Kulaeva OA, Zhernakov AI, Afonin AM, Boikov SS, Sulima AS, Tikhonovich IA, et al. (2017) Pea Marker Database (PMD) – A new online database combining known pea (*Pisum sativum* L.) gene-based markers. PLoS ONE 12(10): e0186713. <https://doi.org/10.1371/journal.pone.0186713>

Editor: Mohar Singh, National Bureau of Plant Genetic Resources, Pusa India, INDIA

Received: April 19, 2017

Accepted: August 17, 2017

Published: October 26, 2017

Copyright: © 2017 Kulaeva et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All relevant data are within the paper, its Supporting Information files and at <http://peamarker.arriam.ru/>.

Funding: The work of AMA and IAT on development of PMD1 and localization of the Sym27 gene was supported by the Russian Science Foundation (<http://www.rscf.ru/en>) (Grant # 14-24-00135). The work of OAK on development of PMD2 was supported by the Russian Foundation for Basic Research (<http://www.rfbr.ru/>)

Abstract

Pea (*Pisum sativum* L.) is the oldest model object of plant genetics and one of the most agriculturally important legumes in the world. Since the pea genome has not been sequenced yet, identification of genes responsible for mutant phenotypes or desirable agricultural traits is usually performed via genetic mapping followed by candidate gene search. Such mapping is best carried out using gene-based molecular markers, as it opens the possibility for exploiting genome synteny between pea and its close relative *Medicago truncatula* Gaertn., possessing sequenced and annotated genome. In the last 5 years, a large number of pea gene-based molecular markers have been designed and mapped owing to the rapid evolution of “next-generation sequencing” technologies. However, the access to the complete set of markers designed worldwide is limited because the data are not uniformed and therefore hard to use. The Pea Marker Database was designed to combine the information about pea markers in a form of user-friendly and practical online tool. Version 1 (PMD1) comprises information about 2484 genic markers, including their locations in linkage groups, the sequences of corresponding pea transcripts and the names of related genes in *M. truncatula*. Version 2 (PMD2) is an updated version comprising 15944 pea markers in the same format with several advanced features. To test the performance of the PMD, fine mapping of pea symbiotic genes *Sym13* and *Sym27* in linkage groups VII and V, respectively, was carried out. The results of mapping allowed us to propose the *Sen1* gene (a homologue of *SEN1* gene of *Lotus japonicus* (Regel) K. Larsen) as the best candidate gene for *Sym13*, and to narrow the list of possible candidate genes for *Sym27* to ten, thus proving PMD to be useful for pea gene mapping and cloning. All information contained in PMD1 and PMD2 is available at www.peamarker.arriam.ru.

Introduction

Modern plant breeding relies on selection of genotypes with desirable traits by means of marker-assisted selection (MAS) and genomics-assisted breeding (GAB) [1,2]. Additionally,

rffi/eng) (Grant # 16-34-60132). The work of AIZ, ASS and VAZ on localization of the Sym13 gene and Sen1 sequencing was supported by the Russian Science Foundation (<http://www.rscf.ru/en>) (Grant # 16-16-00118). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

close attention is paid to identifying genes of interest by the candidate gene approach [3]. In cereals, significant success in both MAS and GAB has already been achieved, but in pulse crops the implementation of these approaches has been limited, mostly due to scarceness of available genomic resources and lack of optimized bioinformatics tools [2,3].

Garden pea (*Pisum sativum* L.) is one of the most valuable pulse crops, an integral part of agricultural systems throughout the world, and a model object in plant genetics since the days of Gregor Mendel [4,5]. However, modern pea genetics is lagging behind that of model plants, despite the significant progress made in discovery of symbiotic genes and the presence of available collections of unique mutant lines, for example, impaired in nitrogen-fixing symbiosis and arbuscular mycorrhiza development [6]. Pea has a large genome (1C = 4300 Mb) [7] congested with repetitive elements [8] complicating genome assembly; therefore, transcriptome assemblies of different tissues and organs are the most comprehensive source of pea gene sequences available [9–12]. Although the transcriptome assemblies alone are not very useful in studies concerning search for particular pea genes, they can provide the basis for construction of high-density genetic maps, which are the crucial tools for identification of loci and markers associated with traits of interest.

The first example of genetic linkage in pea was described in 1912 [13], and the first genetic map was constructed in 1925 [14], but only in the early 2000s complete genetic maps composed of 7 linkage groups (LGs) consistent with the pea karyotype were constructed mostly based on “anonymous” RFLP and RAPD markers [15–18]. Later, with the development of sequencing technologies and the emergence of pea EST databases, the first maps based on genic markers were built [19,20]. High level of genome synteny between *P. sativum* and *Medicago truncatula* Gaertn., a model legume with sequenced and annotated genome, opens the opportunity of comparative analysis between pea linkage groups and chromosomes of *M. truncatula*. This approach makes it possible to determine the nucleotide sequence of specific pea genes through fine gene mapping and subsequent search for candidate genes in *M. truncatula* and, possibly, other related species with sequenced and annotated genomes [5,21].

In recent years, numerous gene-based molecular markers have been designed and mapped in *P. sativum*. Development of “next-generation sequencing” (NGS) technologies allowed identifying thousands of single nucleotide polymorphism sites (SNPs) across a species’ genome, as demonstrated by several works aimed at polymorphism studies and construction of pea genetic maps [22–30]. These works are based either on transcriptome sequencing [22,24–28], or on alternative technologies such as RADSeq (Restriction site Associated DNA Sequencing) method (both reducing genome complexity), or whole genome sequencing technology [23,29,30]. However, the majority of genome-based markers are associated with random, often non-coding parts of genome rather than with particular functional transcripts, thus denying researchers the possibility of exploiting the synteny with *M. truncatula* or other legumes for candidate gene search.

Although a number of online marker databases exists, none of them provide information accumulated from different sources in a simple and convenient format [31,32], while support for others has been dropped entirely [33]. The need to gather information from different papers hinders the work on locus localization within a LG and further searches for candidate genes. Thus, the aim of this work was to integrate the information about pea markers and provide an easy-to-use online tool—the Pea Marker Database (PMD)—combining information about known pea gene-based markers.

At first, PMD1 was constructed to combine the markers from several independent sources [19,24–26,34–37]. By the time the development of PMD1 was finished, a highly comprehensive genetic map based on analysis of 12 pea mapping populations and consisting of new SNP markers and previously obtained markers from different studies [15,17,19,20,24–26,34–36,38–

[51] was constructed [27]. Unfortunately, the resultant marker positions were presented in large data files as supplementary material, and were, in our opinion, not convenient for intensive marker analysis. Thus, we transformed the data obtained by Tayeh and colleagues in 2015 [27] to a web-based, user-friendly form with additional important features, which we called PMD2.

Here, we present the description of the databases and the related online interface, as well as demonstrate the databases usability by (i) marker design for the particular pea genome region, followed by (ii) fine mapping of pea symbiotic genes *Sym13* and *Sym27*, and (iii) subsequent candidate gene search.

Materials and methods

Data sources

Markers for PMD1 were obtained from eight papers [19,24–26,34–36,38] and divided into four groups, or sets. Markers designed using NGS data were labeled according to the name of the first author of the corresponding article: **Dm** set (Duarte marker set) [24], **Sm** set (Sindhu marker set) [25] and **Lm** set (Leonforte marker set) [26]. The sequences of the markers in these sets were obtained from the supplementary files of the corresponding articles [24–26]. The fourth **ESTm** set (EST marker set) included EST markers [19,34–36,38] used to construct the previous pea genetic maps. The sequences of the EST markers were obtained from the NCBI nucleotide database [52].

