UC Davis

UC Davis Previously Published Works

Title

The Co-4 locus on chromosome Pv08 contains a unique cluster of 18 COK-4 genes and is regulated by immune response in common bean

Permalink

https://escholarship.org/uc/item/2ts57030

Journal

Theoretical and Applied Genetics, 128(6)

ISSN

0040-5752

Authors

Oblessuc, Paula Rodrigues
Francisco, Camila
Melotto, Maeli

Publication Date

2015-06-01

DOI

10.1007/s00122-015-2500-6

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Manuscript
Click here to download Manuscript: COK4 manuscript_Final.docx
Click here to view linked References

The *COK-4* gene forms a unique cluster of 18 copies on chromosome Pv08 and is regulated by immune response in common bean

Paula Rodrigues Oblessuc[1,2], Camila Francisco[1*], Maeli Melotto[1]

[1] Department of Plant Sciences, University of California, Davis, CA 95616, USA

[2] CAPES foundation, Brazilian Ministry of Education, Brasília, DF 70040-020, Brazil

Corresponding author: Maeli Melotto
University of California
Department of Plant Sciences
One Shields Avenue
Davis, CA 95616
Phone: +1 530-752-1747
E-mail address: melotto@ucdavis.edu

*Present address: Hospital Municipal Universitário de São Bernardo, São Bernardo do Campo, SP 09735-260, Brazil

Key words: receptor-like kinase, *Phaseolus vulgaris*, resistance locus, genome analysis, *Catharanthus roseus* RLK1, PAMP-triggered immunity.

Abstract

Key message The common bean locus *Co-4*, traditionally referred to as an anthracnose resistance gene, contains a cluster of predicted receptor-like kinases (COK-4 and CrRLK1-like) co-regulated with the plant's basal immunity.

Abstract Genetic resistance to anthracnose, caused by the fungus *Colletotrichum lindemuthianum*, is conferred by major loci throughout the *Phaseolus vulgaris* genome, named *Co*. The complex *Co-4* locus was previously reported to have several copies of the *COK-4* gene that is predicted to code for a receptor-like kinase (RLK). In general, plant RLKs are involved in pathogen perception and signal transduction; however the molecular function of *COK-4* remains elusive. Using newly identified molecular markers (PvTA25 and $PvSNP_{COK-4}$), the SAS13 marker, *COK-4* sequence and phylogeny analyses, and the recently released bean genome sequence, we determined the most probable boundaries of the *Co-4* locus; a 325-Kbp region on the chromosome Pv08. Out of the 49 predicted transcripts in that region, 24 are putative RLKs (including 18 *COK-4* copies) with high similarity to members of the *Catharanthus roseus* RLK1 (CrRLK1) protein family from different plant species, including the well-described FERONIA (FER) and ANXUR. We also determined that two RLK-coding genes in the *Co-4* locus (*COK-4-3* and *FER-like*) are transcriptionally regulated when bean plants are challenged with the flg22 peptide, a commonly used elicitor of plant immunity, or the bacterium *Pseudomonas syringae* pv. *phaseolicola*, causal agent of halo blight. While *COK-4-3* is activated during immune response, *FER-like* is down-regulated suggesting that these genes may work together to fine tune plant responses to biotic stress. These results highlight the importance of dissecting the regulation and molecular function of individual genes within each locus, traditionally referred to as resistance gene based on genetic segregation analysis.

Introduction

Plants have the innate ability to recognize conserved microbial molecular patterns and establish immune responses that can be triggered by a broad range of pathogens or highly specific to a particular pathogen. These responses can be addressed in two major layers of plant immunity: pathogen-associated molecular pattern (PAMP)-triggered immunity (PTI) and effector-triggered immunity (ETI) (Jones and Dangl 2006; Spoel and Dong 2012). PTI is induced by perception of PAMPs through pattern-recognition receptors (PRRs) located at the plant cell surface and ETI is mediated by resistance (R) genes leading to hypersensitive response (HR). All kinds of phytopathogens can potentially activate PTI and/or ETI (Thomma et al. 2011), which may result in systemic plant responses such as induced systemic resistance (ISR) or systemic acquired resistance (SAR). These immune responses involve intricate metabolic pathways mediated by several plant hormones, such as jasmonic acid (JA) and salicylic acid (SA) (Thomma et al. 2011).

Among many pathosystems used to study the molecular process involved in plant-pathogen interaction, *Colletotrichum* species have long served as a model for hemibiotrophic fungal pathogens (O'Connell et al. 2012), being used in the early studies on phytoalexins in its interaction with common bean (*Phaseolus vulgaris* L.) (Kuć 1982). Besides its scientific importance, common bean is also the most economically important species of the genus *Phaseolus* and the primary dietary protein source for several populations, mainly in the developing countries (Broughton et al. 2003). *Colletotrichum lindemuthianum* (Sacc. & Magnus) Briosi & Cavara is the causal agent of anthracnose in common bean, one of the most serious diseases in this crop throughout the world; not only because of its seed-borne nature, but also for the great variability of this pathogen (Melotto et al. 2000). This disease is responsible for great losses on common bean yield (up to 100%) and, therefore it is one of the longest studies diseases of this crop (Kelly and Vallejo 2004; Singh and Schwartz 2010; Ferreira et al. 2013).

Understanding common bean resistance against anthracnose is one of the main goals in breeding programs as genetic resistance is the most-efficient and environmentally friendly control of crop diseases

(Dodds and Rathjen 2010). Until now, 14 anthracnose resistance loci were discovered (*Co-1* to *Co-14*) in common bean genome (Ferreira et al. 2013). The *Co-4* locus, first described in the genotype TO (Fouilloux 1979; Awale and Kelly 2001), confers resistance against several races of *C. lindemuthianum* (Balardin and Kelly 1998). A second allele, *Co-4*^{2}, was identified in the resistant differential cultivar G2333 that possesses a combination of three independent resistance loci, *Co-4*^{2}, *Co-5* and *Co-7* (Young et al. 1998). The single dominant *Co-4*^{2} locus for anthracnose resistance present in the G2333-derived breeding line SEL 1308 (Young et al. 1998) provides greater resistance than the original *Co-4*, and it is recognized among the broadest-based resistance genes described in bean (Balardin and Kelly 1998; Silverio et al. 2002). The molecular structure of the difference alleles of *Co-4* remains to be determined.

The genomic structure of *Co-4* locus has been defined using genetics and genomics tools. Sequencing of the bacterial artificial chromosome BAC 78L$_{17}$ (Vanhouten and MacKenzie 1999) identified with the SAS13 molecular marker (Young et al. 1998) revealed that the *Co-4* locus contains several putative orthologs of *Pto*-like kinase genes, named *COK-4* (Melotto and Kelly 2001; Melotto et al. 2004). *In silico* analysis suggests that *COK-4* is a member of the receptor-like kinase (RLK) family that codify for a 369-amino acid protein with a superfamily kinase domain, including AT-binding and transmembrane domains (Melotto and Kelly 2001).

RLKs are important PRRs that play a role in self- and non-self-recognition, including the perception of hormones (Shiu and Bleecker 2001), PAMPs, and pathogen effectors. Several RLKs involved in plant immunity have been identified, including Xa21 (Song et al. 1995), Pto (Sessa and Martin 2000), FLAGELLIN SENSING 2 (FLS2) (Chinchilla et al. 2006), BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED KINASE 1 (BAK1) (Chinchilla et al. 2007), among others. FLS2 is one of the well-studied RLKs (Zipfel et al. 2004), which is involved in PTI through the perception of the bacterial PAMP flagellin, acting together with BAK1, to activate downstream immune responses (Chinchilla et al. 2007). Thus, mounting evidence suggests that RLKs are part of basal plant immunity against fungal and bacterial pathogens.

Owing to its similarity to RLKs, we reasoned that *COK-4* could be regulated by PAMPs and play a role in basal immunity against other phytopathogens in addition to *C. lindemuthianum*. We first defined the bean genomic region containing the *Co-4* using genetics and genomics analysis; the locus is now placed in a 325-Kbp region close to the telomere of the Pv08 chromosome. Out of the 24 RLK-coding genes at the *Co-4* locus, 18 showed high nucleotide sequence similarity to the originally identified *COK-4* from the bean genotype SEL 1308. Functional analysis of two kinases in this locus (referred to as *COK-4-3* and *FER-like*) revealed that they are regulated upon leaf treatment with the PAMP flg22 and infection with *Pseudomonas syringae* pv. *phaseolicola* (Pph). These findings suggest that the *Co-4* locus not only confers resistance against the anthracnose fungus, but it is also involved in the early stages of PTI in common bean.

