

# Geographical Gradient of the *eIF4E* Alleles Conferring Resistance to Potyviruses in Pea (*Pisum*) Germplasm

Eva Konečná<sup>1,3</sup>, Dana Šafářová<sup>2</sup>, Milan Navrátil<sup>2</sup>, Pavel Hanáček<sup>1,3</sup>, Clarice Coyne<sup>4</sup>, Andrew Flavell<sup>5</sup>, Margarita Vishnyakova<sup>6</sup>, Mike Ambrose<sup>7</sup>, Robert Redden<sup>8</sup>, Petr Smýkal<sup>9\*</sup>

**1** Department of Plant Biology, Mendel University in Brno, Brno, Czech Republic, **2** Department of Cell Biology and Genetics, Palacky University in Olomouc, Olomouc, Czech Republic, **3** CEITEC MENDELU, Mendel University in Brno, Brno, Czech Republic, **4** Western Regional Plant Introduction Station - USDA, Pullman, Washington, United States of America, **5** Division of Plant Sciences, University of Dundee at James Hutton Institute, Invergowrie, United Kingdom, **6** Vavilov Institute of Plant Industries, Saint Petersburg, Russian Federation, **7** John Innes Centre, Norwich, United Kingdom, **8** Australian Grains Genebank, Horsham, Victoria, Australia, **9** Department of Botany, Palacky University in Olomouc, Olomouc, Czech Republic

## Abstract

**Background:** The eukaryotic translation initiation factor 4E was shown to be involved in resistance against several potyviruses in plants, including pea. We combined our knowledge of pea germplasm diversity with that of the *eIF4E* gene to identify novel genetic diversity.

**Methodology/Principal findings:** Germplasm of 2803 pea accessions was screened for *eIF4E* intron 3 length polymorphism, resulting in the detection of four *eIF4E<sup>A-B-C-5</sup>* variants, whose distribution was geographically structured. The *eIF4E<sup>A</sup>* variant conferring resistance to the P1 PSbMV pathotype was found in 53 accessions (1.9%), of which 15 were landraces from India, Afghanistan, Nepal, and 7 were from Ethiopia. A newly discovered variant, *eIF4E<sup>B</sup>*, was present in 328 accessions (11.7%) from Ethiopia (29%), Afghanistan (23%), India (20%), Israel (25%) and China (39%). The *eIF4E<sup>C</sup>* variant was detected in 91 accessions (3.2% of total) from India (20%), Afghanistan (33%), the Iberian Peninsula (22%) and the Balkans (9.3%). The *eIF4E<sup>S</sup>* variant for susceptibility predominated as the wild type. Sequencing of 73 samples, identified 34 alleles at the whole gene, 26 at cDNA and 19 protein variants, respectively. Fifteen alleles were virologically tested and 9 alleles (*eIF4E<sup>A-1-2-3-4-5-6-7</sup>*, *eIF4E<sup>B-1</sup>*, *eIF4E<sup>C-2</sup>*) conferred resistance to the P1 PSbMV pathotype.

**Conclusions/Significance:** This work identified novel *eIF4E* alleles within geographically structured pea germplasm and indicated their independent evolution from the susceptible *eIF4E<sup>S1</sup>* allele. Despite high variation present in wild *Pisum* accessions, none of them possessed resistance alleles, supporting a hypothesis of distinct mode of evolution of resistance in wild as opposed to crop species. The Highlands of Central Asia, the northern regions of the Indian subcontinent, Eastern Africa and China were identified as important centers of pea diversity that correspond with the diversity of the pathogen. The series of alleles identified in this study provides the basis to study the co-evolution of potyviruses and the pea host.

**Citation:** Konečná E, Šafářová D, Navrátil M, Hanáček P, Coyne C, et al. (2014) Geographical Gradient of the *eIF4E* Alleles Conferring Resistance to Potyviruses in Pea (*Pisum*) Germplasm. PLoS ONE 9(3): e90394. doi:10.1371/journal.pone.0090394

**Editor:** João Pinto, Instituto de Higiene e Medicina Tropical, Portugal

**Received:** October 7, 2013; **Accepted:** January 30, 2014; **Published:** March 7, 2014

This is an open-access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the Creative Commons CC0 public domain dedication.

**Funding:** This work was supported by the Ministry of Agriculture IQ91A229 project. CC was supported by USDA (United States Department of Agriculture) project #5348-2100=017-00D. 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.

\* E-mail: petr.smykal@upol.cz

## Introduction

Crop genetic diversity is an important pre-requisite for improving crop traits through breeding, particularly since the presence of closely related and genetically uniform varieties provides an ideal genetic environment for disease epidemics to occur, as evidenced by several historical and also recent events (the 1846 potato blight in Ireland, the 1970 corn blight in the USA, or the 1999 wheat rust in Africa). Available crop genetic resources function as reservoirs of often yet undiscovered allelic variants that provide an opportunity for genetic improvement of a given cultivated species [1,2]. However, the identification of specific, often rare traits requires specific screening and testing of entire, often very large collections. This is a time- and resource-intensive process. This situation improves, however, once a respective underlying gene is identified, especially in the case of monogenic

traits. Various available genomic technologies [3–7] can be applied to uncover such variation. Such screening of wild and cultivated germplasm for allelic variation of identified resistance genes has been receiving increased attention [2,8], as it can be efficiently substituted for phenotypic characterization.

Pea is an ancient legume crop, originating and domesticated in the Middle East and Mediterranean regions, from where its cultivation spread to today's Russia, and westwards through the Danube Valley and/or through ancient Greece and Rome into Europe. Pea likewise spread eastward into Persia, India and China [9]. Archaeological evidence indicates that the first instance of pea dates back to 8000 B.C. [10] in the Near East. Later, during the Stone and Bronze Ages, the existence of pea was documented in Europe and India. From there, it entered China by the first century B.C. [11]. Wild *Pisum sativum* subsp. *elatius* and subsp.

*sativum* are found naturally in Europe, Northwestern Asia and in temperate Africa, while *P. fulvum* range is restricted to the Middle East. *Pisum abyssinicum* is found in Ethiopia and Yemen [12,13]. There are 25 larger sized germplasm collections preserving pea diversity, which collectively hold around 72,000 accessions, while an additional 27,000 accessions are maintained in 146 smaller collections around the world [14,15]. The diversity of these collections has been studied using both by morphological descriptors and agronomical traits as well as molecular markers (reviewed in Smýkal et al. [14]), and diversity core collections were formed thereafter [16–19]. These studies showed that although *Pisum* is a comparably small genus containing two or three species, it has a wide and structured diversity, showing a range of degrees of relatedness that reflect taxonomic identifiers, eco-geography and breeding gene pools [13,17].

Along with abiotic stresses, plant pathogens are a major detriment to agriculture and threaten global food security. The use of genetic resistance is considered to be the most effective and sustainable strategy for controlling plant pathogens in agricultural practice, as it is environmentally friendly, targets specific pathogens, and provides reliable protection without additional labor or material costs [20]. Long before plants were domesticated and grown as monocultures, plant pathogens were co-evolving with wild plants growing in mixed-species communities. Evolution has continued to occur within domesticated plants growing as selected genotypes in denser populations than in the wild. Furthermore, the domestication of wild plants has distributed crops far from their places of origin [21,22], and the introduction of pathogens eventually accompanied this distribution. This co-evolutionary process shaped both plants and their pathogens, including viruses [23–25].

*Pea seed borne mosaic virus* (PSbMV), member of the genus *Potyvirus*, has been since identification in Czechoslovakia [26] reported worldwide and causes serious yield losses in a broad spectra of legumes including the most economically important like pea, lentil, faba bean, and chickpea [27]. The virus causes various symptoms, depending on the host and virus isolate/pathotype, such as the downward rolling of leaflets, the transient clearing and swelling of leaf veins, chlorotic mosaics, stunting, and delayed flowering. PSbMV is transmitted between plants in a non-persistent manner by aphids and then infects seeds [28]. Due to its seed-borne transmission, PSbMV presents a serious phytosanitary risk both for germplasm maintenance [29] and seed production.

The pea genome contains two virologically defined clusters of recessive resistance genes that are responsive to various potyviruses. One cluster (on pea linkage group II) includes *bcm*, *cym-1*, *mo*, *sbm-2* and *sbm-3* loci, conferring resistance to *Bean common mosaic virus* (BCMV), *Clover yellow vein virus* (CIYVV), *Bean yellow mosaic virus* (BYMV-S pathotype) and *Pea seed-borne mosaic virus* (PSbMV, pathotype P2) respectively. The second cluster (on pea linkage group VI) includes *cym-2*, *wlv* and *sbm-1* loci, conferring resistance to *Clover yellow vein virus* (CIYVV), *Bean yellow mosaic virus* (BYMV-W pathotype) and the *Pea seed-borne mosaic virus* (PSbMV, pathotype P1), respectively [30,31]. With exception of *cym-2* and *sbm-1* loci, shown to be identical [32,33], it is not clear if these clusters are closely linked separate genes or the same gene with alleles of different specificity.

The pea *eIF4E* gene was identified as a susceptibility factor corresponding to the recessive resistance gene, *sbm-1* locus, and a homologue *eIF(iso)4E* is presumed to be *sbm-2* locus [31–35]. PSbMV has been well-studied genomically [36–38] and viral P3-6K1 and VPg proteins have been identified as PSbMV determinants [30] responsible for physical interaction with host

*eIF4E* or *eIF(iso)4E* proteins and critical for viral infection. Studies on pepper and *Arabidopsis* suggest that potyviruses may selectively use either *eIF4E* and/or *eIF(iso)4E* proteins to achieve infection [39,40].