The whole set of markers for PMD2 development was obtained from [27]. Of those, markers based on the transcripts from the work of Alves-Carvalho and colleagues in 2015 [12] were labeled **ACm** (Alves-Carvalho marker set) and used for PMD2 development along with **Dm**, **Sm** and **ESTm** sets, as well as with several markers from a range of previous works [15,17,20,39–51,53] not combined into a set (“separate markers”). The sequences for the markers in the **ACm** set were obtained from the pea transcriptome assembly available online at <http://bios.dijon.inra.fr/FATAL/cgi/pscaml.cgi> [54]. The sequences of the nucleotide-based separate markers [20,39–48,53] were obtained from the NCBI database.

PMD1 construction

As the first step of PMD1 construction, BLASTN analysis against the *M. truncatula* genome (Mt 4.0 v.1) [37] with a threshold E-value of 10^{-10} was conducted for all sequences from the **Dm**, **Sm** and **Lm** sets. This revealed a number of duplicated (i.e. corresponding to the same *M. truncatula* gene) markers in each set. The **Dm** set had the best combination of the overall number and distribution of markers and the minimal number of identified duplicated markers, which made it the most suitable “backbone” for artificial map construction.

In total, highly similar *M. truncatula* gene sequences were found for 96.7% of the pea marker genes. Genomic synteny analysis showed that the pea LGs I, II, III, V, and VI corresponded to the *M. truncatula* chromosomes 5, 1, 2/3, 7, and 2/6, respectively, consistent with previously reported results [5,20]; pea LG IV corresponded to regions of chromosomes 8, 5 and 4, and LG VII—to regions of chromosomes 8 and 4 of *M. truncatula*.

Further analysis showed that LGs from different studies were inverted relative to each other. For example, the order of the markers in the LGs I, III and VII in the **Lm** map was inverted relative to the corresponding LGs in the **Dm** and **Sm** maps; the marker order in the LGs IV and VI in the **Sm** map was also inverted relative to the corresponding LGs in the **Dm** and **Lm** maps. In order to construct a comprehensive database, LGs from all maps were oriented uniformly.

Markers from different datasets corresponding to the same transcript of *M. truncatula* were deemed “identical” and used as anchor markers for map joining. Since the locations of the markers in each set were determined using different mapping populations and, moreover, different joining algorithms for each map, the distances between the anchor markers varied in each set. To equalize map scales, the distances between markers belonging to the **Sm** and **Lm** sets were scaled to the dimension of the **Dm** set; then, markers lying between each pair of anchor markers were placed accordingly on the constructed “artificial” LGs.

Finally, by aligning the sequences from the **ESTm** set to the *M. truncatula* genome, correspondence between the NGS-based and EST-based markers was identified. All identified marker names were included in the description of the corresponding loci on the resulting PMD1 map. This pipeline (Fig 1) combining the locations of markers from different studies resulted in a more comprehensive albeit potentially less accurate map; small discrepancies were found in the interposition of some markers which are most probably the result of combining maps from three different studies that varied in marker saturation in some regions and were based on different genetic lines (possibly carrying minor chromosome rearrangements).

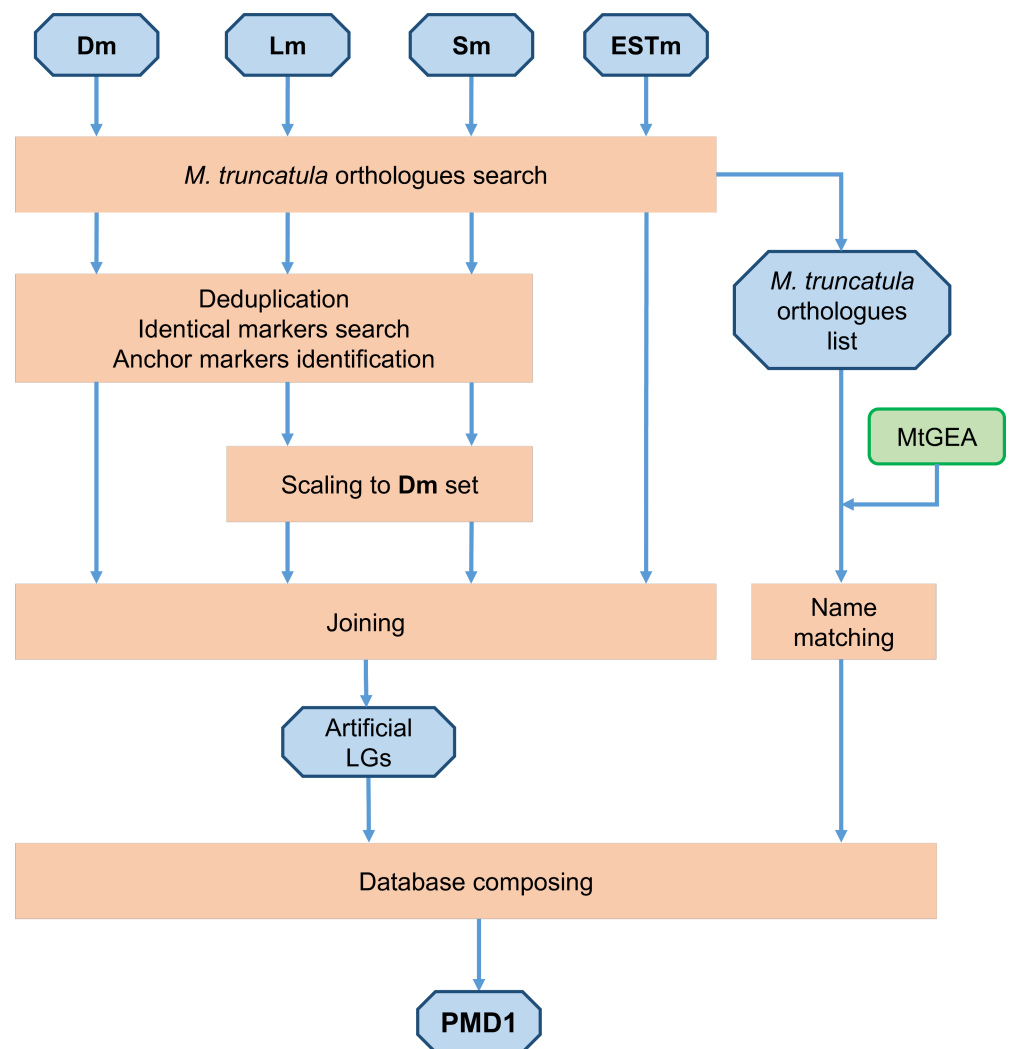


Fig 1. Pipeline of the PMD1 development. Initial and resulting datasets are placed in blue octagons. Operations are placed in peach-colored rectangles. Steps of PMD1 development are indicated by arrows. MtGEA means *Medicago truncatula* Gene Expression Atlas.

<https://doi.org/10.1371/journal.pone.0186713.g001>

Functional annotation of the markers included searching for homologous sequences in the *M. truncatula* genome and identifying the relevant entries (referred to as “Name matching” in Fig 1) in the *M. truncatula* Gene Expression Atlas (MtGEA) [55,56].

The artificial LGs construction was performed using an original program developed within the frame of this study in Visual Basic for Applications [57]. Visualization of the database content was conducted using D3.js, a JavaScript library for manipulating documents based on [58].

PMD2 construction

For PMD2 construction, the markers positions, corresponding transcript names, references to previously developed markers [15,17,19,20,24,25,34–36,38–51], *M. truncatula* homologous sequences for the ACm set and information concerning the quality of some markers were taken from the supplementary material data files of Tayeh and colleagues [27] and processed (Fig 2).

As gene and QTL mapping studies in pea had been conducted by different research groups using independently developed markers [19,20,24,25,34–36,38–48,53], the correspondence between these markers and the ACm set required examination. Table of correspondence between markers from the ACm, Dm, Sm sets and markers from [20] was extracted from the supplementary material data files of [27]. Markers found to be developed from the same sequence were labeled “identical” (as in PMD1). Comparative analysis of the markers showed a high number of duplications in [26] with up to 6 distinct markers corresponding to a single marker from [27] (consistent with the analysis carried out for markers from Lm set during PMD1 development). To avoid complicating the database data from [26] was not included in PMD2.