Material and methods

Mapping population and *C. lindemuthianum* pathogenesis assay

The common bean breeding line SEL 1308 was used as the source of the *Co-4*^{2} gene in a cross with Black Magic, an anthracnose susceptible black bean cultivar. Hybrid seeds were advanced to the F_2 generation and 98 randomly selected F_2 individuals were used as a mapping population (Melotto and Kelly 2001). Plants were grown in controlled environment at 22^{o}C, 80% relative humidity, and 16h of daily light. Ten-day old seedlings were spray-inoculated with race 73 of *C. lindemuthianum*, which is avirulent on bean plants carrying the *Co-4*^{2} gene (Young et al. 1998). Inoculum preparation, inoculation methods, and disease symptoms evaluation were conducted as described by Young and Kelly (1996).

Molecular marker development

Simple sequence repeats (SSR) markers were developed based on the DNA sequence of the clone BAC 78L$_{17}$ that was mapped to the *Co-4* locus (Melotto et al. 2004). SSRs were searched in the sequence using

the SSRIT software (http://www.gramene.org/db/markers/ssrtool; Temnykh et al. 2001). Among the SSR markers (Table S1), PvTA25 showed polymorphism between the SEL 1308 and Black Magic and was used to genotype the F_2 segregating population. The PCR reaction (25 μl) consisted of 1.5 mM $MgCl_2$, 1x enzyme buffer, 200 μM dNTP, 1U *Taq* polymerase (Promega, Madison, WI), 50 ng DNA, and 25 ng of each primer (Table S2). The PCR cycle was 2 min at 94 °C, plus 13 cycles of 30 sec at 94 °C, 30 sec at 70 °C (with 1 °C decrease per cycle), 2 min at 72 °C, and 20 cycles of 30 sec at 94 °C, 30 sec at 57 °C, 2 min at 72 °C, followed by a final extension cycle of 7 min at 72 °C. PCR products were resolved in 6% polyacrylamide gel fixed in 1% acetic acid and 10% ethanol solution for 10 min, followed by a wash with distilled water for 1 min. The gel was soaked in 1.5% nitric acid for 3 min and rinsed with distilled water for 1 min. Gel was stained with 0.2% silver nitrate for 20 min followed by two washed with distilled water for 30 sec each. Developing was conducted with a solution of sodium carbonate (30g/L) and 37% formaldehyde (0.54 ml/l). Blocking was performed with 5% acetic acid.

Amplified fragment length polymorphism (AFLP) markers were developed by using bulk segregant analysis (BSA; Michelmore et al. 1991). DNA from F_2 individuals were bulked in resistant and susceptible pools based on the anthracnose response of each individual. Bulked DNA was digested with *Eco*RI and *Mse*I restriction enzymes, followed by adaptor ligation, and pre-selective amplification using adaptor-specific primers containing one additional base (Table S2). Selective amplification was performed with primers containing two more random bases. PCR conditions were exactly as described by Hazen et al. (2002). Amplicons were resolved in 6% polyacrylamide gel following the same protocol described for the SSR analysis. The AFLP which showed good amplification pattern and polymorphic bands between parents and bulks were used to genotype the F_2 population individuals.

Single nucleotide polymorphism (SNP) makers were developed with *COK-4* open reading frame (ORF) sequences of contrasting bean genotypes (Melotto and Kelly 2001). Primers were designed to detect both parental alleles of the *COK-4* gene in the F_2 mapping population. One forward primer was designed to anneal with both homologs and two reverse primers were designed to specifically anneal to

one of each homolog (Table S2). The PCR was optimized to amplify both homologs in the same reaction for the heterozygous genotypes. The reaction consisted of 1x enzyme buffer, 3.5 mM $MgCl_2$, 200 µM dNTP, 1.5U *Taq* DNA polymerase (Gibco), 15 ng of forward primer, 15 ng of the Black Magic homolog reverse primer, and 30 ng of the SEL 1308 reverse primer in a 30 µl reaction. The thermocycling profile consisted of one cycle of 94°C for 4 min, 30 cycles of 94 °C for 10 sec, 70 °C for 30 sec, and 72 °C for 2 min, followed by an extension cycle of 72 °C for 7 min.

Genetic linkage analysis

In addition to the newly developed markers, individual F_2 plants from the Black Magic x SEL 1308 population were also screened with the previously found SCAR markers linked to the *Co-4* locus: SAS13 (Young et al. 1998), SBB14 and SH18 (Awale and Kelly 2001). The amplification conditions were the same as described on the respective publications for each marker. The linkage map was obtained based on the inheritance of both disease phenotype and molecular markers, which was confirmed in F_2 plants using the chi-square test. Linkage analysis was performed using the MAPMAKER 3.0b software (Lander et al. 1987) with thresholds of 3.0 LOD score value and a 37.5 centiMorgan (cM) of maximum genetic distance. Loci were ordered using the "*order*" command, and the final order was tested using the "*ripple"* command with a window of six markers. Finally, multipoint distance estimates were obtained using the "*map"* command, with the cM distance among markers and resistance locus being calculated by the Kosambi mapping function. The linkage map diagram was created with the MapChart 2.2 program (Voorrips 2002).

Physical localization of DNA markers and sequence analysis

All molecular markers with known sequence and tightly linked to the *Co-4* locus were used to define its physical location on the G19833 reference bean genome sequence v1.0 available at Phytozome

(http://www.phytozome.net/; Schmutz et al. 2014). The markers used were: PvTA25 and $PvSNP_{COK-4}$ from this study; SAS13 (a 978-bp sequence obtained from the genotype SEL 1308; Melotto et al. 2004); SH18 and SBB14 kindly provided by James Kelly and Halima Awale; and SCARY20 (phaseolusgenes ID 548) and SCARC08 (phaseolusgenes ID 334) (Queiroz et al. 2004; http://phaseolusgenes.bioinformatics.ucdavis.edu/). The SEL 1308 *COK-4* ORF (NCBI accession number GI:9796477; http://blast.ncbi.nlm.nih.gov/) and the whole sequence of the clone BAC $78L_{17}$ (NCBI accession number GI:38194906) were also aligned to the bean genome. Alignment between DNA marker and the reference genome sequences was performed using BLASTN with default parameters (E-value < 1 x 10^{-5} and identity ≥ 70%) to define marker location. In addition, pair-wise alignments between each marker and the *Co-4* region were performed using the BLASTN tool available at NCBI (bl2seq; Tatusova and Madden 1999) was used to refine the E-value for each marker. All the predicted transcripts in the *Co-4* locus were obtained from the Phytozome website and the putative functions of the genes were inferred with the Pfam annotation also available through the Phytozome database.

Phylogenetic analysis

First, we identified the top 100 hits of putative paralogs of the COK-4 kinase in common bean using the predicted COK-4 protein from the bean line SEL1308 (COK-4_SEL1308) as query against common bean proteome database available at Phytozome (BLASTP, threshold E-value ≤ 1 x 10^{-20} and identity > 30%). These 100 sequences and the SEL 1308 COK-4 were aligned with CLUSTALW as part of the MEGA 5.05 software (Tamura et al. 2011). The conserved catalytic tyrosine kinase domain of these predicted bean kinases were identified by searching the COK-4 protein sequence against the NCBI protein conserved domain database (CDD) (Marchler-Bauer et al. 2013) and the phylogenetic tree was created with MEGA 5.05 using the maximum parsimony method. Bootstrap support values were obtained over 1,000 replications.

The phylogeny described above was also performed with the top 100 hits of COK-4_SEL1308 against the non-redundant (nr) protein database of all species available at NCBI (BLASTP, threshold E-value $\leq 1 \times 10^{-20}$ and identity > 30%).