Recessive resistance of pea to PSbMV corresponds with the matching-allele model, explaining the interaction between potyviruses and plant hosts [41]. The mutant resistant allele of pea *eIF4E* (named *eIF4E<sup>A</sup>* in this study) differs from its wild-type (sensitive) counterpart by five non-conservative amino acids [32,31,42]. Natural variation and functional analysis have revealed evidence of co-evolution between *eIF4E* and potyviral VPg [3,43,44]. The *eIF4E* allelic diversity has been systematically screened in various crop collections, such as pepper [5,45,46], melon [4], tomato [47] and barley [3]. A possible link between the spread of potyviruses and the origin of agriculture was demonstrated by Gibbs et al. [48], who clearly showed that the human-mediated spread of crop hosts was followed by the further diversification of viruses.

In this study, we have combined our knowledge of pea germplasm diversity with that of the *eIF4E* resistance gene and systematically screened 2803 accessions with known geographical origins including *Pisum* species with the aim of identifying allelic diversity present within the broader pea germplasm held in *ex situ* collections.

## Results

### Geographical distribution of *eIF4E* variants

A total of 2803 pea accessions were screened for *eIF4E* intron 3 length polymorphism using two sets of primer combinations, resulting in the detection of four respective *eIF4E<sup>A-B-C-S</sup>* variants (Table 1, Fig. 1). The first corresponded to the already known *eIF4E<sup>A</sup>* resistance [31] variant (amplified 243 bp fragment with primer combination A and 536 bp with primer combination B), the second to the susceptible *eIF4E<sup>S</sup>* [31] (293 and 586 bp fragments), the third to the novel *eIF4E<sup>B</sup>* (293 and 536 bp fragments) and the fourth to the novel *eIF4E<sup>C</sup>* (293 and 592 bp fragments). The resistance *eIF4E<sup>A</sup>* variant was found in seven accessions from Ethiopia-Sudan, fourteen accessions from India, five accessions from Pakistan-Nepal and one accession each from China, Russia and Afghanistan. In addition, this variant was detected in fifteen USA and Canadian, two South American and eleven European modern varieties or breeding lines, which have the Ethiopian line PI193835 in their pedigree. In total, the *eIF4E<sup>A</sup>* variant was detected in 53 accessions, which represent 1.9% of those tested. The *eIF4E<sup>B</sup>* variant was abundant in accessions of Ethiopian (29%), Afghan (23%), Nepal-Indian (20%), Caucasus (6.4%) and particularly Chinese (36.5%) origin (Table 2, Fig. 2). Moreover, it was found in five (e.g. 20%) accessions of South American origin. In contrast, this variant was under-represented in landraces of European origin, except in the Balkans (4 acc.), and absent from modern pea varieties. The *eIF4E<sup>C</sup>* variant was the most frequent in Iberian peninsula (25%) and India-Nepal (16.6%), followed by Balkan (9%), Ethiopia (6%), and Afghan (6.2%) locations, and occurring in total of 91 (3.2%) accessions (Table 2). Due to bulking of 10 plants per sample in the case of Chinese origin accessions (ATFCC), we detected high proportion (15%) of sample heterogeneity. This heterogeneity was not possible to test in USDA, IPK, CGN and JIC samples, as they originated from single plants, while in CzNPC and VIR samples (originated again from bulks of 10 plants per samples) prevailed susceptible alleles. Finally, the susceptible *eIF4E<sup>S</sup>* variant was found in 53.8% to 100% proportion within studied regions, in total 2331 accessions (83%). The lowest occurrence was in 145 studied

accessions from India (53.8%), followed by China (63%) and Ethiopia (63%). In contrast, within the 1145 analyzed European origin accessions, including modern pea varieties, it was predominant (97%) (Fig. 2, Table 2). It is clear that there is bias towards European (1145 acc.) compare to other geographical regions, affecting likely allele distribution resulting in S allele over-representation.

### Isolation of *eIF4E* sequences from the selected accessions

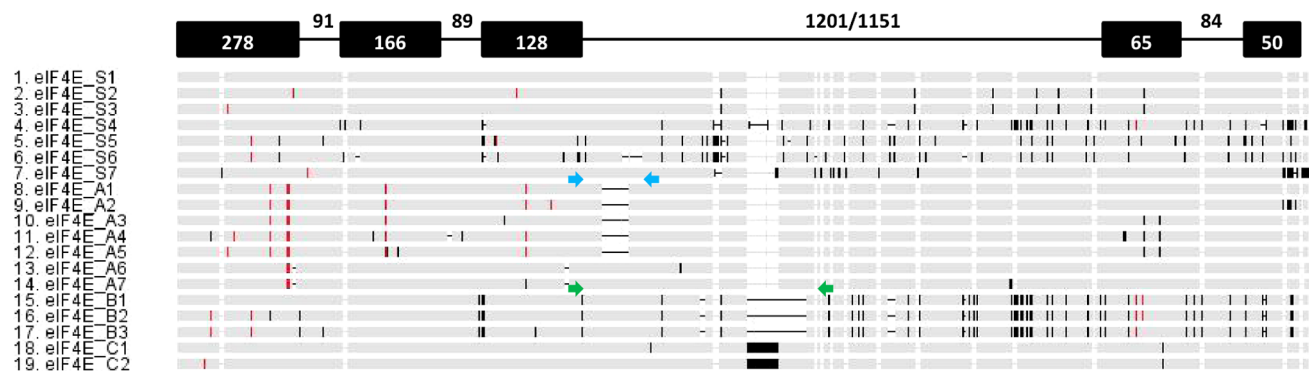
In order to assess the genetic diversity within the four classes (A, B, C and S) and to test whether intron 3 lengths correspond to DNA and amino acid haplotypes; we sequenced 32, 25, 9 and 7 accessions with respective *eIF4E<sup>A-B-C-S</sup>* variants from selected geographically diverse regions: Turkey, Nepal, Pakistan, Afghanistan, Ethiopia and China (Fig. 2, Table 2, 3, Table S2). The total length of genomic DNA clones of *eIF4E<sup>A-B-C-S</sup>* variants from ATG to stop codons was 2102, 2080, 2208, 1973–2892 bp long, respectively (Table S2), and length polymorphism was largely conferred by 50 and 56 bp insertions/deletions of minisatellite-like repeat sequences located at 102 and 421 bp respectively from the beginning of intron 3 (Fig. 1, Table 1). The extensive variation of *eIF4E<sup>S</sup>* was due to wild *Pisum* samples, especially of more distant subspecies (*P. abyssinicum*) and species (*P. fulvum*). Beside intron 3 polymorphism, extensive nucleotide variation was found; altogether, there were 156 SNPs identified in 73 sequences (Fig. 1, Table S2) resulting in 34 alleles at whole gene (exons-introns), 26 at cDNA and 19 alleles at protein levels, respectively. Of these 156 SNPs, 62 were unique, e.g. occurred only in once within the sequenced samples.

There were four different sequences of susceptible *eIF4E<sup>S</sup>* found within 25 sequenced samples of the cultivated *Pisum sativum* gene pool and three additional sequences within 7 samples of wild pea species (Table 3, Table S2). Translation of computationally spliced cDNA has resulted in four (*eIF4E<sup>S-1-2-3</sup>* and 7) protein variants within the cultivated gene pool and an additional three (*eIF4E<sup>S-4-5-6</sup>*) in *P. fulvum* and *P. abyssinicum* (Fig. 3, Table S2). The typically resistant *eIF4E<sup>A-1</sup>* allele [31,39], as represented by the PI193835 accession from Ethiopia (genebank sequence number GU289735), differed from the *eIF4E<sup>S-1</sup>* allele (represented by JI194, genebank sequence number KF053441) by five nucleotide (G185T, C218A, C221A in exon 1, G410A in exon 2 and T687G in exon 3) exchanges, while the *eIF4E<sup>A-6</sup>* allele (PI269818 accession, genebank

sequence number KF053455) differed by an additional two nucleotide (G217C in exon 1, C993T in intron 3) exchanges, one AGC triplet (AGC229-231) deletion, and one single base (C768T) deletion, in addition to 1200 bp intron 3 length similar to susceptible 1201 bp (Table 2, Table S2). Within the 28 sequenced accessions with *eIF4E<sup>B</sup>* intron length variant, we detected seven alleles, both at nucleotide (Fig. 1) and amino acid (Fig. 3) levels. The three *eIF4E<sup>B-1-2-3</sup>* alleles found in nine sequenced samples differed from the *eIF4E<sup>S</sup>* alleles in 40 SNPs and one nucleotide insertion, resulting in four DNA and three protein variants. The deduced amino acid sequence of the *eIF4E<sup>B-1</sup>* allele differed from the susceptible *eIF4E<sup>S</sup>* allele at three amino acid (M207I and LD218-219QE) exchanges in exons 4 and 5. Two accessions from Turkey (JI1370, JI1090) differed by an additional three SNPs in exon 1, leading to two amino acid exchanges, V23D and V49A, resulting in two *eIF4E<sup>B-2-3</sup>* alleles (Table 3). We have detected two *eIF4E<sup>C</sup>* alleles among seven sequenced samples. The *eIF4E<sup>C-2</sup>* allele of the VIR1859 accession, besides having a 1257bp long intron 3 due to its extra copy of a second 56 bp repeat, differed by two (A56T in exon 1, G1872A in intron 3) nucleotides from the *eIF4E<sup>S</sup>* allele, leading to single amino acid (N19I) exchange at exon 1 (Fig. 3). The remaining six sequences of *eIF4E<sup>C</sup>* were identical to the *eIF4E<sup>S</sup>* allele, except for a 56 bp insertion in intron 3 and a single nucleotide exchange at intron 3 in PI505122 and PI639981, accessions from Albania and Bulgaria (Table 2, S2). We were interested in establishing which, if any, of the *eIF4E<sup>A, B</sup>* or *C* alleles can be found in wild pea. It turned out that all analyzed wild pea accessions displayed the *eIF4E<sup>S</sup>* variant as assessed by intron 3 length. Similarly, sequencing analysis of the seven selected accessions had shown that all *P. sativum* subsp. *elatius*, except for *P. fulvum* JI1007, JI1010 and *P. abyssinicum* JI1632 accessions with one or two amino acid exchanges, had the typically susceptible *eIF4E<sup>S-1</sup>* allele (Table 2, 3, S2). All had substantial sequence polymorphism compare to cultivated pea except for *P. sativum* subsp. *elatius* JI2630 from the Crimea.