For marker sets from [19,34–36,38–48,53], comparative analysis between marker reference sequences (EST or nucleotide) and transcript sequences from the ACm set was performed

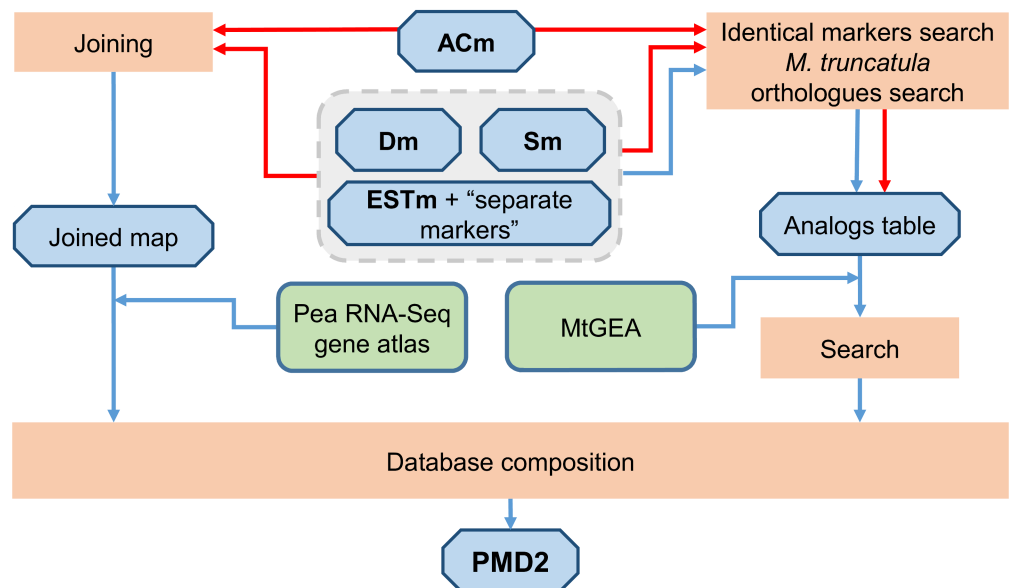


Fig 2. Pipeline of the PMD2 development. Initial and resulting datasets are placed in blue octagons. Operations are placed in peach-colored rectangles. Steps performed previously by Tayeh and colleagues in 2015 and described in [27] are indicated by red arrows. Steps performed in the course of the present study are indicated by blue arrows. MtGEA—*Medicago truncatula* Gene Expression Atlas.

<https://doi.org/10.1371/journal.pone.0186713.g002>

using the BLASTN algorithm (construction of “Analogous table” in Fig 2). Markers with a single hit with more than 95% coverage were marked as “identical” (3411 markers in total corresponding to 1586 distinct transcripts in PMD2).

The similarity between marker sequences [19,20,24,25,34–36,38–48,53] and *M. truncatula* sequences was assessed using the BLASTN algorithm (construction of “Analogous table” in Fig 2). Based on the results of this analysis, correspondence to the relevant *M. truncatula* genes was found for 75.8% of the investigated sequences. Functional annotation of the markers also included identifying the relevant entries in the *M. truncatula* Gene Expression Atlas (MtGEA) [55,56]. The correspondence between all genic markers from PMD2 and the expression profiles of the corresponding *M. truncatula* transcripts available in MtGEA was found for 61% of the sequences.

The functional annotation of markers from the **ACm** set includes information about the expression profile of the corresponding transcript from Pea RNA-seq Atlas available online at <http://bios.dijon.inra.fr/FATAL/cgi/pscaml.cgi> [12].

Visualization of the database content was conducted using D3.js, a JavaScript library for manipulating documents based on [58]. The databases have been tested on Mozilla Firefox (version 49.0.1) and Google Chrome (version 53.0.2785.143) browsers.

Plant growth conditions and gene mapping

The pea mutant line E135F, obtained after ethyl methanesulfonate (EMS) treatment of cv. Sparkle seeds [59] and carrying the recessive allele of *sym13* gene [60] corresponding to manifestation of the mutant phenotype (white nodules not capable of nitrogen fixation, so-called Fix^- phenotype), was crossed with the multiply marked line JI73 (formerly NGB1238) characterized by normal nodulation. After self-pollination of the resultant F_1 plants, the mapping population was obtained. For analysis of the symbiotic phenotype, F_2 plants were grown in 5 L plastic pots containing quartz sand with mineral nutrition lacking combined nitrogen [61] and inoculated with *Rhizobium leguminosarum* bv. *viciae* strain RCAM 1026 [62]. After 28 days of growth, the plants were taken from the pots and the phenotype of root system was examined. In total, 72 plants were analyzed (18 Fix^- and 54 Fix^+ , Chi-square for 1:3 = 0.00, p-value = 1.00).

The pea mutant line RisFixQ obtained after EMS mutagenesis of cv. Finale seeds [63] carries the recessive allele of *sym27* gene causing formation of non-fixing, prematurely senescent nodules [64]. The *Sym27* mapping population consisted of 50 F_2 plants obtained by crossing of the RisFixQ (*sym27*) line and the multiply marked line JI73 (NGB1238), followed by self-pollination of F_1 plants. The symbiotic phenotypes of F_2 plants were assessed by analyzing F_3 families originating from each F_2 plant (10 plants from each F_3 family were planted per one 5 L plastic pot and grown in the same conditions as for *Sym13* mapping experiment). At the 28th day after planting and inoculation, the root system phenotype was examined; families containing only plants with normal nodules were scored as descendants of *Sym27* F_2 plants, families containing only plants with defective nodules were scored as descendants of *sym27* F_2 plants, and families containing plants with either normal or defective nodules were scored as descendants of F_2 plants heterozygous for *Sym27*. Thus, the allelic state of *Sym27* was analyzed as a co-dominant marker (12 +/+, 24 +/-, 11 -/-, Chi-square for 1:2:1 = 0.06; p-value = 0.97).

DNA was extracted from the plant leaves according to [65] with slight modifications. RNA was extracted from nodules of pea mutant line E135F collected on the 21st day after inoculation. The growth conditions of plants used for cDNA analysis and protocols of RNA extraction and cDNA synthesis were described previously [66,67]. PCR was performed in iCycler (Bio-Rad, USA) or Dyad (Bio-Rad, USA) with use of the ScreenMix-HS kit (Evrogen, Russia) in the

conditions as follows: 95°C (5 min), 35 x [95°C (30 sec), 58–60°C (various for different primer pairs) (30 sec), 72°C (1 min)], 72°C 5 min. Primer design was performed with help of the Primer-BLAST [68]. For CAPS marker analysis, the proper restriction enzyme for SNP site recognition was chosen using dCAPSfinder 2.0 (<http://helix.wustl.edu/dcaps/dcaps.html>) [69], or was extracted from the list of available transcriptome-based pea markers for the lines Sparkle (parent of E135F (*sym13*)), Finale (parent of RisFixQ (*sym27*)) and JI73 (= NGB1238) [28]. Restriction digest of the PCR product was performed with use of Fermentas enzymes (Thermo Fischer Scientific, USA). Information regarding PCR primers and restriction enzymes is listed in S1 and S2 Tables. Primers for amplification of the candidate gene *Sen1* were designed on the basis of transcript GDTM01047803 found in pea nodule transcriptome assembly [11]: PsSen1_fw1: 5′ -TAAACAGATCAATCAAGCATTCATG-3′, and PsSen1_rv1: 5′ -AT TGGTTCAACATGAAGTATACG-3′.

The genetic linkage maps were constructed using JoinMap 4.1 software with default parameters [70]. For visualization of the maps, the MapChart program [71] was used.

The sequence of pea *Sen1* determined for cv. Sparkle is deposited in GenBank under the accession number KY888171.

The molecular biology procedures were performed using equipment of the Core Center “Genomic Technologies, Proteomics and Cell Biology” in ARRIAM, Saint-Petersburg, Russia.