Pathogenesis assay

Seeds of the bean genotypes G2333 and Beluga (used as susceptible control) were germinated on filter paper in a growth chamber at 28°C with 12 hour photoperiod for three days. Seedling were transplanted to 1:1:1 v:v:v mixture of growing medium (Redi-earth plug and seedling mix, Sun Gro), fine vermiculite, and perlite and grown in controlled environmental chambers at 28°C, 60±5% relative humidity, and a 12h photoperiod under light intensity of 100 µmol/m^2/s. *Pseudomonas syringae* pv. *phaseolicola* (Burkn.) Downs (Pph) strain NPS3121 was grown in low-salt Luria-Bertani medium (Katagiri et al. 2002) at 30°C supplemented with 100 µg.ml^{-1} rifampicin. Young, fully expanded primary leaves were dip-inoculated into 10^8 CFU/ml aqueous suspension of containing 0.03% of Silwet L-77 (Lehle Seeds Co., Round Rock, TX). Inoculum preparation and bacterial population counts in the leaf apoplast were performed as previously described (Katagiri et al. 2002). Statistical significance of the mean difference between the bean genotypes was detected with two-tailed Student's *t* test. Symptoms were recorded 7 days after inoculation.

Callose deposition assay

G2333 seeds were germinated and grown as describe above. Fully expanded first trifoliate leaves were syringe-infiltrated with 1 µM of flg22 (Alpha Diagnostics, Inc., Santa Monica, CA) or water. Infiltrated leaves were collected 12 and 24 h post infiltration (hpi) and incubated in 90% ethanol at 37°C on an orbital shaker (30 rpm). After the chlorophyll had been removed, leaves were rinsed in 50% ethanol followed by a final rinse in water. Cleared leaves were stained with 0.1 µM aniline blue for 30 min and

maintained in 50% glycerol. Images (12 to 15 per sample) were captured with a Nikon Eclipse 80i fluorescent microscope (Nikon Corporations, Shinagawa-ku, Tokyo) equipped with DAPI filter (358 nm excitation and 461 nm emission) and a digital camera. Callose deposits were counted using the DotCount v1.2 software (Reuter 2012; http://reuter.mit.edu/software/dotcount/), using an image intensity threshold of 100 and dot sizes ranging from 5 to 500. Experiments were performed two times independently.

Gene expression analysis

Gene-specific primers were designed based on the common bean gene sequences from the Phytozome database (Table S3). The efficiency of each primer set was verified using a five-fold serial dilution of G2333 cDNA. Linear regression between the amount of cDNA template and the C_T values was calculated based on the efficiency standard curves for each primer to obtain the correlation coefficient ($R^2 > 0.95$) according to Schmittgen and Livak (2008).

Fully expanded primary leaves of G2333 were dip-inoculated with either 10^8 CFU/ml Pph suspension with 0.03% Silwet or 0.03% Silwet alone (mock-inoculation). Leaves were collected at 6, 12, and 24 h post-inoculation (hpi). Additionally, young first trifoliate leaves were immersed in 5 µM flg22 or water control for 30 min. Leaves were maintained in high humidity using a sterile petri dish and a humid paper, in a growth chamber at 28°C with 12 hour photoperiod, and were collected 6, 12 or 24 h after flg22 treatment. Total leaf RNA was extracted using the RNAeasy Plant mini kit (Qiagen, Valencia, CA) following the manufacturer's recommendations. The total RNA was quantified using a NanoDrop spectrophotometer (Thermo 367 Scientific, Rockford, IL).

Reverse transcription (RT) was performed using Takara RNA PCR kit (Clontech, Montain View, CA), 150 ng/µl of total RNA and 0.125 µM of oligo-dT primer, following the manufacturer's protocol. RT reaction was carried out at 50°C for 30 min and at 95°C for 5 min. Quantitative PCR (qPCR) reactions were carried out using 1 µl of cDNA (RT reaction above), 200 mM of each primer (Table S3),

and iTaq Fast SYBR green supermix (BioRad, Hercules, CA) reagents in a final volume of 20 µl. qPCR cycles consisted of one cycle of 95°C for 5 min, 40 cycles of 95 °C for 10 sec, 60 °C for 30 sec, followed by the dissociation curve default parameters using the Applied Biosystems 7300 thermocycler (Applied Biosystems, Foster City, CA). The gene expression levels of the treated samples relative to control samples was determinate with the $2^{-\Delta\Delta C}{}_{T}$ method (Livak and Schmittgen 2001). The *P. vulgaris INSULIN DEGRADING ENZYME (PvIDE;* Phvul.001G133200) gene (Borges et al. 2011) was used as the reference gene for the amount of RNA template across different reactions. The genes analyzed were: *FLS2-like (LRR)* (Phvul.005G149200), *FLS2-like (RLK)* (Phvul.002G196200), *COK-4-3* (Phvul.008G026900), *FER-like* (Phvul.008G030800), *NB-LRR* (Phvul.008G031200), *FUL-like* (Phvul.008G027800). All experiments were performed in three biological replicates and statistical analyses were conducted according using two-tailed Student's *t* test.

Results

Molecular markers define the genetic boundaries of the *Co-4* locus

Several markers closely linked to the anthracnose resistance *Co-4* locus of common bean have been identified (Vallejo and Kelly 2004; Ferreira et al. 2013); however the genomic boundaries of this locus are still elusive. Thus, we sought to refine its genetic structure by saturating this locus with new molecular markers and determine the segregation ratio of all possible polymorphic markers using the SEL 1308 x Black Magic F_2 population (Melotto and Kelly 2001). A SSR marker was developed based on the sequence of the previously identified clone (BAC $78L_{17}$) that spans part of the complex *Co-4* locus (Melotto et al. 2004). A 149-bp TA-repeat marker, named PvTA25, showed a co-dominant polymorphism between the parents of the mapping population as well as among bean lines carrying contrasting alleles at the *Co-4* locus (Fig. 1a). All three lines known to carry the resistant *Co-4* allele (G2333, SEL 1308, and TO) showed the same PvTA25 marker allele, while the susceptible genotypes Black Magic and SEL 1360 shared a PvTA25 DNA fragment of higher molecular weight (Fig. 1a). A dominant single locus AFLP

marker, named $E_{TGC}M_{GGT}$(135), was also identified to be linked to the *Co-4* locus, with presence or absence of a 135-bp PCR amplicon in the resistant and susceptible genotypes, respectively (Fig. 1b). Finally, an allele-specific SNP marker, named $PvSNP_{COK-4}$, was developed by aligning the *COK-4* sequences from SEL 1308 and Black Magic. Segregation analysis of this marker revealed its co-dominant nature, in which heterozygous *COK-4* individuals carried both the 700-bp susceptible and 1000-bp resistance alleles (Fig. 1c).

The PvTA25, AFLP $E_{TGC}M_{GGT}$(135), and $PvSNP_{COK-4}$ markers, as well as the previously SAS13 molecular marker, segregated as a single locus in our mapping population and were closely linked to the *Co-4* resistance gene as determined by chi-square statistical analysis (Table 1). The previously identified SBB14 and SH18 markers (Awale and Kelly 2001) were also tested in our segregating population and they showed a low chi-square *P*-value, indicating that they might not be a single locus (Table 1); nonetheless, both were found to be linked to *Co-4* based on linkage mapping analysis (Fig. 1d). The genetic order of all markers around the *Co-4* locus was estimated using the genotypic and phenotypic data of F_2 individuals. SAS13, PvTA25 and $PvSNP_{COK-4}$ were the closest markers to *Co-4*, all within 0.7 cM from each other, whereas $E_{TGC}M_{GGT}$(135) and SBB14 markers mapped 6.6 cM from *Co-4*, and SH18 mapped 10.4 cM apart of *Co-4* (Fig. 1d).

Physical location of the marker linked to *Co-4* in the bean genome

The most distant markers from the *Co-4* locus, SH18 showed sequence similarity with many regions in different chromosomes ($1 \times 10^{-149} \leq$ E-value $\leq 1 \times 10^{-142}$), including Pv08 at 54,381,505..54,382,406 region (Fig. 2). The markers SCARY20 and SCARC08, previously mapped at 1.2 cM and 7.8 cM apart the *Co-4* locus from TO genotype, respectively (Queiroz et al. 2004), show best alignment scores (BLASTN E-value = 0.0) on Pv08 at positions 28,034,637..28,034,904 and 7,414,123..7,415,017, respectively (Fig. 2). These markers also aligned with different regions on Pv08 as well as other bean chromosomes ($0.0 \leq$ E-value $\leq 1 \times 10^{-134}$). SH18, SCARY20, and SCARC08 were located at genome

regions with no predicted coding sequences. The SBB14 marker sequence, however, was found at the position 2,809,493..2,810,488 of Pv08 (BLASTN E-value = 0.0), approximately 240 Kb apart from *Co-4* locus (Fig. 2), in the 5'UTR from the Phvul.008G033800, a predict amylase gene, and does not have significant similarity to other regions of the bean genome.