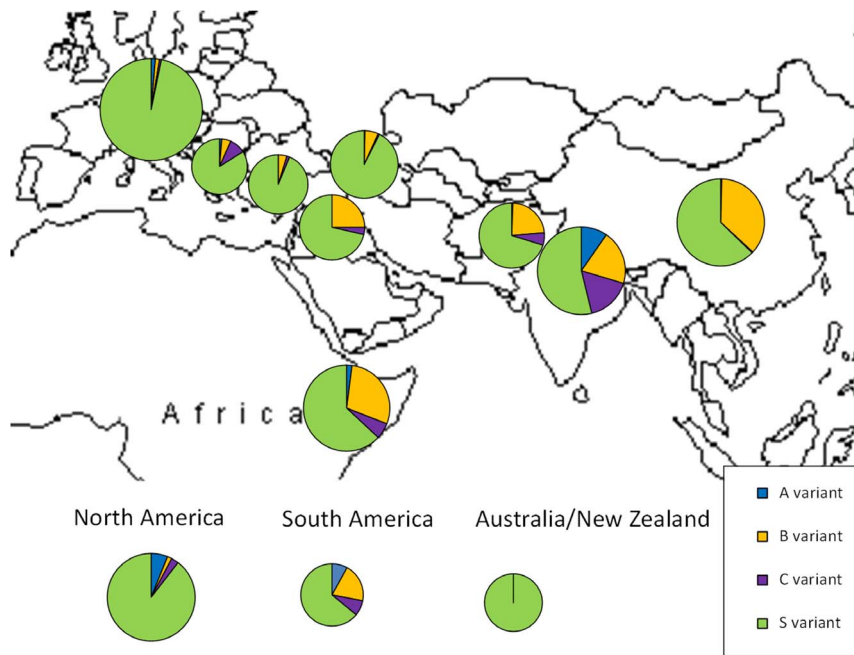
### Haplotype network analysis

Haplotype network analysis was used to reveal relationships between sequences. This showed the clear separation of *Pisum sativum* subsp. *elatius* (JI3157, JI1091, JI2630), *Pisum fulvum* (JI1007, JI1010) and *Pisum abyssinicum* (JI1632) accessions from cultivated germplasm (Fig. 4). The only exception was *Pisum sativum* subsp.



**Figure 1. Schematic representation of sequence alignment of the all 19 identified protein *eIF4E* alleles.** Four principal *eIF4E* variants identified by intron 3 polymorphism are designated A,B,C and S, while numbers indicate the respective allelic variant. Black bars indicate polymorphism nucleotides within both exons and introns, while red bars indicate polymorphism leading to amino acid exchanges. Horizontal lines indicate insertions/deletions. The heading line indicates nucleotide numbers and exon (solid black boxes)-intron (lines) positions and sizes. Blue arrows indicate primer A combination (Ps-eIF4E-750F and Ps-eIF4E-586R) and green arrows indicate primer B combination (Ps-eIF4E-750F and Ps-eIF4E-1270R).

doi:10.1371/journal.pone.0090394.g001



**Figure 2. Geographical distribution of four *eIF4E* alleles expressed as percentage of total.** Comparison of allele frequencies in 13 geographical regions, as detailed in Table 2.  
doi:10.1371/journal.pone.0090394.g002

*elatius* JI261 (genbank sequence number KF053440) from Turkey, which had a sensitive allele at the amino acid level, while exhibiting a transitory stage towards the *eIF4E<sup>B</sup>* allele at the nucleotide level (Table S2). Moreover, intron 3 was 36 bp longer, with a length of 1187 bp. Interestingly, the JI261 sequence was identical to *eIF4E<sup>S</sup>* until position 956 of intron 3, from which point forward it was similar to the *eIF4E<sup>B</sup>* allele. An important difference was the absence of the exon 4 and 5 exchanges, where it was again identical to the susceptible *eIF4E<sup>S</sup>* allele. The JI261 sequence occurred between the *eIF4E<sup>S</sup>* and *eIF4E<sup>B</sup>* alleles, with 23 mutational steps. A further 24 mutations separated the *eIF4E<sup>B</sup>* alleles of JI1194 (genbank sequence number KF053439) of Afghanistan, ATC7134 (KF053433) of China, and JI1090 (KF053434) and JI1370 (KF053432) of Turkey (Fig. 4). The *eIF4E<sup>A</sup>* alleles identified in Indian accessions (PI347464, PI347494, PI378158) proved to be derived from Ethiopian (PI193835, PIS479), Afghan (PIS468) and Indian (PI356991, ATC1044) accessions, and were identical. A further cluster contained Chinese (ATC6931, ATC6927) and Indian (PI347484, PI347422 and PI347328) origin S alleles, both at nucleotide and amino acid levels (Fig. 4).

#### Testing for response to PSbMV infection

The functional status of the 15 alleles of all four detected variants was verified through biological testing of their response to P1 PSbMV infection. All 19 of the pea accessions bearing *eIF4E<sup>A-1</sup>* tested showed resistance to viral infection i.e. the same phenotype as the non-infected or resistant controls (Table S2). DAS-ELISA testing proved negative, indicating the absence of viral coat protein. Six samples had *eIF4E<sup>A-2-3-4-5-6-7</sup>* alleles with single (PI116056, PI193584, JI1788) or multiple (PI269818, PI269774) amino acid exchanges, but none of these changed the expression of resistance in these accessions. The eight tested accessions bearing the *eIF4E<sup>B-1</sup>* allele split into two phenotypic groups. Three accessions (ATC7134 from China and CGN3311, PIS477 from Pakistan, all *eIF4E<sup>B-1</sup>*) were resistant to the P1 pathotype of PSbMV, as the virus was not able to replicate and spread. The second group of accessions (ATC7140, ATC6928 from China, CGN3319 from Pakistan, all *eIF4E<sup>B-1</sup>*; and JI1090, *eIF4E<sup>B-2</sup>*, from Turkey), showed the typical susceptibility symptoms of infection, vein clearing and chlorotic leaf mosaics, leaf size decrease, internode shortening, and sporadically mild leaf-rolling, 14 days after inoculation by PSbMV. The symptoms were comparable

**Table 1. Allele designation based on intron 3 length polymorphism, indicating size of obtained fragments with primer combinations A and B, number of repeat motives, intron 3 difference and total length.**

<i>eIF4E</i> variant	Primer combination A product (bp)	Primer combination B product (bp)	copies of 50 bp repeat	copies of 56 bp repeat	intron 3 difference	intron 3 length (bp)
S	293	586	2	2	0	1201
A	243	586	1	2	-50 bp	1151
B	293	536	2	1	-70 bp	1131
C	293	592	2	3	+56 bp	1257

doi:10.1371/journal.pone.0090394.t001

**Table 2.** Summary of germplasm screening of 2803 pea accessions, indicating four principal *elF4E* intron 3 length variants (A-B-C and S), distribution in total numbers and percentage over 13 geographical regions.

Regions	Number of accessions	Percentage								
		A variant	B variant	C variant	S variant	A variant	B variant	C variant	S variant	
Ethiopia - East Africa	349	7	101	21	220	2.0	28.9	6.0	63.0	
Afghanistan-Pakistan	178	1	41	11	125	0.6	23.0	6.2	70.2	
India-Nepal	145	14	29	24	78	9.7	20.0	16.6	53.8	
China-Mongolia	271	1	99	1	170	0.4	36.5	0.4	62.7	
Turkey-Syria	64	0	3	1	60	0.0	4.7	1.6	93.8	
Israel-Lebanon-Jordan	28	0	7	1	20	0.0	25.0	3.6	71.4	
Balkan	75	1	4	7	63	1.3	5.3	9.3	84.0	
Russia-Caucassus	295	1	19	2	273	0.3	6.4	0.7	92.5	
Iberian Peninsula	36	0	2	9	25	0.0	5.6	25.0	69.4	
Australia-New Zealand	11	0	0	0	11	0.0	0.0	0.0	100.0	
South America	25	2	5	2	16	8.0	20.0	8.0	64.0	
Europe	1145	15	15	7	1108	1.3	1.3	0.6	96.8	
USA-Canada	181	11	3	5	162	6.1	1.7	2.8	89.5	
<b>Total</b>	<b>2803</b>	<b>53</b>	<b>328</b>	<b>91</b>	<b>2331</b>	<b>Average</b>	<b>1.9</b>	<b>11.7</b>	<b>3.2</b>	<b>83.2</b>