Results and discussion

Database description

PMD resource was developed to provide information about pea gene-specific markers in a simple and convenient format. Both databases allow selecting linkage groups and markers along with searching for specific markers with the “search” function. For each LG, a specific region can be quickly selected and the scale can be adjusted. Specific marker can be selected by clicking on the square adjacent to the marker. Marker selection evokes a table containing information about the name of the marker, LG, literature references and the sequence of the transcript. Additionally, the table provides information about the homologous sequence from *M. truncatula* with reference to the dedicated transcript page in the Phytozome resource [72], and information about the expression profile of the *M. truncatula* transcripts in the MtGEA [56]. Note that for transcripts corresponding to a large number of entries in MtGEA, PMD returns the first record.

In total, PMD1 comprised 2484 markers: 336 in LG I, 328 in LG II, 404 in LG III, 317 in LG IV, 331 in LG V, 304 in LG VI and 464 in LG VII. The marker distribution across the artificial LGs is shown in Fig 3. For two or more markers representing the same pea transcript, all marker names were assigned to one locus in PMD. The average distance between markers was

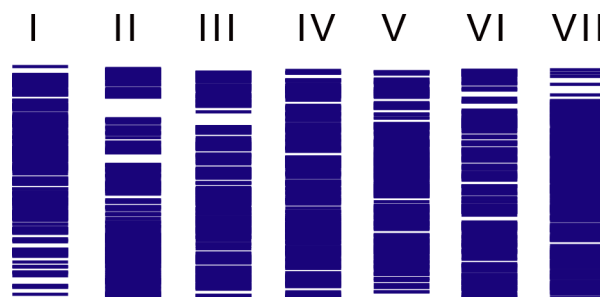


Fig 3. Marker distribution across the artificial LGs presented in PMD1.

<https://doi.org/10.1371/journal.pone.0186713.g003>

0.6 cM 98.8% of the markers were spaced less than 4 cM apart. Most of the discrepancies resulted from map joining (see [Materials and Methods](#); PMD1 development) are not exceed 3 cM and therefore are not critical for primary mapping. Nevertheless, markers with a dubious location are colored red in PMD1 and should be used cautiously in fine mapping studies.

The later version, PMD2, contains a total of 15944 pea markers and has a few advanced features compared to PMD1. The linkage groups in PMD2 are arranged horizontally ([Fig 4](#)) for better representation of markers information. In PMD2, an advanced “search” function is implemented making it possible to search not only for a marker name (even incomplete), but also for a *M. truncatula* gene identifier. With *M. truncatula* gene identifier as a query the search is performed for a complete match. For identifiers not found in the database, the identifier of a nearest (calculated using *M. truncatula* genome map) gene present in PMD2 is shown in the output, a useful feature for researchers possessing only the information about the homology of investigated gene to *M. truncatula* genes.

PMD2 includes the information about “identical” markers and marker type (gene-based, AFLP, SSR, morphological, protein, RAPD, SNP). For some markers which were marked as “identical”, but had not been included in joined LGs in [27], only information about “identity” and literature reference is presented in PMD2. According to Tayeh and colleagues [27], some markers displayed a high level of segregation distortion and, therefore, were labeled as “HLSO” in PMD2. Moreover, by searching for *M. truncatula* sequences homologous to pea transcripts from [12], several highly similar sequences not located in a conserved syntenic block were revealed [27]. Such possible “not real orthologues” of *M. truncatula* are marked as “NRO” in PMD2. Also, for some markers labeled as “identical”, different *M. truncatula* sequences were revealed as the closest homologues, perhaps due to minute differences in hit scores obtained in BLAST search. Such markers are marked as “DH” in PMD2.

All information contained in PMD1 and PMD2 is available at www.peamarker.arriam.ru. Detailed instructions on the use of PMD1 and PMD2 are presented in the “Help” section at www.peamarker.arriam.ru.

Comparative analysis of the marker positions in PMD1 and PMD2

Marker placement in the artificial LGs developed for PMD1 was compared to that in LGs presented in [27]. The analysis was carried out only for markers present in both PMD1 and PMD2. For all artificial LGs, the average number of markers with altered positions did not exceed 5%. Most of these markers were placed at the same site, possibly shifted by one or two markers in comparison with markers from [27], due to the different mapping scales in different studies, which does not strongly affect mapping success.

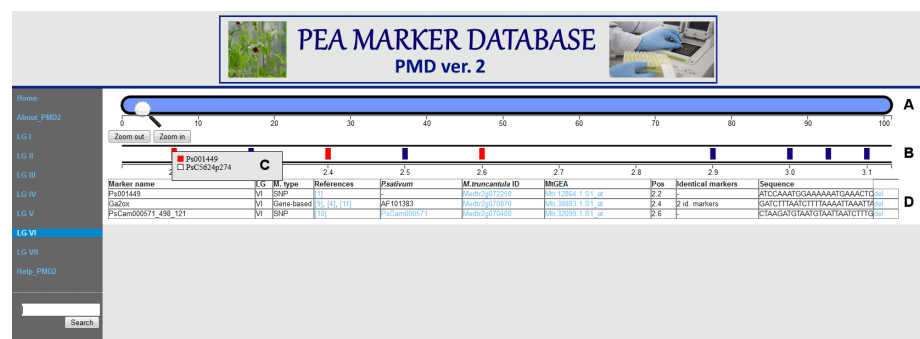


Fig 4. Interface of PMD2. (A) General view of selected LG. (B) Scaled region of selected LG. Previously selected loci are colored in red. (C) Markers situated in selected locus. (D) Table containing information about selected markers.

<https://doi.org/10.1371/journal.pone.0186713.g004>

Although LGs IV, V and VI in PMD2 were found to be inverted in comparison with PMD1, it was decided to retain the order of the PMD2 markers as in the original paper of Tayeh and colleagues [27].

Validation and applications

Mapping of *Sym13* gene. To test the usability of the PMD, we performed precise mapping of pea symbiotic gene *Sym13* using the gene-based markers developed in accordance to PMD1. Earlier, *Sym13* was shown to be genetically linked to allozyme markers *Skdh* and *Est-2* [60], which position is now considered as the middle of linkage group VII [20]. Therefore we (i) selected four markers from the middle part of LG VII, i.e. spanning the region from 110 cM to 140 cM according to PMD1 (S1 Table), (ii) investigated the polymorphism of the marker sequences using transcriptomic RNAseq datasets for Sparkle (parent of E135F (*sym13*)) and JI73 (= NGB1238) [28] (see **Materials and Methods** for details), (iii) found SNP sites suitable for CAPS marker design and made PCR primers flanking these sites (S1 Table). Segregation of the CAPS markers was then tested on F₂ population, along with analysis of the gene-based marker designed for pea homologue of *M. truncatula* DEFECTIVE IN NITROGEN FIXATION 2 (*DNF2*), which was initially chosen as a candidate for *Sym13* gene (S1 Table). As a result, pea *Sym13* was placed between genic markers *PsC5588p480* and *PsC908p622*, apart from *Dnf2* (Fig 5). The positions of all markers on the resulted map are in agreement with PMD1 markers order.