Consistently, the tightly linked PvTA25, $PvSNP_{COK-4}$, and SAS13 marker sequences were also found at unique regions of that chromosome (Fig. 2). SAS13 is located in the Phvu1.008G028500 gene (BLASTN E-value = 0.0), PvTA25 is 650 bp apart from the Phvu1.008G029500 gene (BLASTN E-value = 4×10^{-40}), and the $PvSNP_{COK-4}$ primers align within the Phvul.008G028400 gene (BLASTN E-value = 1×10^{-7}), with a predicted DNA fragment of 724 bp, similar to the one amplified from Black Magic (Fig. 1C). None of these markers aligned at a different genomic location, confirming the single locus segregation data analysis (Table 1; Awale and Kelly 2001; Melotto and Kelly 2001).

Previously, we have physically located the BAC $78L_{17}$ at the chromosome Pv08 as revealed by FISH analysis (Melotto et al. 2004). BLASTN analysis of the BAC $78L_{17}$ against the common bean genome v1.0 (Schmutz et al. 2014) located this BAC in the 2,345,000..2,464,000 region of Pv08 (BLASTN E-value = 0.0). BLASTN alignment between the previously identified *COK-4* gene sequence from SEL 1308 (Melotto et al. 2004) and the bean genome revealed 20 significant hits on Pv08 (E-value $\leq 1 \times 10^{-64}$ and identity > 70%; Table 2), extending our previous description of the BAC $78L_{17}$ region that contains ten sequences with similarity to *COK-4* (Fig. 2). Out of the 20 hits, 18 lye on a genomic region with predicted transcripts encoding a tyrosine kinase domain (Table S4). The COK-4 copies were numbered according to their order of location in the genome (Table 2). Outside the Pv08 chromosome, only one match for the SEL 1308 *COK-4* gene was found on chromosome Pv05 (Table 2), in a region with no predicted transcripts. Furthermore, none of the genetic markers tightly-linked to *Co-4* were found on chromosome Pv05; thus we have not considered it as a possible location for this anthracnose resistance locus. Therewith, we predict that the *Co-4* gene is most likely to be within the 325 Kbp region (Chr08:2,245,000..2,570,000) adjacent to the Pv08 telomere (Richard et al. 2013) based on genetic

linkage (Fig. 1), the genomic locations (Fig. 2) of the markers closed linked to *Co-4* (SAS13, PvTA25, $PvSNP_{COK-4}$), and the presence of multiple potentially *COK-4* paralogs in that region (Fig. 2 and Table 2).

The *Co-4* locus is enriched with putative kinases member of the CrRLK1 family

Once we determined the 325 Kbp region on Pv08 (Chr08:2,245,000..2,570,000) as the most likely to contain the *Co-4* locus, we sought to characterize its gene content. Fourty-nine transcripts were identified (Fig. 2 and Table S4) with support of expression data such as RNA-seq and EST (Phytozome). Function annotation of the transcripts revealed three putative transcription factors next to each other (two SRF-type transcription factors and one Myb-like domain), three DSBA-like (disulfide oxidoreductase-like), three COBRA-like, one NB-LRR (Nucleotide Binding–Leucine-Rich Repeat) domains gene, eleven genes showing various putative functions, and four with unknown function (Table S4). Twenty-four transcripts in the *Co-4* region are predicted to encode protein kinases, with significant similarity to the predicted SEL 1308 COK-4 protein (BLASTP E-value $\leq 2 \times 10^{-31}$; Table S5). Four of the *COK-4* gene copies showed the highest similarity (BLASTP E-values = 0.0) to the protein COK-4 from SEL 1308: Phvul.008G028300 (identity = 81.5%), Phvul.008G028400 (identity = 78.4%), Phvul.008G028500 (identity = 83.4%) and Phvul.008G028600 (identity = 84.0%) (Table S5). All predicted kinases in the *Co-4* region showed significant similarity with members of the Arabidopsis CrRLK1 family, FERONIA (FER), ANXUR2 and AT5G39000 (Table S5). In addition, BLASTP analysis of COK-4_SEL1308 against the non-redundant database of NCBI showed high similarity with CrRLK1 family members from different plant species, with FER and ANXUR being also overrepresented (Table S6). Phylogenetic analysis of these proteins showed that COK-4 form a major clade with serine/threonine kinases from *Glycine max*, *Cicer arietinun*, *Lotus japonicus*, *Theobroma cacao*, and *Malus domestica*, as well as Pto-like proteins from three *Solanum* species and *Capsicum chinense* (Fig. 3).

Among the 24 putative kinases in the *Co-4* locus, 20 are predicted to encode a single kinase domain protein, and four seem to encode both a kinase and a malectin domain (Table S5). Malectin is an

endoplasmic reticulum membrane-anchored domain, and is found in proteins of the CrRLK family (Kessler et al. 2010), among other protein families. Three of these proteins located on one edge of the *Co-4* locus (Fig. 2b) showed high similarity with the *Arabidopsis thaliana* CrRLK family member FERONIA (FER) (BLASTP E-value = 0.0; Table S5). The fourth putative protein with a kinase and malectin domain, encoded by Phvul.008G030200, is similar to a malectin/receptor-like protein kinase from Arabidopsis with no specific function established yet (Table S5).

Kinase proteins in the *Co-4* locus seems to be evolutionarily related to COK-4

The great number of copies of the *COK-4* genes in the region of the *Co-4* resistance locus indicates that gene duplication events may have taken place in this region during the course of common bean evolution, resulting in the genetic and phenotypic variations observed among bean lines, including TO and G2333 (Long et al. 2013). Thus, we investigated the phylogenetic relationship of proteins similar to COK-4 using the common bean proteome (Phytozome).

Owing to its highly conserved kinase domain, the COK-4_SEL1308 protein showed significant similarity to protein kinases throughout the common bean genome, including the kinases at the *Co-4* locus. Therefore, we used these first 100 best hits identified by BLASTP (E-value $\leq 2 \times 10^{-37}$) to identify the ones that formed a single clade with COK-4_SEL1308. All of these proteins contain a kinase domain annotated as belonging to the protein superfamily PTKc cd14066 conserved kinase domain (NCBI conserved domain database), which was considered to perform the phylogeny analysis (Fig. S1). One of the predicted kinase on *Co-4* locus (Phvul.008G029800) showed low similarity to the SEL 1308 COK-4 protein and it was not in the best 100 kinase matches used for the phylogeny analysis. Also, another kinase at the *Co-4* locus clustered close to the putative COK-4 paralogs (Fig. S1), however its encoding gene (Phvul.008G031100.1) was not considered a *COK-4* copy as it does not have significant nucleotide similarity to the SEL 1308 *COK-4* gene. Interestingly, all 18 *COK-4* copies located on the *Co-4* locus (Table S5) formed a single cluster with COK-4_SEL 1308 (Fig. 4). Four proteins showed to be the

closest related to COK-4 form SEL 1308, forming a small sub-clade, which included Phvul.008G028300, Phvul.008G028400, Phvul.008G028500 and Phvul.008G028600 (Fig. 4), confirming the BLASTP results. The four kinases predicted to encode a malectin-kinase protein (Table S5) also formed a single cluster with other RLK proteins from Pv04 (Fig. S1). These results indicate that the kinases present at the *Co-4* locus are closer related to each other than they are to other kinases in the bean genome.

Co-4 locus seems to be involved in bean innate immune response

Phylogeny and BLAST analyses indicate that the majority of putative proteins in the *Co-4* locus are similar to CrRLK1 proteins as described above. Members of the CrRLK1 family, such as *FER*, *ANXUR*, *HERCULES* and *THESEUS* are known to be involved in plant growth and reproduction (Lindner et al. 2012), but recent results have shown that *FER* in particular, is involved also in PAMP-triggered immunity (Keinath et al. 2010). Thus, we reasoned that the predicted kinases at the *Co-4* locus could be regulated by pathogens other than *C. lindemuthianum* as originally identified, and play a role in broad immune response. To test this hypothesis, we used the flg22 peptide found at the N-terminus of bacterial flagellin, which is typically used to assess the PTI response in plants, such as Arabidopsis, *Lotus japonicus*, and common bean (Navarro et al. 2004; Hou et al. 2011; Lopez-Gomez et al. 2011).