doi:10.1371/journal.pone.0090394.t002

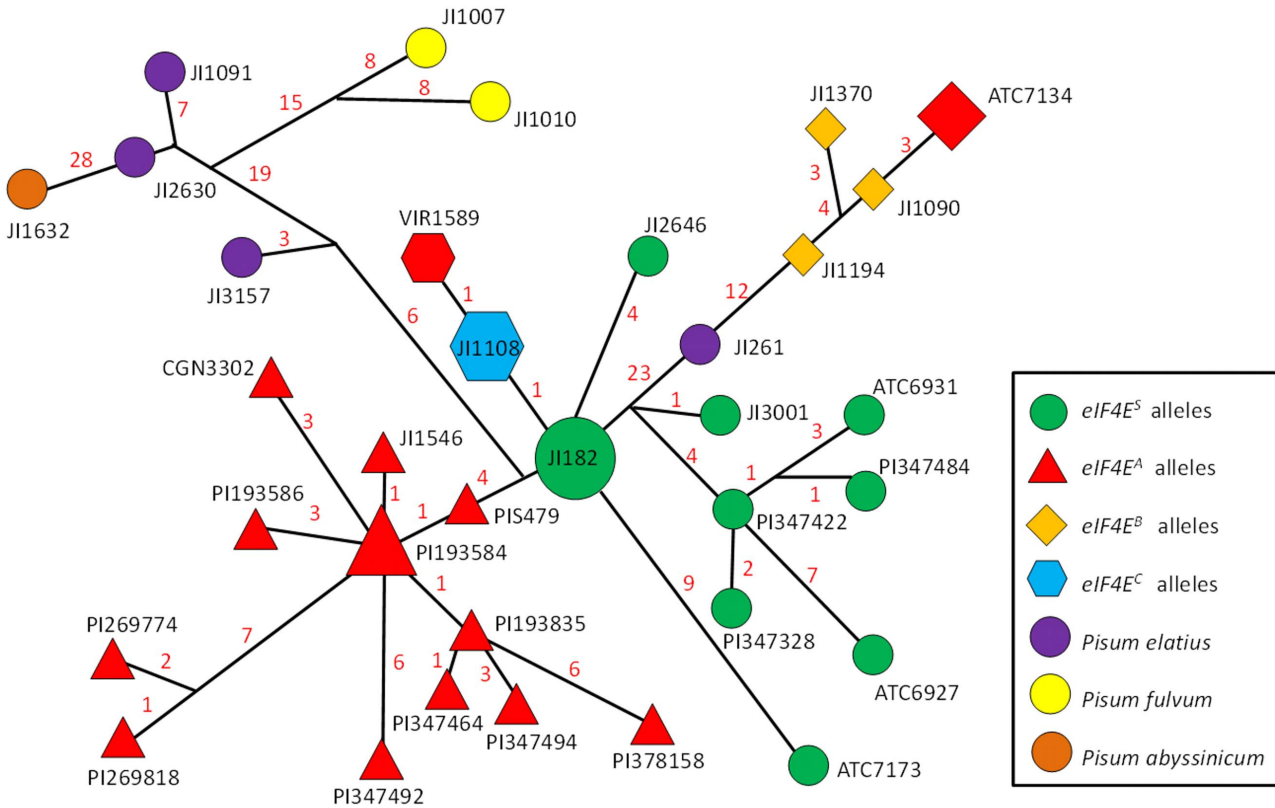
**Table 3.** List of sequenced 73 accessions assigned to identified 19 protein alleles of four *elF4E*<sup>A-B-C-S</sup> intron 3 length variants with indicated country of origin.

<i>elF4E</i> allele	Number of accessions with given allele	Accession(s)	Origin
<i>elF4E</i> <sup>S-1</sup>	24	Jl 182, Jl 1785, Jl 1030, Jl 190, PI 357290, Jl 193, Jl 194, Jl 205, Jl 1107, Jl 1085, Jl 2065, Jl 1845, Jl 1532, Jl 2607, Jl 1121, Jl 1756, Jl 2571, Jl 3001, Jl 2630, Jl 1091, Jl 3157, Jl 261, ATC_7173, ATC_6927	Nepal, Nepal, Iran, Sudan, Macedonia, Sudan, Sudan, Russia, Nepal, Turkey, Tanzania, Greece, Hindukush, Lybia, Nepal, Nepal, Gruzia, Iran, Crimea, Greece, Turkey, Turkey, China, China
<i>elF4E</i> <sup>S-2</sup>	1	PI 347328	India
<i>elF4E</i> <sup>S-3</sup>	3	PI 347422, PI 347484, ATC_6931	India, India, China
<i>elF4E</i> <sup>S-4</sup>	1	Jl 1632	Ethiopia
<i>elF4E</i> <sup>S-5</sup>	1	Jl 1007	Israel
<i>elF4E</i> <sup>S-6</sup>	1	Jl 1010	Israel
<i>elF4E</i> <sup>S-7</sup>	1	Jl 2646	Malawi
<i>elF4E</i> <sup>A-1</sup>	19	Jl 967, Jl 467, Jl 1790, ATC_1044, PIS_468, PIS_479, CGN_3302, PI 116056, PI 193584, PI 249645, PI 356991, PI 356992, PI 347494, PI 347492, Jl 1260, PI 193586, Jl 2643, Jl 1787, Jl 1788	Ethiopia, Europe, Afghanistan, Afghanistan, Afghanistan, Ethiopia, Ethiopia, India, Ethiopia, India, India, India, India, India, India, Ethiopia, Malawi, Nepal, Nepal
<i>elF4E</i> <sup>A-2</sup>	1	Jl 1546	Ethiopia
<i>elF4E</i> <sup>A-3</sup>	1	PI 193835	Ethiopia
<i>elF4E</i> <sup>A-4</sup>	1	PI 378158	India
<i>elF4E</i> <sup>A-5</sup>	1	PI 347464	India
<i>elF4E</i> <sup>A-6</sup>	1	PI 269818	England
<i>elF4E</i> <sup>A-7</sup>	1	PI 269774	England
<i>elF4E</i> <sup>B-1</sup>	7	ATC_7134, CGN_3311, CGN_3319, ATC_7140, PIS_477, Jl 1194, ATC_3275	China, Pakistan, Pakistan, China, Hindukush-Pakistan, Afghanistan, China
<i>elF4E</i> <sup>B-2</sup>	1	Jl 1370	Turkey
<i>elF4E</i> <sup>B-3</sup>	1	Jl 1090	Turkey
<i>elF4E</i> <sup>C-1</sup>	6	Jl 267, PI 505122, PI 639981, Jl1109, Jl1104, Jl1108	Greece, Albania, Bulgaria, Nepal, Nepal, Nepal
<i>elF4E</i> <sup>C-2</sup>	1	VIR 1589	Russia, Sverdlovsk

doi:10.1371/journal.pone.0090394.t003

	19	23	34	49	62*	73-74*	77	86	107*	108	115	148	163	169*	186	207	218-219
<b>S allele</b>																	
<i>eIF4E</i> <sup>S-1</sup>	N	V	E	V	W	AA	S	S	G	A	H	L	I	N	N	M	LD
<i>eIF4E</i> <sup>S-2</sup>	N	V	E	V	W	AA	R	S	G	A	H	L	T	N	N	M	LD
<i>eIF4E</i> <sup>S-3</sup>	N	V	A	V	W	AA	S	S	G	A	H	L	I	N	N	M	LD
<i>eIF4E</i> <sup>S-4</sup>	N	V	E	V	W	AA	S	S	G	A	H	L	I	N	N	I	LD
<i>eIF4E</i> <sup>S-5</sup>	N	V	E	A	W	AA	S	S	G	A	H	W	I	N	N	M	LD
<i>eIF4E</i> <sup>S-6</sup>	N	V	E	A	W	AA	S	S	G	A	H	L	I	N	N	M	LD
<i>eIF4E</i> <sup>S-7</sup>	N	V	E	V	W	AA	S	C	G	A	H	L	I	N	N	M	LD
<b>A allele</b>																	
<i>eIF4E</i> <sup>A-1</sup>	N	V	E	V	L	DD	S	S	R	A	H	L	I	K	N	M	LD
<i>eIF4E</i> <sup>A-2</sup>	N	V	E	V	L	DD	S	S	R	A	H	L	I	K	H	M	LD
<i>eIF4E</i> <sup>A-3</sup>	N	V	E	V	L	DD	S	S	R	A	H	L	I	N	N	M	LD
<i>eIF4E</i> <sup>A-4</sup>	N	A	E	V	L	DD	S	S	R	A	H	L	I	K	N	M	LD
<i>eIF4E</i> <sup>A-5</sup>	N	V	A	V	L	DD	S	S	R	P	R	L	I	K	N	M	LD
<i>eIF4E</i> <sup>A-6</sup>	N	V	E	V	W	PD	-	S	G	A	H	L	I	N	N	M	LD
<i>eIF4E</i> <sup>A-7</sup>	N	V	E	V	W	PD	-	S	G	A	H	L	I	K	N	M	LD
<b>B allele</b>																	
<i>eIF4E</i> <sup>B-1</sup>	N	V	E	V	W	AA	S	S	G	A	H	L	I	N	N	I	QE
<i>eIF4E</i> <sup>B-2</sup>	N	D	E	A	W	AA	S	S	G	A	H	L	I	N	N	I	QE
<i>eIF4E</i> <sup>B-3</sup>	N	D	E	A	W	AA	S	S	G	A	H	L	I	N	N	I	LD
<b>C allele</b>																	
<i>eIF4E</i> <sup>C-1</sup>	N	V	E	V	W	AA	S	S	G	A	H	L	I	N	N	M	LD
<i>eIF4E</i> <sup>C-2</sup>	I	V	E	V	W	AA	S	S	G	A	H	L	I	N	N	M	LD