The successful localization of *Sym13* allowed us to carry out the search for candidate gene in *M. truncatula* genome. Since the markers flanking *Sym13* correspond to *Medtr4g096700* and *Medtr4g094252* genes, all genes in between were considered as possible candidates. Note that, as some recombination events were detected between *Dnf2* marker and *Sym13*, *DNF2* was excluded from the candidate genes list. Among the candidates, the gene *Medtr4g094335* attracted our attention as it was a homologue of known symbiotic gene *STATIONARY ENDO-SYMBIONT NODULE 1 (SEN1)* of *Lotus japonicus* (Regel.) K. Larsen [73], and, according to Pea RNA-seq Atlas (S3 Table), the expression of the corresponding pea gene was nodule-specific. We sequenced the pea *SEN1* homologue in E135F (*sym13*) and wild-type line Sparkle and found a nucleotide substitution c.C197T leading to amino acid change S66L in the case of

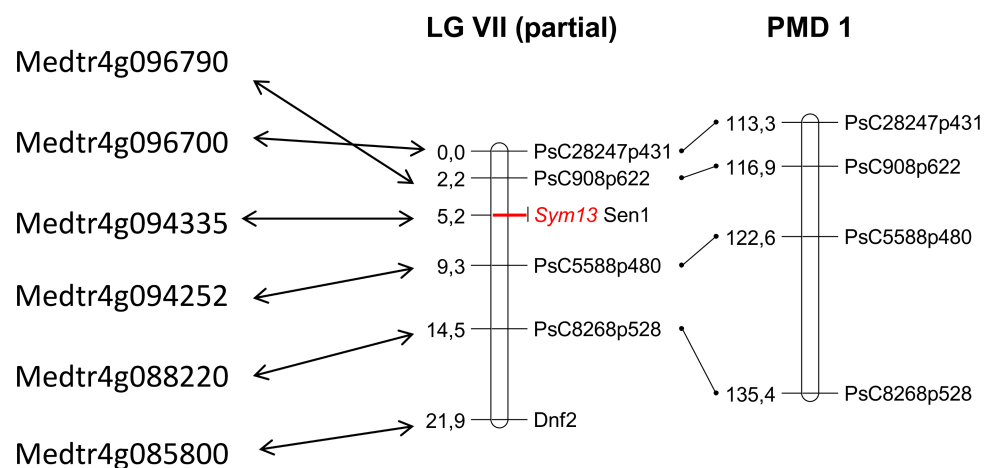


Fig 5. Localization of *Sym13* gene in relation to gene-based markers. From left to right: names of *M. truncatula* genes homologous to pea markers used for *Sym13* mapping, genetic map containing *Sym13*, artificial map from PMD1.

<https://doi.org/10.1371/journal.pone.0186713.g005>

mutant, which was predicted to be potentially damaging for protein function according to SIFT program [74]. Sequencing of the corresponding transcript on cDNA identified the same substitution, as well as the absence of introns, in pea gene *Sen1*, which is consistent with structure of *L. japonicus SEN1* gene and predicted structure of *M. truncatula* homologue *Medtr4g094335*. Thus, use of PMD for marker selection and candidate gene search in *M. truncatula* allowed us to infer that pea *Sym13* most likely encodes symbiosis-specific transporter of iron ions, as the presumable orthologous *SEN1* gene in *L. japonicus* does; however, mutant phenotype complementation studies are required to support the proposition about orthology of pea *Sym13* and *L. japonicus SEN1*.

Mapping of *Sym27* gene. As a second case study aimed at testing the features and facilities of PMD1 and PMD2, the pea symbiotic gene *Sym27* was fine-mapped in LG V based on the information provided in the database. Previously, the *Sym27* locus was roughly localized in LG V in relation to the gene-based markers *Pgd*, *Pme1* and *Met2* [53]. In the present study, we (i) identified *M. truncatula* genes homologous to the specified markers, (ii) found adjacent *M. truncatula* genes for which homologous pea markers were listed in PMD, (iii) outlined the region of LG V where, according to PMD, *Sym27* is located, (iv) selected markers lying 2–5 cM apart, and (v) adapted the selected markers for our mapping population (i.e., identified marker sequence variants between the parental lines of the mapping population and designed CAPS markers) and finally used them for *Sym27* fine mapping (S2 Table).

As a result of mapping, the *Sym27* gene was localized in relation to nine gene-based markers between *Ps001440* and *Met2*, implying that the *Sym27* homologue in *M. truncatula* is located between the corresponding *Medtr7g058640* and *Medtr7g061018* genes (Fig 6). Note that, apart from LG V being inverted in PMD1 compared with PMD2, the order of markers on our genetic map was in good agreement with the “artificial” PMD1 and “real” PMD2 (Fig 6). According to the most recent *M. truncatula* genome release Mt4.0v1 [37], this region contains 121 genes, of which 10 can be considered promising candidates based on their nodule-specific expression, involvement in nitrogen assimilation or microbial hosting processes in the plant tissues (S4 Table). By contrast to *Sym13* candidate gene search, none of the *Sym27* candidate genes were previously reported as symbiotic, which complicates the analysis but pointing at the possible novelty of the *Sym27* gene function in nitrogen-fixing symbiosis.

Conclusion

The increasing amount of genetics and genomics data obtained for model and agriculturally important plants provides a reliable basis for using up-to-date methods of molecular biology and genetics like MAS and GAB for crop breeding. Identification of particular genes responsible for desirable agricultural traits is usually performed via genetic mapping followed by candidate gene search, and such mapping is best carried out using gene-based molecular markers. Development of molecular markers leads to the construction of highly saturated genetic maps, which are essential tools for undertaking MAS and GAB. However, the usefulness of genetic and genomic data is highly dependent on availability of resources combining different sorts of such information.

This work was an attempt to integrate a large amount of information for pea gene-based markers into one database with a clear and user-friendly interface. Since the pea genome assembly is not available yet, comparative analysis between *P. sativum* and *M. truncatula* sequences is routinely used (for example, when searching for exon-intron junctions), making usability the defining and most important trait of any marker database. The Pea Marker Database (PMD) is a convenient tool that, as was demonstrated in the case study, facilitates marker development and gene mapping in pea. Indeed, combining the data on marker location that

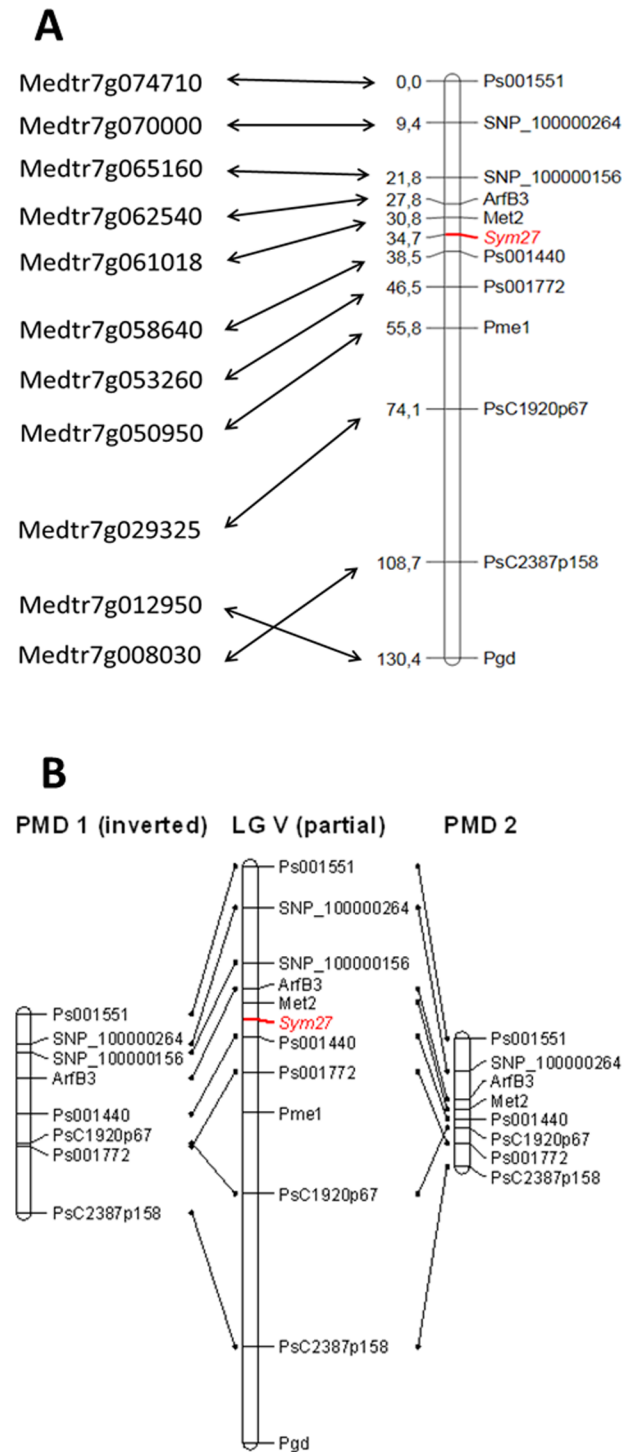


Fig 6. Localization of *Sym27* gene in relation to gene-based markers. A—*M. truncatula* genes homologous to pea markers used for *Sym27* mapping (right) and pea genetic map with *Sym27* position (left). B—Order of markers on the resulting map (center), PMD1 (left) and PMD2 (right).