First, we determined whether flg22 could induce PTI in G2333 by assessing callose deposition in treated leaves, a hallmark PTI response in plants (Boller and Felix 2009; Hou et al. 2011). In fact, G2333 leaves showed high numbers of callose deposits 12 hpi. The number of callose deposits decreased after 24 h after flg22 treatment; nonetheless it was still higher than that of the water control (Fig. 5a and b). To further confirm that flg22 can trigger defense responses in G2333, we assessed the expression of two putative Arabidopsis FLS2 orthologs in beans. The *FLS2-like (LRR)* (Phvul.005G149200) is predicted to have only the LRR (leucine-rich repeat) domain, while the *FLS2-like (RLK)* (Phvul.002G196200) has both LRR and kinase domains similar to FLS2 (Zipfel et al. 2004) and is the protein with highest similarity (BLASTP E-value = 0.0 and 44% identity) to the Arabidopsis FLS2 in the bean reference

genome. In agreement with the induction of PTI assessed by callose deposition, both *FLS2-like (LRR)* and *FLS2-like (RLK)* were significantly induced by flg22. The up-regulation of the *FLS2-like (RLK)* gene lasted until 24 h after flg22 treatment (Fig. 5c).

Next, primers were designed for all genes in the *Co-4* locus, including all *COK-4* copies, however gene-specific and/or efficient primers could be obtained for all of them. Thus, we were able to selected fours genes, representing different function in the locus and for which gene-specific and efficient RT-qPCR primers could be designed, to test their expression after flg22 treatment. Three of them were found to be modulated by flg22: *COK-4-3* (Phvul.008G026900) was significantly induced at 24hpi, while *FER-like* (Phvul.008G030800) and the putative transcription factor *FUL-like* (Phvul.008G027800) showed transient repression at early time points and returned to basal levels at 24 hpi (Fig. 5c). Finally, the only *NB-LRR* domains coding gene found at the *Co-4* locus (Phvul.008G031200) was not responsive to flg22 treatment (Fig. 5c).

To determine whether live bacteria also regulate the expression of these genes, we inoculate G2333 plants with the bacterium *P. syringae* pv. *phaseolicola* (Pph). The G2333 seems to be tolerant to Pph, as these plants supported a large bacterial population in their leaf apoplast since the first day after inoculation (Fig. 6a) and yet, no symptoms were observed even after 7 days post inoculation (Fig. 6b). By contrast, the susceptible cultivar Beluga supported high bacterial titers in the apoplast and showed typical halo blight symptoms later in the infection cycle (Fig. 6a and b). Analysis of gene expression in inoculated G2333 plants revealed repression of both *FLS2-like (LRR) and FLS2-like (RLK)* genes as bacterial infection progressed (*i.e.*, 12 and 24 hpi) suggesting a low level of defense response that correlated well with high bacterial titer in the leaves. Similarly, SEL 1308 incompatible response to *C. lindemuthianum* showed to involve repression of PTI pathway and down-regulation of the *FLS2-like (LRR)* after fungus infection (Oblessuc et al. 2012). The *COK-4-3* gene was also down-regulated as early as 6 hpi, returning to normal levels at 24 hpi. In contrast, the *FER-like* gene was up-regulated after Pph infection, also returning to normal levels after 24 hpi (Fig. 6c). The putative transcription factor *FUL-like*

showed no change in transcript levels in response to Pph, while the *NB-LRR* domain coding gene was slightly repressed in the initial phase of Pph infection (6 hpi), maintaining normal levels after 12 hpi (Fig. 6c). Altogether, these results suggest that the kinases in the *Co-4* locus are involved in basal immunity as they are inversely regulated in plants undergoing immune response (*i.e*. flg22 treatment) or infected with a phytopathogen (*i.e*. large Pph population in the leaves). Additionally, our results suggest that other genes in the *Co-4* locus (transcription factors and NB-LRR-resistance gene analogs) might be involved with either resistance or susceptibility as they are only regulated by either fgl22 or Pph infection.

Discussion

Plant responses to pathogens implicate in drastic changes in host genes expression and protein turnover resulted from pathogens recognition and activation/inactivation of a complex chain of metabolic pathways. Broadly, the final outcome of the plant response is resistance or susceptibility to the pathogen (Spoel and Dong 2012). Understanding these molecular mechanisms involved in plant immunity is crucial for crop improvement. In the present study, we have determined the most probable location of the *Co-4* locus of common bean, assessed the phylogenetic relationship of the predicted kinases found in the locus, and provided genetic evidence that *Co-4* may have a role in basal immunity in addition to its originally assigned function in resistance to anthracnose.

The genomic structure of *Co-4* was first analyzed through the molecular mapping of new markers linked to anthracnose resistance. Linkage analysis showed that the newly developed markers PvTA25 and $PvSNP_{COK-4}$, as well as SAS13 are closely linked to each other and to *Co-4* resistance gene. The identified genetic distance between the markers, however, may be overestimated due to inherent restrict recombination frequency observed in small mapping populations (Liu 1998), such as the one used here. Thus, these markers could be physically closer to each other than the genetic linkage analysis predicted. Indeed, PvTA25, $PvSNP_{COK-4}$, SAS13 markers were located in a small interval also covered by the clone BAC $78L_{17}$ in the common bean chromosome Pv08, confirming that they form a unique locus on the

genome. In addition, mismatches between the markers primers sequences and the genome of the bean genotype G19833 indicate that these three markers either could not be amplified by PCR in this bean line or would show different allele size for the PCR amplicon, supporting the linkage of these markers to the resistance *Co-4* locus.

The G19833 common bean genotype is resistant to some races of *C. lindemuthianum*, but seems not to contain resistant alleles of the *Co-4* gene (Kelly and Vallejo 2004; Ferreira et al. 2013). Our results showed that the new markers PvTA25 and $PvSNP_{COK-4}$ amplify the susceptible alleles of the bean line Black Magic showing their transferability across bean genotypes in addition to be breeder-friendly markers, in which polymorphism could be easily observed by PCR technique. Thus, these markers may be an important tool to be applied in molecular breeding for the development of cultivars containing the resistant allele in the *Co-4* locus.

Genetically linked markers together with the recent release of the common bean genome (Schmutz et al. 2014) enabled us to further refine the genomic structure of the *Co-4* locus. In addition to the markers PvTA25, $PvSNP_{COK-4}$, SAS13 (Young et al. 1998), all publicly available sequences linked to *Co-4* were located in the common bean genome including the markers SBB14, SH18 (Awale and Kelly 2001), SCARY20 and SCARC08 (Queiroz et al. 2004), as well as the BAC $78L_{17}$ clone and *COK-4* gene (Melotto and Kelly 2001; Melotto et al. 2004). With the results of this alignment analysis we could establish the most probable region for the locus containing the functional *Co-4* gene, a 325 Kbp-long sequence at the end of chromosome 8 (Pv08) of the common bean reference genome (G19833 genotype; Schmutz et al. 2014).

Interestingly, this region contains 18 copies of the *COK-4* coding sequence originally identified by Melotto and Kelly (2001), considerably extending the physical boundaries of the *Co-4* locus beyond the clone BAC 78_{L17}. This clone, isolated from the bean cultivar Sprite, was reported to have five COK-4 kinases (Melotto et al. 2004). However, the corresponding region in the G19833 genome contains ten COK-4 kinases confirming the evolutionary complexity of the locus, where the number of putative *COK*-

4 paralogs varies according to the bean genotype. Furthermore, all of the *COK-4* coding sequences found on bean reference genome are transcribed into RNA as confirmed by RNA-seq and EST mapping analysis (Schmutz et al. 2014) suggesting that they are all active genes. A single *COK-4* related sequence was found in another chromosome (Pv05); however, it may have lost its function during the translocation as no *COK-4* transcript mapped to Pv05. The putative COK-4 parologs may altogether contribute to *Co-4*-based resistance or at least one of these genes might be the single functional *Co-4* resistance gene. These alternatives remain to be experimentally validated.