**Figure 3. Summarized amino acid exchanges of identified 19 *eIF4E* alleles.** Asterisk indicates the position of exchanges previously identified as crucial for resistance. doi:10.1371/journal.pone.0090394.g003



**Figure 4. Haplotype network of 34 alleles identified at *eIF4E* whole gene level, using median-joining network algorithm, implemented in NETWORK.** It is based on total of 156 SNP characters, and excludes 50, 56 bp indels in intron 3, as these would compromise analysis. Red numbers above lines indicate mutational steps. The shortest tree has 235 mutations. Triangles indicate A, squares B, hexagons C and circles S variants as detailed in Table 1. Red colour of symbols indicates accessions tested resistant to P1 PSbMV. Size of symbols is proportional to number of accessions with given haplotype. doi:10.1371/journal.pone.0090394.g004

with the response of the sensitive genotypes with the *eIF4E*<sup>S-1</sup> allele. All tested *eIF4E*<sup>S-1</sup> (18 accessions), *eIF4E*<sup>S-2</sup> (1 acc.) and *eIF4E*<sup>S-3</sup> (3 acc.) displayed sensitive response to P1 PSbMV pathotype (Table S2). We have not been able to test remaining *eIF4E*<sup>S-4,5,6</sup> and <sup>7</sup> alleles, due to the shortage of homogenous plant material. Finally, of the three tested accessions with the *eIF4E*<sup>C-1-2</sup> alleles, only VIR1589 from the Sverdlovsk region of Russia was resistant, while JI1104 and JI1108 from Nepal proved susceptible (Table S2). This corresponded to an amino acid exchange at position N191 in exon 1 between the *eIF4E*<sup>C-1</sup> and *eIF4E*<sup>C-2</sup> alleles (Table S2).

## Discussion

In this study, we combined our knowledge and access to a wide range of pea germplasm [13,17] with an interest in PSbMV resistance [31] and focused our analysis on the geographical distribution of selected, easily detectable alleles of the *eIF4E* gene using the polymorphism of intron 3.

### *eIF4E* gene structure and diversity

The pea *eIF4E* gene is comprised of five exons, interrupted by four intron sequences, with a total length of 2.1 kb [31], which is spliced into a 687 nucleotide long (open reading frame) mRNA [32]. The size polymorphism of intron 3 is caused by 50 and 56 bp insertions of minisatellite-like repeat sequences [31], separated by 259 bp of common intron sequence. This polymorphism led to the initial mapping of the *sbm-1* locus [35] and the development of a co-dominant marker for breeding [31]. Previous sequencing of 43 accessions of pea varieties showed four haplotypes of intron 1, eight haplotypes of the intron 2 and three haplotypes of intron 4, while 17 haplotypes were detected in intron 3 sequences ([31] and unpublished). Since all, except PI269774 (Sankia) and PI269818 (Aa134), previously tested susceptible and resistant accessions differed by 50 or 56 bp indels in intron 3 [31], we exploited this length polymorphism to screen a wider pea germplasm population. This led to the identification of novel polymorphisms in four principal variants (*eIF4E*<sup>S</sup>, *eIF4E*<sup>A-B-C</sup>). Another functional *eIF4E* resistance allele found in the PI269818 and PI269774 accessions was not directly tested in this study, as its polymorphism is more difficult to analyze. It is clear that this type of analysis does not provide a comprehensive dataset, as more subtle, single-nucleotide mutations are missed, as documented by sequenced samples, whereas in 73 sequences, 19 *eIF4E* alleles were found (Fig. 1, Table 3, S1).

The mutant *eIF4E*<sup>A-1</sup> allele of pea gene *eIF4E* differs from its *eIF4E*<sup>S-1</sup> counterpart in five non-conservative amino acid exchanges in and around the cap-binding pocket of pea *eIF4E*, which impacts infection by PSbMV [32]. Previously reported mutations in pea *eIF4E* were localized on the  $\beta$ 1 (W62L),  $\beta$ 1- $\beta$ 2 (AA73-74DD),  $\beta$ 3- $\beta$ 4 (G107R) and  $\beta$ 5 (N169K) loops [42]. Positions of the newly identified amino acid (V23D, V49A and LD218-219QE) changes in *eIF4E*<sup>B-1-2-3</sup> alleles do not correspond with the known mutations leading to resistance and are not localized on the variable  $\beta$  loops regions, with the exception of the M207I change located on the  $\alpha$ 3- $\beta$ 7 loop. The only resistance determinant mapped to the  $\beta$ 7 loop is for BaYMV infection in barley [54]. In the *Capsicum* species, several *eIF4E* haplotypes with amino acid changes were reported, some of which included amino acid changes in positions 218 and 219 similar to those in the pea *eIF4E*<sup>B</sup> allele. Systematic testing of individual mutations by *in planta* assay showed that only a specific combination of mutations leads to resistance [42]. This hypothesis confirms previously reported amino acid changes in pea *eIF4E* in PI269818 and PI269774

accessions [31] with *eIF4E*<sup>A6-7</sup> alleles, which have amino acid changes in the cap-binding pocket [31] and lack one (S78) amino acid (Fig. 3, Table 3 and Table S2, Supporting information). Amino acid changes in the loop near the cap-recognition pocket were shown to be directly associated with resistance to potyvirus in pepper, lettuce, and pea [35,55]. Moreover, not all *eIF4E* proteins encoded by the resistance alleles are defective in their ability to bind the m7-GTP cap, suggesting that disrupted cap binding is not always required for potyvirus resistance [56]. Thus mutations conferring *sbm-1* resistance to PSbMV act combinatorially [42]. The interactions between particular *eIF4E* and particular viral VPgs are highly specific for host-virus interactions. Higher plants are unique in that they encode two distinct isoforms of *eIF4F* that have both overlapping and isoform-specific roles, *eIF4F* protein complex, which contains *eIF4E* and *eIF4G*, and *eIF(iso)4F* protein complex, which contains *eIF(iso)4E* and *eIF(iso)4G* [25]. Although these two complexes seem to be equivalent for the *in vitro* translation, they differ in their *in vivo* expression patterns and specificity for cellular mRNAs and likely also viral RNA. Potyviruses could selectively use either *eIF(iso)4E* or *eIF4E* to infect plants, while some are able to use both [39]. In pepper, the combination of *eIF4E* (*pvv2* locus on chromosome 4) and *eIF(iso)4E* (*pvv6* locus on chromosome 3) resistance alleles showed a complementary effect [45]. Also in pea, the *eIF(iso)4E* homologous gene exists, mapped to LG II ([35] and close to the *sbm-2* locus (not shown). However, its direct involvement in potyvirus resistance still needs to be demonstrated. Thus we cannot exclude the possibility that interaction between *eIF4E* and *eIF(iso)4E* proteins may also play a role in pea. This might partly explain the discrepancies in virological testing of *eIF4E*<sup>B</sup> accessions. Finally, the result of the interaction, the resistance or tolerance, depends not only on the genetic background but also on the virus pathotype [43,57]. It would be interesting to test region-specific potyvirus isolates [58] on the alleles identified in this study. The current phylogeny of potyviruses indicates that the progenitor probably infected plants growing in southwestern Eurasia and northern Africa, which evolved from a rymovirus by acquiring the ability to be transmitted by aphids [48]. It is notable that we did not find the resistant *eIF4E*<sup>A</sup> variant in any of the 146 tested wild pea samples, nor have we been able to detect any intermediate alleles between the susceptible *eIF4E*<sup>S-1</sup> and resistant *eIF4E*<sup>A-1</sup>, except in JI261 accessions already separated by 23 mutations between *eIF4E*<sup>S-1</sup> and *eIF4E*<sup>B-1</sup>. Such an intermediate allele was found for *Pm3* powdery mildew resistance in wheat from the Himalayan Range, supporting the hypothesis that a recent evolution of these alleles took place in the hexaploid wheat gene pool [8]. This is a departure from the situation with other diseases for which resistance can often be found within the wild genepool, in contrast to the domestic gene pool. i.e. pathogens overcame resistance in the domestic gene pool. Haplotype analysis indicated the separate evolution of *eIF4E*<sup>A</sup> alleles, with Ethiopian and Indian accessions being the most distantly related, while Chinese and some Afghan and Nepalese accessions were closer to *eIF4E*<sup>S-1</sup> (Fig. 4). Thus the evolution of the *eIF4E* gene is demonstrated in the available pea genepool.