<https://doi.org/10.1371/journal.pone.0186713.g006>

are collected in PMD with recently developed set of potential RNAseq-based markers [28] for some pea genotypes allowed us to identify the prominent candidate gene *Sen1* for pea

symbiotic gene *Sym13* described ca. 30 years ago [60], a result important to further the understanding of the molecular mechanisms underlying nodule senescence in pea [66,67]. The fine mapping of the pea symbiotic gene *Sym27* based on the information provided in the PMD allowed us to narrow the list of possible candidate genes for *Sym27* to approximately 10. Thus, PMD will be useful for geneticists and breeders alike, in particular, for mapping newly obtained mutations, for selecting appropriate markers in accordance with breeding programs, and for studying pea genetic diversity using particular marker sequences.

Supporting information

S1 Table. CAPS markers used for *Sym13* genetic mapping.

(DOCX)

S2 Table. CAPS markers used for *Sym27* genetic mapping.

(DOCX)

S3 Table. List of *M. truncatula* genes located between *Medtr4g092760* and *Medtr4g096790*, including possible *Sym13* candidate genes.

(XLSX)

S4 Table. List of *M. truncatula* genes located between *Medtr7g058640* and *Medtr061018*, including possible *Sym27* candidate genes.

(XLSX)

Acknowledgments

We are grateful to Prof. Jari Valkonen (University of Helsinki, Helsinki, Finland) for critical reading of the manuscript and MSc. Jaroslava Fedorina (ARRIAM, Saint-Petersburg, Russia) for assistance in case study experiments. We thank Dr. Alexey Borisov and Dr. Viktor Tsyganov (ARRIAM, Saint-Petersburg, Russia) for help in obtaining the *Sym27* mapping population, and Msc. Tatiana Serova (ARRIAM, Saint-Petersburg, Russia) for providing cDNA synthesized from RNA of E135F (*sym13*) nodules.

Author Contributions

Conceptualization: OAK IAT VAZ.

Data curation: OAK AIZ SSB.

Formal analysis: AMA SSB.

Funding acquisition: OAK IAT VAZ.

Investigation: OAK AIZ ASS VAZ.

Methodology: OAK SSB.

Project administration: VAZ.

Resources: IAT VAZ.

Software: SSB.

Supervision: IAT VAZ.

Validation: AIZ AMA.

Visualization: AMA ASS.

Writing – original draft: OAK VAZ.

Writing – review & editing: AMA ASS IAT VAZ.