Previously, *COK-4* was regarded as a *Pto-like* gene, in which all the COK-4 homologs studied formed a cluster with the Pto protein of tomato (Melotto et al. 2004). Surprisingly, the present analysis showed that the majority (65.2%) of putative COK-4 kinases showed highest similarity to the RLK FERONIA (FER, At3g51550) of Arabidopsis, in addition to FER-like from others species, including tomato. Other members of the CrRLK1 kinase subfamily, such as ANXUR, were also found to be similar to COK-4. However, COK-4 seems to be closely related to Pto-like kinases of other *Solanum* species. This apparent discrepancy between previous and current analysis may be due to the much smaller database available at the time of the first study. In addition, COK-4 may have clustered with Pto-like protein from *Solanum ssp.* because these plants are not well studied as *Solanum lycopersicum* and fully sequenced and/or annotated genomes are still not available. This high phylogenetic relationship among the COK-4 kinase and only two members of the CrRLK1 family, FER and ANXUR, suggests that the COK-4 encoding gene underwent extensive duplication that may or may not have retained the kinase function. Further biochemical analyses are needed to verify the activity of COK-4 proteins in Pv08.

Among subfamilies of RLK, the CrRLK1 family has emerged recently as sensors for cell wall integrity that is involved in cell growth in different physiological contexts (Wolf et al. 2012). FER and ANXUR are very similar to each other in the CrRLK1 family of proteins in Arabidopsis; both are necessary for fertility through self- and non-self-recognition (Wolf et al. 2012). In plants, genes located in the same region of the genome can be involved in the same pathway and be co-regulated, forming operon-

like gene clusters (Zmasek and Godzik 2011; Boycheva et al. 2014). The *Co-4* locus might be such an example where *COK-4* paralogs and *FER*-like genes may have evolved from similar functions and some other genes, such as the *FUL-like* gene, in this locus may be involved in the same pathway. The *FUL-like* (*FRUITFULL*-like) is a putative MADS box transcription factor, and its homolog in Arabidopsis (AT5G60910) is an *AGAMOUS-like 8* called *FRUITFULL* (*FUL*) because of its involvement in the control of flowering time, fruit development, and determinacy (Pabón-Mora et al. 2012). In addition, the soybean MADS-box transcription factor modulates floral organ numbers, petal identity, and sterility (Huang et al. 2014). Here the common bean *FUL-like* was co-regulated in the same direction as *FER-like* during PTI induction, suggesting that *FUL-like* may have evolved to perform a related function.

In general, CrRLK1 family members seem to be involved, at least in part, in modulation of ROS production to regulate cell growth in different developmental stages and hormone signaling pathways, such as ethylene (ET), jasmonate (JA), and salicylic acid (SA), after cell wall damage perception (Wolf et al. 2012). ROS production and activation of ET, JA, and SA pathways are well-known plants responses to pathogens, thus CrRLK1 family members are potential membrane receptors that could be active during plant-pathogen interaction. In fact, FER is the only member of CrRLK1 family that has been associated with plant immunity so far (Wolf et al. 2012).

The striking similarity between COK-4 and FER prompted us to check whether the *Co-4* locus could be involved in the bean innate immunity. In conditions where bean immunity was activated, *i.e.* flg22 treatment, we observed that expression of the *COK-4-3* was significantly up-regulated along with both *FLS2-like* genes. On the other hand, the *FER-like* gene was strongly repressed during PTI, that is, at the same time points of induction of *FLS2-like* and high callose deposition after flg22 treatment. These data suggest that both *COK-4-3* and *FER-like* genes play distinct roles in PTI responses. While *COK-4-3* may be a positive regulator of PTI, the *FER-like* gene on the *Co-4* locus may be involved in repression of PTI. To further support this hypothesis, we assessed the expression of these genes in bean plants that are tolerant to the bean pathogen Pph (*i.e.*, these plants are symptomless but support high bacterial population

in their leaves typical of susceptible interactions). At the same time that bacterial population was high and *FLS2-like* genes were repressed in these plants indicating low level of PTI, *COK-4-3* was also repressed and *FER-like* was up-regulated. Taken together, these findings provide strong genetic evidence that both *COK-4* and *FER-like* may be involved in the basal immune response to different pathogens.

COK-4 may have an evolution history with *FER* but both assumed different functions by either *COK-4* losing the malectin domain or *FER* gaining that domain. Although evolutionary events in eukaryotes that distinguish a protein from its closest ancestor have been studied, it was found that in general domain loss is more common than domain gain and that the exchange of a domain is rare (Björklund et al. 2005; Zmasek and Godzik 2011). Therewith, *COK-4* may be a PTI defense response activator, while *FER-like* may acts as PTI inhibitor. Continuous studies on the evolution of new biochemical functions emerging in the *Co-4* locus through the *FER-like* and *COK-4* genes should further the current understanding of the molecular pathways underlying bean immunity against a broad range of pathogens. Nonetheless, our results come up as important directions, establishing the boundaries of the *Co-4* locus, providing additional markers for molecular breeding as new tools for employing anthracnose resistant in beans, and reinforcing the role of the putative COK-4 kinases in common bean basal immunity.

Author Contributions

Performed experiments: CF, PRO, MM. Analyzed data: PRO, MM. Conceived and coordinated the project: MM. Wrote the manuscript: PRO, MM. All authors have read and approved the final version of the manuscript.

Acknowledgements

The authors thank Dr. L.E.A. Camargo for hosting CF and MM in his lab during the screening for AFLP markers, Dr. J.D. Kelly for common bean seeds and the *Colletotrichum lindemuthianum* isolate, Dr.

David Guttman for the *Pseudomonas phaseolicola* strain, CAPES/Science Without Borders for the post-doctoral fellowship to PRO (award #9773-13-4), and FAPESP for the Undergraduate Research Assistantship to CF (award #01/11218-0). This study was funded by FAPESP – Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil (Grant #00/09049-2) to MM.

Conflict of Interests

The authors declare that they have no conflict of interests.

Ethical standards

All experiments described in this manuscript comply with the current laws of the country in which they were performed.

Figure legends

Fig. 1 Molecular marker banding pattern in resistant and susceptible genotypes of common bean. G2333, SEL 1308, and TO are homozygous anthracnose resistant genotypes, Black Magic and SEL1360 are homozygous anthracnose susceptible genotypes. **a** and **b** Silver-stained polyacrylamide gel showing polymorphisms (scored bands indicated by the arrows) detected with the PvTA25 SSR primers (**a**) and the $E_{TGC}M_{GGT}$(135) AFLP (**b**) markers. **c** Ethidium bromide-stained agarose gel showing PCR-amplified DNA fragments with the $PvSNP_{COK\text{-}4}$ marker (700 bp lower band and 1000 bp upper band); legends on top of the lanes indicate the genotype of F_2 individuals from the cross between SEL 1308 and Black Magic. **d** Predicted genetic distance among the markers linked to the *Co-4* locus. The numbers on the top are estimated distances between molecular markers in centiMorgans (cM) calculated with the MapMaker

software considering LOD score > 3 as threshold. The linkage map diagram was created with the MapChart 2.2 program (Voorrips 2002) and the scale was set as 10 mm/cM.

Fig. 2 Genomic boundaries and structure of the bean anthracnose resistance locus *Co-4*. **a** Modified genome browser representation of the 325 Kbp in the chromosome 8 (Chr08) spanning the *Co-4* locus that contains the markers SAS13, PvTA25, $PvSNP_{COK-4}$, as well as the BAC $78L_{17}$ sequences. Physical location was determined by BLASTN analysis (threshold E-value $\leq 1 \times 10^{-5}$) of the marker sequences against the common bean genome v.1.0 (Schmutz et al. 2014; http://www.phytozome.net/). **b** Predicted transcripts in *Co-4* region are shown below its genomic positions in (**a**). Color codes, as indicated in the legend below the figure, represent predicted gene functions. Asterisks above the transcript indicate genes that were analyzed by RT-qPCR.

Fig. 3 COK-4_SEL 1308 predicted protein clustered with members of the *Catharanthus roseus* RLK1 (CrRLK1) protein family (mainly FERONIA-like and ANXUR-like) from diverse plant species. Phylogenetic analysis was performed with the maximum parsimony method using the MEGA 5.05 software (Tamura et al. 2011). Bootstrap support values are adjacent to the tree nodes.

Fig. 4 COK-4_SEL 1308 predicted protein clustered with common bean kinases located in the *Co-4* genomic region of Pv08. Phylogenetic analysis of predicted amino acid sequence was performed with the maximum parsimony method using the MEGA 5.05 software (Tamura et al. 2011). Bootstrap support values are provided adjacent to nodes. The diagram shown in front of the transcript name represents the single kinase domain (gray rectangles) within the protein. The numbers indicate the total amino acids of each protein. Diagram was adapted from the protein domain view of Phytozome. Only the clade containing the COK-4 is shown (refer to Fig. S1 for the entire tree with the top 100 kinases most similar to COK-4_SEL1308 in the G19833 reference genome).