### Geographical distribution of *eIF4E* variants versus PSbMV diversity

Pea genotypes resistant to PSbMV were identified by Hagedorn and Gritton [59] in two Ethiopian lines (PI193586 and PI193835). Later, Hampton [60] found several resistant accessions from India. Since they are all identical in sequence [31], it raises the question of their origin. Hampton [60] proposed that India may have been the center of origin for the virus, and consequently also the center

of PSbMV-immune germplasm. Despite this resistant germplasm, pea has not been systematically bred to include PSbMV resistance until recently; as a consequence, the vast majority of pea varieties are susceptible to this virus [31], especially when contrasted with landraces where the possibility of both natural and human selection for resistance exists. A relationship between pea host and PSbMV pathogen diversity in northern Pakistan and Afghanistan has been proposed [60,61]. More recently, phylogenetic analysis of PSbMV isolates from Australia and China grouped them on separate clades [55]. These results suggest the operation of co-evolutionary forces, which could be tested using the *eIF4E* alleles identified in this study. Interestingly, similar results were found in a study of barley *eIF4E* haplotype diversity, which was also found to be considerably higher in Central and East Asia, both regions with a long history of the bymovirus disease [3]. It is intriguing to see that in different crops, the resistance diversity center is outside the original place of domestication, especially in Central and Eastern Asia. It might be hypothesized that only some environments may be conducive for pathogen development in crops, hence co-evolution may be geographically limited. This agrees with results showing that the *eIF4E* gene is under positive selection pressure [3,44]. There is strong selection for the *eIF4E<sup>B</sup>* allele in China and less elsewhere (Fig. 2, Table 2), but no A or C alleles in China, possibly consistent with a genetic bottleneck in the later migration of pea to China, compared with central/western Asia [18]. Moreover, assignment of Chinese accessions to provinces showed preferential occurrence of B alleles in the autumn/winter-sown provinces of Henan, Anhui, Hubei and Sichuan (Table S1) that experience high rainfall, long growing seasons and cold, frost-prone vegetative phases. Conversely, S alleles predominated in spring-sown types grown in northern provinces with low rainfall and high, variable temperatures that result in severe drought [62]. Analysis of pea germplasm diversity, using retrotransposon-based insertion polymorphism [13,14,63] showed that although there is substantial genetic diversity present, it is only partially geographically structured. Especially in case of the *eIF4E<sup>B</sup>* alleles, there is substantial clustering of Indian and Ethiopian accessions, indicating a close relationship between materials from these areas (Fig. 5). The number of mutations makes independent origin unlikely, unless stringent positive selection is operating, as detected by Cavatorta [44]. We have also found evidence of positive selection in a set of 27 tested pea cDNA haplotypes (not shown). As proposed for barley *eIF4E*, the explanation for the unusually high overall degree of *eIF4E* sequence variation and haplotype diversity may be that these are important for adaptation to different local habitats [3]. The second, more plausible scenario is introgression and maintenance in populations by selective advantage. Based on the number of accessions with detected *eIF4E<sup>A</sup>* and *eIF4E<sup>B</sup>* alleles in the northern Indian subcontinent versus Ethiopia/eastern Africa, we speculate that these more likely originated in the Africa and were then brought by oversea trade to Afghanistan-Pakistan-Nepal-India region. The Silk Route, Amber Road and trans-Saharan trade routes were all instrumental in establishing links between Africa, India and beyond. This scenario, while not proven, is supported by a haplotype analysis network, where Indian *eIF4E<sup>A</sup>* alleles were shown to be derived from Ethiopian accessions (Fig. 4). There is bias towards European origin accessions, comprising 40.8% of samples, followed by Ethiopia and East Africa (12.4%) and Russia (10.5%) which consequently affects proportion (57%) of S alleles (Table 2). However, these samples were included to provide more complete view on *eIF4E<sup>A</sup>* diversity, including modern pea varieties. However this affects only

percentage in total dataset and not within studied regions (Fig. 2, Table 2).

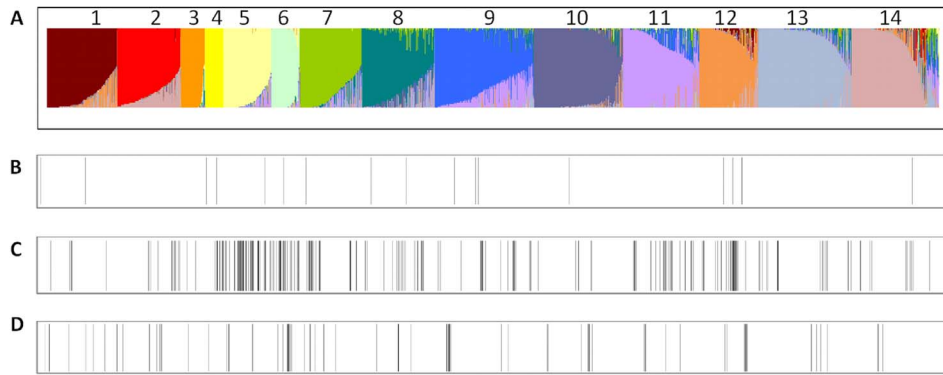
Various approaches have been applied to identify variants of the *eIF4E* gene in several crops. These showed different levels of diversity [3,43,46,47]. In the most comprehensive study, 1090 barley samples from 84 countries screened using the high-resolution melting PCR method of cDNA, led to the identification of 47 *eIF4E* haplotypes [3]. In our study, we have identified 34 alleles on exon-intron level with 156 single-nucleotide polymorphic sites in total of 73 sequences. This suggests that substantial diversity within the *eIF4E* gene can be expected within available pea germplasm, which might be fully revealed only by sequencing. It is notable that none of the 146 tested wild pea samples had *eIF4E<sup>A</sup>* or *eIF4E<sup>B</sup>* alleles, but at the amino acid level, all corresponded to the susceptible *eIF4E<sup>S-1</sup>* allele, despite substantial polymorphism both in introns and exons, indicating selection for functionality. The discovery of the intermediate allele in wild *P. sativum* subsp. *elatius* JI261 from Turkey is similar to data obtained in *Arabidopsis thaliana* [25], which suggests a distinct mode of evolution of resistance in wild species as opposed to crop species and supports the scenario in which potyviruses spread at the advent of agriculture, which brought plants into dense monocultures [48]. Very little is known about the selective forces that drive viral evolution in natural ecosystems, which contrasts with the more detailed population genetics studies in crop plants that have revealed the importance of mutation rates, recombination, genetic drift and migration in virus evolution [64]. Although we can not fully excluded the possibility, that JI261 allele could have resulted from domestic to wild introgression and not independent mutations, the analysis of 45 RBIP loci (Fig. 5) clustered it with typical *P. s.* subsp. *elatius* accessions [63]. The existence of resistance alleles only in the domesticated pea gene pool leads us to speculate that the mutation originated during early cultivation. The series of alleles identified in this study provide an excellent basis for testing various potyviruses and pathotypes in which to study the co-evolution of potyviruses and their pea host.

## Materials and Methods

### Plant material

The 2803 pea accessions used in this study were drawn from the following collections: the Czech National Pea Collection (CzNPC, 1252 accessions); the USDA core collection (384); 119 accessions of Chinese origin from the Australian Temperate Field Crops Collection [(ATFCC) now reconstituted as Australian Grains Genebank, (AGG)], the Vavilov Institute (VIR) Russian Federation (69); the Centre for Genetic Resources (CGN), The Netherlands (14); Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Germany (35); and John Innes Centre (JIC), UK (836). The permission from the relevant institutions was obtained to access the collections, and the respective pea germplasm accessions were donated for this study under Standard Material Transfer Agreement of germplasm resources. Although we specifically targeted local landraces rather than globally grown, modern varieties, the analyses of 1252 accessions from the Czech National Pea collection represent the diversity present in the gene pool of cultivated pea in the 19<sup>th</sup> to 21<sup>st</sup> centuries [16,49]. This collection is composed of 972 commercial varieties, 226 breeding lines and 54 landraces, which originate largely from Europe (925), the former Soviet Union (177), and the USA and Canada (77). In contrast, the John Innes Centre pea collection has, in addition to varieties (1071) and breeding lines (61), a large proportion of landraces (600), mutant stocks (585) and wild peas (445). Geographically, this collection is dominated by accessions of





**Figure 5. Visualization of genetic relationship of four identified *eIF4E* variants on the background of 14 BAPS identified clusters of 3,029 JIC accessions genotyped by 45 retrotransposon-insertion loci [13,63].** Notably, clusters 3 and 4 contain 145 acc. of wild *P. fulvum*, *P. s. subsp. elatius* and *P. abyssinicum*; cluster 7 contains 95 acc. from Afghanistan; cluster 8 has 225 acc. from Ethiopia; and cluster 9 contains 247 acc. from India and Ethiopia, while remaining clusters are more diverse (A). 18 accessions with the *eIF4E<sup>A</sup>* (B), 241 accessions with the *eIF4E<sup>B</sup>* (C), 81 accessions with the *eIF4E<sup>C</sup>* variant (D). doi:10.1371/journal.pone.0090394.g005

Ethiopian (388), Mediterranean (199), Indian (53) and Chinese (37) origin. The Australian ATFCC and USDA pea collections are partially complementary. The former includes the diversity core set of Chinese origin [18], of which 119 were studied here, while the latter has over 6000 accessions, of which a core set of 384 [19] was analyzed in this study. Finally, 69 accessions selected from the VIR collection originated in the Caucasus and Central Asia (Turkmenistan, Kazakhstan, Georgia, and Armenia) as well as in North Africa (Morocco and Algeria). It is worth mentioning that although Ethiopia is known as the origin and occurrence for *P. abyssinicum*, the accessions used in this study were mainly *P. sativum* subsp. *sativum*. This set was complemented with wild *P. fulvum* (9), *P. sativum* subsp. *elatius* (86), *P. sativum* subsp. *sativum* (19) (formerly *P. humile/P. syriacum*) (see Smýkal et al. [13] for taxonomical classification) and *P. abyssinicum* (32) (Table S1). These represent both primary and secondary centers of pea diversity as well as primary and secondary gene pools.

#### DNA isolation and PCR analysis

Young leaves were harvested from ten (in the case of the Czech, ATFCC, CGN, IPK and VIR collections) or single (in the case of the JIC and USDA collections) randomly chosen plants per accession and stored at  $-80^{\circ}\text{C}$ . Genomic DNA was isolated and PCR performed using standard protocols [31]. Products were resolved on 1.5% TBE agarose gel and visualized using ethidium bromide staining under UV-light.