References

1. Collard BCY, Mackill DJ. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. *Philos Trans R Soc Lond B Biol Sci. The Royal Society*; 2008; 363: 557–572. <https://doi.org/10.1098/rstb.2007.2170> PMID: 17715053
2. Bohra A, Pandey MK, Jha UC, Singh B, Singh IP, Datta D, et al. Genomics-assisted breeding in four major pulse crops of developing countries: present status and prospects. *Theor Appl Genet. Springer Berlin Heidelberg*; 2014; 127: 1263–1291. <https://doi.org/10.1007/s00122-014-2301-3> PMID: 24710822
3. Varshney RK. Gene-Based Marker Systems in Plants: High Throughput Approaches for Marker Discovery and Genotyping. *Molecular Techniques in Crop Improvement. Dordrecht: Springer Netherlands*; 2010. pp. 119–142. https://doi.org/10.1007/978-90-481-2967-6_5
4. FAOSTAT [Internet]. Available: <http://www.fao.org/faostat/en/%7B#%7Dhome>
5. Smýkal P, Aubert G, Burstin J, Coyne CJ, Ellis NTH, Flavell AJ, et al. Pea (*Pisum sativum* L.) in the Genomic Era. *Agronomy. Molecular Diversity Preservation International*; 2012; 2: 74–115. <https://doi.org/10.3390/agronomy2020074>
6. Borisov AY, Danilova TN, Koroleva TA, Kuznetsova E V, Madsen L, Mofett M, et al. Regulatory genes of garden pea (*Pisum sativum* L.) controlling the development of nitrogen-fixing nodules and arbuscular mycorrhiza: A review of basic and applied aspects. *Appl Biochem Microbiol. Nauka/Interperiodica*; 2007; 43: 237–243. <https://doi.org/10.1134/S0003683807030027>
7. Royal Botanic Gardens, Kew: Plant DNA C-values database [Internet]. Available: <http://data.kew.org/cvalues/>
8. Macas J, Neumann P, Navrátilová A. Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics. BioMed Central*; 2007; 8: 427. <https://doi.org/10.1186/1471-2164-8-427> PMID: 18031571
9. Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM. Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. *BMC Genomics. 2011*; 12: 227. <https://doi.org/10.1186/1471-2164-12-227> PMID: 21569327
10. Sudheesh S, Sawbridge TI, Cogan NOI, Kennedy P, Forster JW, Kaur S. De novo assembly and characterisation of the field pea transcriptome using RNA-Seq. *BMC Genomics. 2015*; 16: 611. <https://doi.org/10.1186/s12864-015-1815-7> PMID: 26275991
11. Zhukov VA, Zhermakov AI, Kulaeva OA, Ershov NI, Borisov AY, Tikhonovich IA. De Novo Assembly of the Pea (*Pisum sativum* L.) Nodule Transcriptome. *Int J Genomics. Hindawi Publishing Corporation*; 2015; 2015: 1–11. <https://doi.org/10.1155/2015/695947> PMID: 26688806
12. Alves-Carvalho S, Aubert G, Carrère S, Cruaud C, Brochot A-L, Jacquin F, et al. Full-length *de novo* assembly of RNA-seq data in pea (*Pisum sativum* L.) provides a gene expression atlas and gives insights into root nodulation in this species. *Plant J. 2015*; 84: 1–19. <https://doi.org/10.1111/tpj.12967> PMID: 26296678
13. Vilmorin PD, Bateson W. A Case of Gametic Coupling in *Pisum*. *Proc R Soc B Biol Sci. The Royal Society*; 1911; 84: 9–11. <https://doi.org/10.1098/rspb.1911.0040>
14. Wellensiek SJ. Genetic Monograph on *Pisum* [Internet]. Martinus Nijhoff; 1925. Available: <https://books.google.ru/books?id=hjYOWAACAAJ>
15. Weeden NF, Ellis THN, Timmerman-Vaughan GM, Swiecicki WK, Rozov SM, Berdnikov VA. A consensus linkage map for *Pisum sativum*. *Pisum Genet. 1998*; 30: 4.
16. Ellis THN, Poyser SJ. An integrated and comparative view of pea genetic and cytogenetic maps. *New Phytol. Blackwell Science Ltd*; 2002; 153: 17–25. <https://doi.org/10.1046/j.0028-646X.2001.00302.x>
17. Laucou V, Haurogné K, Ellis N, Rameau C. Genetic mapping in pea. 1. RAPD-based genetic linkage map of *Pisum sativum*. *TAG Theor Appl Genet. Springer-Verlag*; 1998; 97: 905–915. <https://doi.org/10.1007/s001220050971>
18. Rameau C, Dénoue D, Fraval F, Haurogné K, Josserand J, Laucou V, et al. Genetic mapping in pea. 2. Identification of RAPD and SCAR markers linked to genes affecting plant architecture. *TAG Theor Appl Genet. Springer-Verlag*; 1998; 97: 916–928. <https://doi.org/10.1007/s001220050972>
19. Aubert G, Morin J, Jacquin F, Lordon K, Quillet MC, Petit A, 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.* 2006; 112: 1024–1041. <https://doi.org/10.1007/s00122-005-0205-y> PMID: 16416153
20. Bordat A, Savoie V, Nicolas M, Salse J, Chauveau A, Bourgeois M, et al. Translational Genomics in Legumes Allowed Placing *In Silico* 5460 Unigenes on the Pea Functional Map and Identified Candidate Genes in *Pisum sativum* L. Andrews BJ, editor. *G3* & (#)58; Genes|Genomes|Genetics. 2011; 1: 93–103. <https://doi.org/10.1534/g3.111.000349> PMID: 22384322
 21. Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, et al. Comparative mapping between *Medicago sativa* and *Pisum sativum*. *Mol Genet Genomics.* 2004; 272: 235–246. <https://doi.org/10.1007/s00438-004-1055-z> PMID: 15340836
 22. Kaur S, Pembleton LW, Cogan NOI, Savin KW, Leonforte T, Paull J, et al. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. *BMC Genomics.* 2012; 13: 104. <https://doi.org/10.1186/1471-2164-13-104> PMID: 22433453
 23. Yang T, Fang L, Zhang X, Hu J, Bao S, Hao J, et al. High-Throughput Development of SSR Markers from Pea (*Pisum sativum* L.) Based on Next Generation Sequencing of a Purified Chinese Commercial Variety. Wu S-B, editor. *PLoS One.* 2015; 10: e0139775. <https://doi.org/10.1371/journal.pone.0139775> PMID: 26440522
 24. Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, et al. Transcriptome sequencing for high throughput SNP development and genetic mapping in Pea. *BMC Genomics.* 2014; 15: 126. <https://doi.org/10.1186/1471-2164-15-126> PMID: 24521263
 25. Sindhu A, Ramsay L, Sanderson L-A, Stonehouse R, Li R, Condie J, et al. Gene-based SNP discovery and genetic mapping in pea. *Theor Appl Genet.* 2014; 127: 2225–2241. <https://doi.org/10.1007/s00122-014-2375-y> PMID: 25119872
 26. Leonforte A, Sudheesh S, Cogan NOI, Salisbury PA, Nicolas ME, Materne M, et al. SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). *BMC Plant Biol.* 2013; 13: 161. <https://doi.org/10.1186/1471-2229-13-161> PMID: 24134188
 27. Tayeh N, Aluome C, Falque M, Jacquin F, Klein A, Chauveau A, 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.* 2015; 84: 1257–1273. <https://doi.org/10.1111/tpj.13070> PMID: 26590015
 28. Zhernakov A, Rotter B, Winter P, Borisov A, Tikhonovich I, Zhukov V. Massive Analysis of cDNA Ends (MACE) for transcript-based marker design in pea (*Pisum sativum* L.). *Genomics data.* Elsevier; 2017; 11: 75–76. <https://doi.org/10.1016/j.gdata.2016.12.004> PMID: 28050346
 29. Boutet G, Alves Carvalho S, Falque M, Peterlongo P, Lhuillier E, Bouchez O, et al. SNP discovery and genetic mapping using genotyping by sequencing of whole genome genomic DNA from a pea RIL population. *BMC Genomics.* BioMed Central; 2016; 17: 121. <https://doi.org/10.1186/s12864-016-2447-2> PMID: 26892170
 30. Ma Y, Coyne CJ, Grusak MA, Mazourek M, Cheng P, Main D, et al. Genome-wide SNP identification, linkage map construction and QTL mapping for seed mineral concentrations and contents in pea (*Pisum sativum* L.). *BMC Plant Biol.* 2017; 17: 43. <https://doi.org/10.1186/s12870-016-0956-4> PMID: 28193168
 31. Welcome to the Public Portal for the University of Saskatchewan Pulse Crop Research Group | Know-Pulse [Internet]. Available: <http://knowpulse.usask.ca/portal/>
 32. CMap Legume Information System [Internet]. Available: <http://cmap.comparative-legumes.org/>
 33. Legume portal [Internet]. Available: http://www.thelegumeportal.net/pea%7B_%7Dmtr%7B_%7Dtranslational%7B_%7Dtoolkit
 34. Burstin J, Deniot G, Potier J, Weinachter C, Aubert G, Baranger A. Microsatellite polymorphism in *Pisum sativum*. *Plant Breed.* Blackwell Publishing Ltd; 2001; 120: 311–317. <https://doi.org/10.1046/j.1439-0523.2001.00608.x>
 35. Deulvot C, Charrel H, Marty A, Jacquin F, Donnadiou C, Lejeune-Hénaut I, et al. Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. *BMC Genomics.* 2010; 11: 468. <https://doi.org/10.1186/1471-2164-11-468> PMID: 20701750
 36. Hecht V, Foucher F, Ferrándiz C, Macknight R, Navarro C, Morin J, et al. Conservation of Arabidopsis flowering genes in model legumes. *Plant Physiol. American Society of Plant Biologists;* 2005; 137: 1420–1434. <https://doi.org/10.1104/pp.104.057018> PMID: 15778459
 37. Tang H, Krishnakumar V, Bidwell S, Rosen B, Chan A, Zhou S, et al. An improved genome release (version Mt4.0) for the model legume *Medicago truncatula*. *BMC Genomics.* 2014; 15: 312. <https://doi.org/10.1186/1471-2164-15-312> PMID: 24767513
 38. Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, et al. Gene-based sequence diversity analysis of field pea (*Pisum*). *Genetics.* Genetics Society of America; 2007; 177: 2263–2275. <https://doi.org/10.1534/genetics.107.081323> PMID: 18073431