Fig. 5 G2333 responses to the PAMP flg22. **a** Graph shows the average number of callose deposits per mm^{-2} of G2333 leaf tissue infiltrated with 1 µM flg22 or water. Results are shown as average of 108 to 135 images in three independent biological replicates ± standard error. **b** Representative images (100 x magnification) of aniline blue stained G2333 leaves 12 h or 24 h post incubation (hpi) with flg22 or water. **c** Expression of the indicated genes (x-axis) in G2333 leaves immersed in 5 µM flg22 at 6, 12, and 24 hpi relative to the their expression in water-immersed leaves (control) considered as 1. Data points are average of at least two biological replicates ($n \geq 6$ ± standard error). Asterisk above the bars of all graphs indicate statistical significance calculated with Student's *t* test (**p<0.01, ***p<0.001).

Fig. 6 G2333 responses to *Pseudomonas syringe* pv. *phaseolicola* (Pph). **a** G2333 showed tolerance to Pph (NPS3121), with no bacterial growth in the leaf apoplast of fully expended primary leaves dipped inoculated with 10^8 CFU/ml. **b** Halo blight symptoms were observed after 7 days of inoculation only for Beluga genotype. **c** Expression of the indicated genes (x-axis) in G2333 leaves dipped in 10^8 CFU/ml of Pph relative to the their expression in mock-inoculated leaves (control) considered as 1. Data points are average of at least two biological replicates ($n \geq 6$ ± standard error). Asterisk above the bars of all graphs indicate statistical significance calculated with Student's *t* test (**p<0.01, ***p<0.001).

Table legends

Table 1 Segregation analysis of molecular markers linked to the *Co-4* gene using an F_2 mapping population derived from the SEL 1308 x Black Magic genetic cross.

Table 2 Common bean genome regions similar to the SEL 1308 *COK-4* gene.

Supplemental material legends

Fig. S1 Phylogenetic analysis of the top 100 protein kinases with highest similarity to the predicted COK-4_SEL1308 protein. The top 100 hits were obtained from BLASTP analysis (threshold E value $\leq 1 \times 10^{-20}$ and identity > 30%) using COK-4_SEL1308 as query against the common bean proteome database available at Phytozome. The phylogenetic tree was obtained with the maximum parsimony method using the MEGA 5.05 software (Tamura et al. 2011). Bootstrap support values are adjacent to the tree nodes. *Co-4* locus-associated kinases formed a single cluster (red box) and kinase/malectin proteins formed another sub-cluster (blue box).

Table S1 Simple sequence repeats (SSR) found in the BAC $78L_{17}$ insert sequence using the SSRIT software (http://www.gramene.org/gramene/searches/ssrtool).

Table S2 Primer sequences for the newly developed SSRs, AFLP, and $PvSNP_{COK-4}$ markers.

Table S3 Primer sequences designed to assess gene expression analysis by RT-qPCR.

Table S4 Predicted transcripts of the *Co-4* genomic region in chromosome 8 (325 Kbp; Chr08:2,245,000..2,570,000) based on RNA-seq data for the bean genome (Phytozome). Transcripts in bold letters code for predicted protein kinases. Transcripts in bold underlined letters were identified as copies of *COK-4*. Annotation is based on the Phytozome database (http://www.phytozome.net/).

Table S5 Predicted common bean (Pv) proteins in the 325 Kbp surrounding the *Co-4* genomic region (Pv08, 2,245,000..2,570,000) with significant similarity (BLASTP) to the predicted COK-4_SEL 1308 protein and their putative Arabidopsis (At) orthologs.

Table S6 Putative COK-4 protein orthologs in 35 different plant species. The SEL 1308 COK-4 predicted protein was used as query for BLASTP analysis using the non-redundant (nr) protein database from NCBI (threshold E-values $\leq 1 \times 10^{-20}$ and identity > 30%). Protein domain superfamily was inferred based on the NCBI conserved domain database (http://www.ncbi.nlm.nih.gov/cdd).

References

Awale HE, Kelly JD (2001) Development of SCAR markers linked to *Co-4*^{2} gene in common bean. Annu Rep Bean Improv Coop 44:119–120.

Balardin RS, Kelly JD (1998) Interaction between races of *Colletotrichum lindemuthianum* and gene pool diversity in *Phaseolus vulgaris*. J Am Soc Hort Sci 123:1038–1047.

Björklund ÅK, Ekman D, Light S, Frey-Skött J, Elofsson A (2005) Domain rearrangements in protein evolution. J Mol Biol 353(4):911-923. doi: 10.1016/j.jmb.2005.08.067

Boller T, Felix G (2009) A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. Ann Rev Plant Biol 60: 379–406. doi: 10.1146/annurev.arplant.57.032905.105346.

Borges A, Tsai SM, Caldas DGG (2011) Validation of reference genes for RT-qPCR normalization in common bean during biotic and abiotic stresses. Plant Cell Rep 5:827-838. doi:10.1007/s00299-011-1204-x

Boycheva S, Daviet L, Wolfender JL, Fitzpatrick TB (2014) The rise of operon-like gene clusters in plants. Trends in Plant Sc 7:447-459. doi: 10.1016/j.tplants.2014.01.013

Broughton WJ, Hernandez G, Blair M, Beebe S, Gepts P, Vanderleyden J (2003) Beans (*Phaseolus* spp.)—model food legumes. Plant Soil 252:55–128. doi: 10.1023/A:1024146710611

Chinchilla D, Bauer Z, Regenass M, Boller T, Felix G (2006) The Arabidopsis receptor kinase FLS2 binds flg22 and determines the specificity of flagellin perception. Plant Cell 18: 465–476. doi: http://dx.doi.org/10.1105/tpc.105.

Chinchilla D, Zipfel C, Robatzek S, Kemmerling B, Nurnberger T, Jones JDG, Felix G, Boller T (2007) A flagellin-induced complex of the receptor FLS2 and BAK1 initiates plant defence. Nature 448:497-U412. doi:10.1038/nature05999

Dodds PN, Rathjen JP (2010) Plant immunity: towards an integrated view of plant-pathogen interactions. Nature Rev Genetics 11(8):539–48. doi:10.1038/nrg2812

Ferreira JJ, Campa A, Kelly JD (2013) Organization of genes conferring resistance to anthracnose in common bean. In: Varshney RK, Tuberosa R (eds) Translational genomics for crop breeding, 1st ed. Wiley Blackwell, pp 151-182.

Fouilloux G (1979) New races of bean anthracnose and consequences on our breeding programs. In: Maraitre H, Meyer JA (eds) Disease of tropical food crops. Université Catholique de Louvain la Neuve, Belgium, pp 221–235.

Hazen SP, Leroy P, Ward R (2002) AFLP in *Triticum aestivum* L.: Patterns of genetic diversity and genome distribution. Euphytica 125:89–102. doi: 10.1023/A:1015760802026

Hou S, Mu R, Ma G, Xu X, Zhang C, Yang Y, Wu D (2011) *Pseudomonas syringae* pv. *phaseolicola* effector HopF1 inhibits pathogen-associated molecular pattern-triggered immunity in a RIN4-independent manner in common bean (*Phaseolus vulgaris*). FEMS Microbiol Letters 323(1):35–43. doi:10.1111/j.1574-6968.2011.02356.x

Huang F, Xu G, Chi Y, Liu H, Xue Q, Zhao T, Gai J, Yu D (2014) A soybean MADS-box protein modulates floral organ numbers, petal identity and sterility. BMC Plant Biol 14(1):89. doi:10.1186/1471-2229-14-89

Jones JDG, Dangl JL (2006) The plant immune system. Nature 444(7117):323–9. doi:10.1038/nature05286

Katagiri F, Thilmony R, He SY (2002) The *Arabidopsis thaliana-Pseudomonas syringae* interaction. In: Somerville CR, Meyerowitz EM (eds) The Arabidopsis Book. American Society of Plant Biologists, Rockville, pp 1-35. doi:10.1199/ tab.0039

Keinath NF, Kierszniowska S, Lorek J, Bourdais G, Kessler SA, Shimosato-Asano H, Panstruga R (2010) PAMP (pathogen-associated molecular pattern)-induced changes in plasma membrane compartmentalization reveal novel components of plant immunity. J Biol Chem 2*85*(50):39140–39149. doi:10.1074/jbc.M110.160531

Kelly JD, Vallejo VA (2004) A Comprehensive review of the major genes conditioning resistance to anthracnose in common bean. HortSc 39(6):1196–1207.