#### *eIF4E* variants amplification

Intron 3 length polymorphism was used as screening criteria, as developed by Smýkal et al. [31]. In short, differences in intron 3 length were assayed using PCR with the following pairs of primers: *P<sub>s</sub>-eIF4E-750F* (5'-GGACTAAGAAATGCTTCAAATGAAGCTGC-3') and *P<sub>s</sub>-eIF4E-586gR* (5'-GAATCATTTAA-GAAGCTCGTGAAGTG-3') primers (combination A, nested within the combination of B) that amplified 243 bp in the resistant and 293 bp in the susceptible accessions [31]; and (combination B) *P<sub>s</sub>-eIF4E-750F* and *P<sub>s</sub>-eIF4E-1270R* (5'-ATTCTCGATCACAC-TAGCCCCCTCC-3') [35] primers that amplified 536 bp versus 586 bp fragment. The combination of these two assays resulted in the detection of respective *eIF4E<sup>A-B-C-S</sup>* variants (Fig. 1, Table 1). PCR amplification of 1000 bp product from start codon to intron 3 was carried out using primers *P<sub>s</sub>-eIF4E-1F* start codon (5'-ATGGTTGTAGAAGACACCCCAAAATC-3') and *P<sub>s</sub>-eIF4E-*

586gR. Primers *P<sub>s</sub>-eIF4E-794F* (5'-GCTAGATGGTTGTTATGATGTTTATCAG-3') and *P<sub>s</sub>-eIF4E-2188R* stop codon (5'-TTGCTAGTTTGCTACCATGTAAGAACG-3') were used to amplify a 1500 bp PCR product spanning the rest of *eIF4E* gene from intron 3 to exon 5. The PCR products were purified according to Werle et al. [50] protocol and sequenced using a BigDye Terminator kit (Applied Biosystems, UK) by MacroGene (Amsterdam, The Netherlands).

#### Bioinformatics

Primer design and restriction analyses were performed using FastPCR software version 5.1.83 (PrimerDigital Ltd., Finland). The DNA sequences were viewed and edited using Sequence Scanner version 2.0 (Applied Biosystems, UK). CLUSTALW alignment was performed using BioEdit version 7.09.0 [51]. The haplotype alignments were performed using a median-joining network algorithm, implemented in NETWORK 4.5.1.6 [52]. To reveal the genetic relationship of samples, we used previously made Bayesian clustering analysis of the genetic diversity of 3,029 JIC accessions genotyped by 45 retrotransposon-insertion loci [13]. Identified clusters were used as framework to visualize individual *eIF4E* alleles within the set of 836 JIC accessions analyzed in this study. Single nucleotide polymorphism mapping to reference sensitive allele was done using Geneious 6.1.6 analysis software (Biomatters, USA).

#### Virological testing

Plants were grown in a substrate Klassman no. 4 (Klasmann-Deilmann GmbH, Germany) in a growth chamber (Microclima 1000, Snijders Scientific, Holland) under a 16/8-h and 22/18°C day/night cycle. Evaluation of resistance/susceptibility to the PSbMV pathotype P-1 was conducted through mechanical inoculation using isolate PSB117CZ [31,53]. All together, 50 accessions were tested. Ten plants per accession were tested in same block, under the same conditions. Sensitive pea cultivars, Merkur and Raman, were used as a sensitive and B99 as a resistant controls [31]. Symptoms were observed at one-week intervals, and systemic infection was confirmed three weeks after infection with DAS-ELISA (Loewe Biochemica, Germany).

## Supporting Information

**Table S1 List of studied accessions, divided by 7 studied germplasm collection with indicated accession number, name, origin and *eIF4E* variant.** (XLSX)

**Table S2 Table of sequenced 73 accessions indicating single-nucleotide polymorphic (SNP) sites in exons and introns (SNP sheet) as well as resulting amino acid exchanges (protein sheet).** Species name, germplasm, country of origin and *eIF4E* allele assignment are shown. Length of all introns and exons is shown in base pairs, with variable intron 3 being highlighted. Results of virological testing with P1 PSBMV

## References

- Tanksley SD, McCouch SR (1997) Seed banks and molecular maps: Unlocking genetic potential from the wild. *Science* 277: 1063–1066.
- Bhullar NK, Street K, Mackay M, Yahiaoui N, Keller B (2009) Unlocking wheat genetic resources for the molecular identification of previously undescribed functional alleles at the *Pm3* resistance locus. *Proc Natl Acad Sci USA* 106: 9519–9524.
- Hofinger BJ, Russell JR, Bass CHG, Baldwin T, dos Reis M, et al. (2011) An exceptionally high nucleotide and haplotype diversity and a signature of positive selection for the *eIF4E* resistance gene in barley are revealed by allele mining and phylogenetic analyses of natural populations. *Mol Ecology* 20: 3653–3668.
- Nieto C, Piron F, Dalmats M, Marco CF, Moriones E, et al. (2007) EcoTILLING for the identification of allelic variants of melon *eIF4E*, a factor that controls virus susceptibility. *BMC Plant Biology* 7: 34.
- Ibiza VP, Cañizares J, Nuez F (2010) EcoTILLING in *Capsicum* species: searching for new virus resistance. *BMC Genomics* 11: 631.
- Reeves PA, Panella LW, Richards CHM (2012) Retention of agronomically important variation in germplasm core collections: implications for allele mining. *Theor Appl Genet* 124: 1155–1171.
- Robaglia C, Caranta C (2006) Translation initiation factors: a weak link in plant RNA virus infection. *Trends in Plant Sci* 11: 40–45.
- Bhullar NK, Zhang Z, Wicker T, Keller B (2010) Wheat gene bank accessions as a source of new alleles of the powdery mildew resistance gene *Pm3*: a large scale allele mining project. *BMC Plant Biol* 10: 88.
- Chimwamurombe PM, Khulbe RK (2011) Domestication. In: Pratap A, Kumar J, editors. *Biology and Breeding of Food Legumes*. MA, USA: CABI, Cambridge. pp. 19–34.
- Zohary D, Hopf M (1993) Domestication of Plants in the Old World—The Origin and Spread of Cultivated Plants in West Asia, Europe, and the Nile Valley. Oxford: Clarendon Press.
- De Candolle A (1882) Origin of cultivated plants. Whitefish MT: Kesinger Publishing LCC.
- Maxted N, Ambrose M (2001) *Peas (Pisum L.)*. In: Maxted N, Bennett SJ, editors. *Plant Genetic Resources of Legumes in the Mediterranean*. Dordrecht: Kluwer Academic Publishers. pp. 181–190.
- Smykal P, Kenicer G, Flavell AJ, Corander J, Kosterin O, et al. (2011) Phylogeny, phylogeography and genetic diversity of the *Pisum* genus. *Plant Gen Res* 9: 4–18.
- Smykal P, Aubert G, Burstin J, Coyne C, Ellis N, et al. (2012) Pea (*Pisum sativum* L.) in the genomic era. *MDPI Agronomy* 2: 74–115.
- Smykal P, Coyne C, Redden R, Maxted N (2013) Peas. In: Singh M, Bishit IS, editors. *Genetic and Genomic Resources for Grain Legume Improvement*. London: Elsevier Insights. pp. 41–80.
- Smykal P, Hybl M, Corander J, Jarkovsky J, Flavell AJ, et al. (2008) Genetic diversity and population structure of pea (*Pisum sativum* L.) varieties derived from combined retrotransposon, microsatellite and morphological marker analysis. *Theor Appl Genet* 117: 413–424.
- Jing R, Ambrose MA, Knox MR, Smykal P, Hybl M, et al. (2012) Genetic diversity in European *Pisum* germplasm collections. *Theor Appl Genet* 125: 367–380.
- Zong X, Redden R, Liu Q, Wang S, Guan J, et al. (2009) Analysis of a diverse global *Pisum* sp. collection and comparison to a Chinese local *P. sativum* collection with microsatellite markers. *Theor Appl Gen* 118: 193–204.
- Kwon SJ, Brown AF, Hu J, McGee RJ, Watt CA, et al. (2012) Genetic diversity, population structure and genome-wide marker-trait association analysis emphasizing seed nutrients of the USDA pea (*Pisum sativum* L.) core collection. *Genes & Genomics* 34: 305–320.
- Wang A, Krishnaswamy S (2012) Eukaryotic translation initiation factor 4E-mediated recessive resistance to plant viruses and its utility in crop improvement. *Mol Plant Pat* 13: 795–803.
- Vavilov NI (1949-1950) The phytogeographic basis of plant breeding. In Chester KS trans. *The origin, variation, immunity, and breeding of cultivated plants*. Waltham MA: Chronica Botanica. pp. 13–54.
- Abbo S, Lev-Yadun S, Gopher A (2012) Plant domestication and crop evolution in the Near East: on events and processes. *Crit Rev in Plant Sci* 31: 241–257.
- Lovisolo O, Hull R, Rosler O (2003) Coevolution of viruses with hosts and vectors and possible paleontology. *Adv Virus Res* 62: 325–379.
- Jones EI, Ferriere R, Bronstein JL (2009) Eco-evolutionary dynamics of mutualists and exploiters. *American Naturalist* 174: 780–794.
- Le Gall O, Aranda MA, Caranta C (2011) Plant resistance to viruses mediated by translation initiation factors. In: Caranta C, Aranda MA, Tepfer M & López-Moya J, editors. *Recent Advances in Plant Virology*. Norfolk: Caister Academic Press. pp. 177–194.
- Musil M (1966) Über das Vorkommen des Virus des Blattrollens der Erbse in der Slowakei (Vorläufige Mitteilung). *Biologia* 21: 133–138.
- Latham LJ, Jones RAC (2001) Alfalfa mosaic and pea seed-borne mosaic viruses in cool season crop, annual pasture, and forage legumes: susceptibility, sensitivity, and seed transmission. *Austr J Agri Res* 52: 771–790.
- Hampton RO, Mink GI (1975) Pea seed-borne mosaic virus. CMI/AAB Descriptions of Plant Viruses. 146 n.
- Alconero R, Weeden NF, Gonsalves D, Fox DT (1985) Loss of genetic diversity in pea germplasm by the elimination of individuals infected by pea seedborne mosaic virus. *Ann of App Bio* 106: 357–364.
- Hjulsager CK, Olsen BS, Jensen DM, Cordea MI, Krath BN, et al. (2006) Multiple determinants in the coding region of pea seed-borne mosaic virus P3 are involved in virulence against *sbm-2* resistance. *Virology* 355: 52–61.
- Smykal P, Šafařová D, Navrátil M, Dostálová R (2010) Marker assisted pea breeding: *eIF4E* allele specific markers to pea seed-borne mosaic virus (PSBMV) resistance. *Mol Breeding* 26: 425–438.
- Bruun-Rasmussen M, Moller IS, Tulinius G, Hansen KR, Lund OS, et al. (2007) The same allele of translation initiation factor 4E mediates resistance against two Potyvirus spp. in *Pisum sativum*. *Mol Plant Microbe Interact* 20: 1075–1082.
- Andrade M, Abe Y, Nakahara KS, Uyeda I (2009) The *cym-2* resistance to Clover yellow vein virus in pea is controlled by the eukaryotic initiation factor 4E. *J Gen Plant Pathol* 75: 241–249.
- Johansen EI, Lund OS, Hjulsager CK, Laursen J (2001) Recessive resistance in *Pisum sativum* and Potyvirus Pathotype resolved in a gene-for-cistron correspondence between host and virus. *J Virology* 75: 6609–6614.
- Gao Z, Evers S, Thomas C, Ellis N, Maule A (2004) Identification of markers tightly linked to *sbm* recessive genes for resistance to pea seed-borne mosaic virus. *Theor Appl Genet* 109: 488–494.
- Olsen BS, Johansen IE (2001) Nucleotide sequence and infectious cDNA clone of the LI isolate of Pea seed-borne mosaic potyvirus. *Archives in Virology* 146: 15–25.
- Johansen IE, Rasmussen OF, Heide M, Borkhardt B (1991) The complete nucleotide sequence of pea seed-borne mosaic virus RNA. *J Gen Virology* 72: 2625–2632.
- Johansen IE, Keller KE, Dougherty WG, Hampton RO (1996) Biological and molecular properties of a pathotype P-1 and a pathotype P-4 isolate of pea seed-borne mosaic virus. *J Gen Virology* 77: 1329–1333.
- Ruffel S, Gallois J-L, Moury B, Robaglia C, Palloix A, et al. (2006) Simultaneous mutations in translation initiation factors *eIF4E* and *eIF(iso)4E* are required to prevent Pepper vein mottle virus infection of pepper. *J Gen Virol* 87: 2089–2098.
- Sato M, Nakahara K, Yoshii M, Ishikawa M, Uyeda I (2005) Selective involvement of members of the eukaryotic initiation factor 4E family in the infection of *Arabidopsis thaliana* by potyviruses. *FEBS Letters* 579: 1167–1171.
- Fraile A, Garcia-Arenal F (2010) The coevolution of plants and viruses: resistance and pathogenicity. *Adv Virus Res* 76: 1–32.
- Ashby AJ, Stevenson CEM, Jarvis GE, Lawson DM, Maule AJ (2011) Structure-Based Mutational Analysis of *eIF4E* in Relation to *sbm1* Resistance to Pea Seed-Borne Mosaic virus in pea. *PLoS ONE* 6: e15873.
- Charron C, Nicolai M, Gallois JL, Robaglia C, Moury B, et al. (2008) Natural variation and functional analyses provide evidence for co-evolution between plant *eIF4E* and potyviral VPg. *Plant J* 54: 56–68.