39. Tayeh N, Bahrman N, Devaux R, Bluteau A, Prosperi J-M, Delbreil B, et al. A high-density genetic map of the *Medicago truncatula* major freezing tolerance QTL on chromosome 6 reveals colinearity with a QTL related to freezing damage on *Pisum sativum* linkage group VI. *Mol Breed*. 2013; 32: 279–289. <https://doi.org/10.1007/s11032-013-9869-1>
40. Armstead I, Donnison I, Aubry S, Harper J, Hortensteiner S, James C, et al. Cross-Species Identification of Mendel's I Locus. *Science* (80-). 2007; 315: 73. <https://doi.org/10.1126/science.1132912> PMID: 17204643
41. Choi H-K, Mun J-H, Kim D-J, Zhu H, Baek J-M, Mudge J, et al. Estimating genome conservation between crop and model legume species. *Proc Natl Acad Sci U S A*. National Academy of Sciences; 2004; 101: 15289–15294. <https://doi.org/10.1073/pnas.0402251101> PMID: 15489274
42. DeMason DA, Weeden NF. Two Argonaute 1 genes from pea. *Pisum Genet*. 2006; 38: 3–9.
43. Gilpin BJ, McCallum JA, Frew TJ, Timmerman-Vaughan GM. A linkage map of the pea (*Pisum sativum* L.) genome containing cloned sequences of known function and expressed sequence tags (ESTs). *Theor Appl Genet*. 1997; 95: 1289–1299. <https://doi.org/10.1007/s001220050695>
44. Page D, Aubert G, Duc G, Welham T, Domoney C. Combinatorial variation in coding and promoter sequences of genes at the Tri locus in *Pisum sativum* accounts for variation in trypsin inhibitor activity in seeds. *Mol Genet Genomics*. 2002; 267: 359–369. <https://doi.org/10.1007/s00438-002-0667-4> PMID: 12073038
45. Platten JD, Foo E, Elliott RC, Hecht V, Reid JB, Weller JL. Cryptochrome 1 contributes to blue-light sensing in pea. *Plant Physiol*. American Society of Plant Biologists; 2005; 139: 1472–1482. <https://doi.org/10.1104/pp.105.067462> PMID: 16244154
46. Prioul-Gervais S, Deniot G, Receveur E-M, Frankewitz A, Fourmann M, Rameau C, et al. Candidate genes for quantitative resistance to *Mycosphaerella pinodes* in pea (*Pisum sativum* L.). *Theor Appl Genet*. Springer-Verlag; 2007; 114: 971–984. <https://doi.org/10.1007/s00122-006-0492-y> PMID: 17265025
47. Sorefan K, Booker J, Haurogné K, Goussot M, Bainbridge K, Foo E, et al. MAX4 and RMS1 are orthologous dioxygenase-like genes that regulate shoot branching in *Arabidopsis* and pea. *Genes* {&} *Dev*. 2003; 17: 1469–1474. <https://doi.org/10.1101/gad.256603> PMID: 12815068
48. Verdier J, Kakar K, Gallardo K, Le Signor C, Aubert G, Schlereth A, et al. Gene expression profiling of *M. truncatula* transcription factors identifies putative regulators of grain legume seed filling. *Plant Mol Biol*. 2008; 67: 567–580. <https://doi.org/10.1007/s11103-008-9320-x> PMID: 18528765
49. Lorida K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, et al. Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). *Theor Appl Genet*. 2005; 111: 1022–1031. <https://doi.org/10.1007/s00122-005-0014-3> PMID: 16133320
50. Lejeune-Hénaut I, Hanocq E, Béthencourt L, Fontaine V, Delbreil B, Morin J, et al. The flowering locus Hr colocalizes with a major QTL affecting winter frost tolerance in *Pisum sativum* L. *Theor Appl Genet*. 2008; 116: 1105–1116. <https://doi.org/10.1007/s00122-008-0739-x> PMID: 18347775
51. Bourgeois M, Jacquin F, Cassecuelle F, Savoie V, Belghazi M, Aubert G, et al. A PQL (protein quantity loci) analysis of mature pea seed proteins identifies loci determining seed protein composition. *Proteomics*. 2011; 11: 1581–1594. <https://doi.org/10.1002/pmic.201000687> PMID: 21433288
52. National Center for Biotechnology Information [Internet]. Available: <https://www.ncbi.nlm.nih.gov/>
53. Zhukov VA, Kuznetsova EV, Ovchinnikova ES, Rychagova TS, Titov VS, Pinaev AG, et al. Gene-based markers of pea linkage group V for mapping genes related to symbioses. *Pisum Genet*. Pisum Genetics Association, Montana State University-Bozeman; 2007; 39: 19–25.
54. The Pea RNA-Seq gene atlas [Internet]. Available: <http://bios.dijon.inra.fr/FATAL/cgi/pscam.cgi>
55. He J, Benedito VA, Wang M, Murray JD, Zhao PX, Tang Y, et al. The *Medicago truncatula* gene expression atlas web server. *BMC Bioinformatics*. BioMed Central; 2009; 10: 441. <https://doi.org/10.1186/1471-2105-10-441> PMID: 20028527
56. The *Medicago truncatula* Gene Expression Atlas Project [Internet]. [cited 10 Dec 2016]. Available: <http://mtgea.noble.org/v3/>
57. Office VBA language reference [Internet]. Available: <https://msdn.microsoft.com/en-us/library/office/gg264383.aspx>
58. D3.js—Data-Driven Documents [Internet]. Available: <https://d3js.org/>
59. Kneen BE, LaRue TA. Induced symbiosis mutants of pea (*Pisum sativum*) and sweetclover (*Melilotus alba annua*). *Plant Sci*. 1988; 58: 177–182. [https://doi.org/10.1016/0168-9452\(88\)90007-6](https://doi.org/10.1016/0168-9452(88)90007-6)
60. Kneen BE, Larue TA, Hirsch AM, Smith CA, Weeden NF. sym 13-A Gene Conditioning Ineffective Nodulation in *Pisum sativum*. *Plant Physiol*. 1990; 94: 899–905. Available: <http://www.ncbi.nlm.nih.gov/pubmed/16667870> PMID: 16667870

61. Borisov AY, Rozov SM, Tsyganov VE, Morzhina E V, Lebsky VK, Tikhonovich IA. Sequential functioning of Sym-13 and Sym-31, two genes affecting symbiosome development in root nodules of pea (*Pisum sativum* L.). *Mol Gen Genet MGG*. 1997; 254: 592–598. Available: <http://www.ncbi.nlm.nih.gov/pubmed/9197420>
62. Afonin A, Sulima A, Zhernakov A, Zhukov V. Draft genome of the strain RCAM1026 *Rhizobium leguminosarum* bv. *viciae* [Internet]. *Genomics Data*. 2017. pp. 85–86. <https://doi.org/10.1016/j.gdata.2016.12.003> PMID: 28053873
63. Engvild KC. Nodulation and nitrogen fixation mutants of pea, *Pisum sativum*. *Theor Appl Genet*. 1987; 74: 711–713. <https://doi.org/10.1007/BF00247546> PMID: 24240329
64. Morzhina, Tsyganov, Borisov, Lebsky, Tikhonovich. Four developmental stages identified by genetic dissection of pea (*Pisum sativum* L.) root nodule morphogenesis. *Plant Sci*. 2000; 155: 75–83. Available: <http://www.ncbi.nlm.nih.gov/pubmed/10773342> PMID: 10773342
65. Rogers SO, Bendich AJ. Extraction of DNA from milligram amounts of fresh, herbarium and mummified plant tissues. *Plant Mol Biol*. 1985; 5: 69–76. <https://doi.org/10.1007/BF00020088> PMID: 24306565
66. Serova TA, Tikhonovich IA, Tsyganov VE. Analysis of nodule senescence in pea (*Pisum sativum* L.) using laser microdissection, real-time PCR, and ACC immunolocalization. *J Plant Physiol*. 2017; 212: 29–44. <https://doi.org/10.1016/j.jplph.2017.01.012> PMID: 28242415
67. Ivanova KA, Tsyganova A V, Brewin NJ, Tikhonovich IA, Tsyganov VE. Induction of host defences by *Rhizobium* during ineffective nodulation of pea (*Pisum sativum* L.) carrying symbiotically defective mutations *sym40* (PsEFD), *sym33* (PsIPD3/PsCYCLOPS) and *sym42*. *Protoplasma*. Springer Vienna; 2015; 252: 1505–1517. <https://doi.org/10.1007/s00709-015-0780-y> PMID: 25743038
68. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics*. 2012; 13: 134. <https://doi.org/10.1186/1471-2105-13-134> PMID: 22708584
69. Neff MM, Turk E, Kalishman M. Web-based primer design for single nucleotide polymorphism analysis. *Trends Genet*. 2002; 18: 613–615. Available: <http://www.ncbi.nlm.nih.gov/pubmed/12446140> PMID: 12446140
70. Stam P. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *Plant J*. Blackwell Science Ltd; 1993; 3: 739–744. <https://doi.org/10.1111/j.1365-313X.1993.00739.x>
71. Voorrips RE. MapChart: Software for the Graphical Presentation of Linkage Maps and QTLs. *J Hered*. 2002; 93: 77–78. Available: <http://dx.doi.org/10.1093/jhered/93.1.77> PMID: 12011185
72. Phytozome v12.0: Home [Internet]. Available: <https://phytozome.jgi.doe.gov/pz/portal.html>
73. Hakoyama T, Niimi K, Yamamoto T, Isobe S, Sato S, Nakamura Y, et al. The Integral Membrane Protein SEN1 is Required for Symbiotic Nitrogen Fixation in *Lotus japonicus* Nodules. *Plant Cell Physiol*. 2012; 53: 225–236. <https://doi.org/10.1093/pcp/pcr167> PMID: 22123791
74. Kumar P, Henikoff S, Ng PC. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nat Protoc*. 2009; 4: 1073–1081. <https://doi.org/10.1038/nprot.2009.86> PMID: 19561590