are indicated as positive, susceptible (S) and resistant (R) reactions, respectively. (XLSX)

## Acknowledgments

We thank Eva Fialová and Jana Veselá for technical assistance and numerous colleagues for discussion over earlier versions of the manuscript.

## Author Contributions

Conceived and designed the experiments: PS. Performed the experiments: EK PH CC DŠ PS. Analyzed the data: EK PS DŠ MN. Contributed reagents/materials/analysis tools: AF MA MV RR. Wrote the paper: EK MA RR CC PS.

44. Cavatorta JR, Savage AE, Yeam I, Gray SM, Jahn MM (2008) Positive Darwinian selection at single amino acid sites conferring plant virus resistance. *J Mol Evol* 67: 551–559.
45. Rubio M, Nicolai M, Caranta C, Palloix A (2009) Allele mining in the pepper gene pool provided new complementation effects between *pvr2-eIF4E* and *pvr6-eIF(iso)4E* alleles for resistance to pepper vein mottle virus. *J Gen Virol* 90: 2808–2814.
46. Jeong HJ, Kwon JK, Pandeya D, Hwang J, Hoang NH, et al. (2012) A survey of natural and ethyl methane sulfonate-induced variations of *eIF4E* using high-resolution melting analysis in *Capsicum*. *Mol Breeding* 29: 349–360.
47. Rigola DJ, van Oeveren J, Janssen A, Bonne A, Schneiders H, et al. (2009) High throughput detection of induced mutations and natural variation using KeyPoint technology. *PLoS ONE* 4: e4761.
48. Gibbs AJ, Ohshima K, Phillips MJ, Gibbs MJ (2008) The Prehistory of Potyviruses: Their Initial Radiation Was during the Dawn of Agriculture. *PLoS ONE* 3: e2523.
49. Cieslarová J, Hýbl M, Griga M, Smýkal P (2012) Molecular Analysis of Temporal Genetic Structuring in Pea (*Pisum sativum* L.) Cultivars Bred in the Czech Republic and in Former Czechoslovakia Since the Mid-20th Century. *Czech J Gen and Plant Breeding* 48: 61–73.
50. Werle E, Schneider C, Renner M, Völker M, Fiehn W (1994) Convenient single-step, one tube purification of PCR products for direct sequencing. *Nucleic Acids Research* 22: 4354–4355.
51. Hall TA (1999) BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucl Acids Symp Ser* 41: 95–98.
52. Bandelt HJ, Forster P, Röhl A (1999) Median-joining networks for inferring intraspecific phylogenies. *Mol Biol Evol* 16: 37–48.
53. Šafářová D, Navrátil M, Petrusová J, Pokorný R, Piaková Z (2008) Genetic and biological diversity of the pea seed-borne mosaic virus isolates occurring in Czech Republic. *Acta Virologica* 52: 53–57.
54. Stein N, Perovic D, Kumlehn J, Pellió B, Stracke S, et al. (2005) The eukaryotic translation initiation factor 4E confers multiallelic recessive bymovirus resistance in *Hordeum vulgare* (L.). *Plant J* 42: 912–22.
55. Nicaise V, German-Retana S, Sanjuan R, Dubrana MP, Mazier M, et al. (2003) The eukaryotic translation initiation factor 4E controls lettuce susceptibility to the potyvirus Lettuce mosaic virus. *Plant Physiol* 132: 1272–1282.
56. Kang BC, Yeam I, Frantz DJ, Murphy JF, Jahn MM (2005) The *pvr1* locus in pepper encodes a translation initiation factor *eIF4E* that interacts with tobacco etch virus VPg. *Plant J* 42: 392–405.
57. Hjulsgaard CK, Lund OS, Johansen E (2002) A new pathotype of pea seedborne mosaic virus explained by the properties of the p3-6k1 and viral genome-linked protein (VPg) coding regions. *Mol Plant Microbe Interact* 15: 169–171.
58. Wylie SJ, Coutts BA, Jones RAC (2011) Genetic variability of the coat protein sequence of pea seed-borne mosaic virus isolates and the current relationship between phylogenetic placement and resistance groups. *Arch Virol* 156: 1287–1290.
59. Hagedorn DJ, Gritton ET (1973) Inheritance of resistance to the pea seed-borne mosaic virus. *Phytopathology* 62: 1130–1133.
60. Hampton RO (1986) Geographic origin of pea seed-borne mosaic virus: a hypothesis. *Pisum Newsletter* 18: 22–26.
61. Ali A, Randles JW (2001) Genomic heterogeneity in Pea seed-borne mosaic virus isolates from Pakistan, the centre of diversity of the host species, *Pisum sativum*. *Arch Virology* 146: 1855–1870.
62. Ling Li, Redden RJ, Zong X, Berger JD, Bennett SJ (2013) Ecogeographic analysis of pea collection sites from China to determine potential sites with abiotic stresses. *Genet Res Crop Evol* 60: 1801–1815.
63. Jing R, Vershinin A, Grzebyta J, Shaw P, Smýkal P, et al. (2010) The genetic diversity and evolution of field pea (*Pisum*) studied by high throughput retrotransposon based insertion polymorphism (RBIP) marker analysis. *BMC Evol Biol* 10: 44.
64. Roossinck MJ (2011) The big unknown: plant virus biodiversity. *Current Opinions in Virology* 1: 63–67.