Kessler SA, Shimosato-Asano H, Keinath NF, Wuest SE, Ingram G, Panstruga R, Grossniklaus, U (2010) Conserved molecular components for pollen tube reception and fungal invasion. Science 330(6006):968–71. doi:10.1126/science.1195211

Kuć J (1982) Induced immunity to plant disease. Bioscience 32(11):854-860. doi: 10.2307/1309008

Lander ES, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: an interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174–181. doi: 10.1016/0888-7543(87)90010-3

Lindner H, Müller LM, Boisson-Dernier A, Grossniklaus U (2012) CrRLK1L receptor-like kinases: not just another brick in the wall. Curr Opin Plant Biol 15(6):659–69. doi:10.1016/j.pbi.2012.07.003

Livak KJ, Schmittgen TD (2001) Analysis of relative gene expression data using real-time quantitative PCR and the $2^{-\Delta\Delta C}{}_{T}$ method. Methods 25(4):402–8. doi:10.1006/meth.2001.1262

Liu BH (1998) Statistical genomics: linkage, mapping and QTL analysis. Cleveland: Cleveland: CRC Press 611.

Long M, VanKuren NW, Chen S, Vibranovski MD (2013) New gene evolution: little did we know. Ann Rev Genet 47:307-333. doi: 10.1146/annurev-genet-111212-133301

Lopez-Gomez M, Sandal N, Stougaard J, Boller T (2011) Interplay of flg22-induced defence responses and nodulation in *Lotus japonicus*. J Exp Bot. online. doi:10.1093/jxb/err1291

Marchler-Bauer A, Zheng C, Chitsaz F, Derbyshire MK, Geer LY, Geer RC, Gonzales NR, Gwadz M, Hurwitz DI, Lanczycki CJ, Lu F, Shennan Lu, Marchler GH, Song JS, Thanki N, Yamashita RA, Zhang D, Bryant SH (2013) CDD: conserved domains and protein three-dimensional structure. Nucleic Acids Res 41(D1):D384-52. doi: 10.1093/nar/gks1243

Melotto M, Balardin RS, Kelly JD (2000). Host-pathogen interaction and variability of *Colletotrichum lindemuthianum*. In: "*Colletotrichum* host specificity, pathology, and host-pathogen interaction". D. Prusky, S. Freeman, and M.B. Dickman (eds.), pp 346-361. APS Press, St Paul, MN, USA.

Melotto M, Kelly JD (2001) Fine mapping of the *Co-4* locus of common bean reveals a resistance gene candidate, COK-4, that encodes for a protein kinase. Theor Appl Genet 103(4):508–517. doi:10.1007/s001220100609

Melotto M, Coelho MF, Pedrosa-Harand A, Kelly JD, Camargo LEA (2004) The anthracnose resistance locus *Co-4* of common bean is located on chromosome 3 and contains putative disease resistance-related genes. Theor Appl Genet 109(4):690–9. doi:10.1007/s00122-004-1697-6

Michelmore RW, Paran J, Kesseli RV (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregation populations. Proc Natl Acad Sci 88:9828–9832. doi: 10.1073/pnas.88.21.9828

Navarro L, Zipfel C, Rowland O, Keller I, Robatzek S, Boller T, Jones JDG (2004) The transcriptional innate immune response to flg22. Interplay and overlap with Avr gene-dependent defense responses and bacterial pathogenesis. Plant Physiol 135:1113–1128. doi: http://dx.doi.org/10.1104/pp.103.036749

Oblessuc PR, Borges A, Chowdhury B, Caldas DGG, Tsai SM, Camargo LEA, Melotto M (2012) Dissecting *Phaseolus vulgaris* innate immune system against *Colletotrichum lindemuthianum* infection. PLoS ONE *7*(8):e43161. doi:10.1371/journal.pone.0043161

O'Connell RJ, Thon MR, Hacquard S, et al. (2012). Lifestyle transitions in plant pathogenic *Colletotrichum* fungi deciphered by genome and transcriptome analyses. Nature Genetics 44(9):1060-1065. doi:10.1038/ng.2372

Pabón-Mora N, Ambrose BA, Litt A (2012) Poppy APETALA1/FRUITFULL orthologs control flowering time, branching, perianth identity, and fruit development. Plant Physiol 158(4):1685-1704. doi: 10.1104/pp.111.192104

Queiroz VT, Sousa CS, Costa MR, Sanglad DA, Arruda KMA, Souza TLPO, Ragagnin VA, Barros EG, Moreira MA (2004) Development of SCAR markers linked to common bean anthracnose resistance genes *Co-4* and *Co-6*. Ann Rep Bean Improv Coop 47: 249-250.

Reuter M (2012) Image Analysis: Dot Count. DotCount v1.2. http://reuter.mit.edu/software/dotcount/

Richard MMS, Chen NWG, Thareau V, Pflieger S, Blanchet S, Pedrosa-Harand A, Iwata A, Chavarro C, Jackson SA, Geffroy V (2013) The subtelomeric khipu satellite repeat from *Phaseolus vulgaris*: lessons learned from the genome analysis of the andean genotype G19833. Front Plant Sci 4:109. doi:10.3389/fpls.2013.00109

Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative CT method. Nature Protoc 3(6):1101–1108. doi:10.1038/nprot.2008.73

Schmutz J, McClean PE, Mamidi S, et al. (2014) A reference genome for common bean and genome-wide analysis of dual domestications. Nature Genet 46:707–713. doi:10.1038/ng.3008

Sessa G, Martin GB (2000) Signal recognition and transduction mediated by the tomato Pto kinase: a paradigm of innate immunity in plants. Microb Infec 2(13):1591-1597.

Shiu SH, Bleecker AB (2001) Plant receptor-like kinase gene family: diversity, function, and signaling. Sci STKE (113):re22. doi:10.1126/stke.2001.113.re22

Silverio L, Vidigal MC, Vidigal Filho PS, Barelli MAA, Thomazella C, Nunes WMC (2002) Genetic resistance to *Colletotrichum lindemuthianum* race 2047 in G2333. Ann Rep Bean Improv Coop 45:74–75.

Singh SP, Schwartz HF (2010) Breeding common bean for resistance to diseases: a review. Crop Science 50(6):2199. doi:10.2135/cropsci2009.03.0163

Song WY, Wang GL, Chen L, Kim HS, Pi LY, Gardner J, Wang B, Holsten T, Zhai WX, Zhu LH, Fauquet C, Ronald PC (1995) A receptor kinase–like protein encoded by the rice disease resistance gene Xa21. Science 270:1804–1806. doi: 10.1126/science.270.5243.1804

Spoel SH, Dong X (2012) How do plants achieve immunity? Defence without specialized immune cells. Nature Rev Immunol 12(2):89–100. doi:10.1038/nri3141

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S (2011) MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 28:2731–2739. doi: 10.1093/molbev/msr121

Tatusova TA, Madden TL (1999) BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences. FEMS Microbiol Letters 174: 247–250. doi: 10.1111/j.1574-6968.1999.tb13575.x

Temnykh S, DeClerck G, Lukashova A, Lipovich L, Cartinhour S, McCouch S (2001) Computational and experimental analysis of microsatellites in rice (*Oryza sativa* L.): frequency, length variation, transposon associations, and genetic marker potential. Genome Res 11:1441-1452. doi:10.1101/gr.184001. PMID:11483586

Thomma BPHJ, Nürnberger T, Joosten MHAJ (2011) Of PAMPs and effectors: the blurred PTI-ETI dichotomy. Plant Cell 23(1):4–15. doi:10.1105/tpc.110.082602

Vanhouten W, MacKenzie S (1999) Construction and characterization of a common bean bacterial artificial chromosome library. Plant Mol Biol 40(6):977–83. doi:10.1023/A:1006234823105

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78. doi:10.1093/jhered/93.1.77

Young RA, Kelly JD (1996) Characterization of genetic resistance to *Colletotrichum lindemuthianum* in common bean differential cultivars. Plant Disease 80(6):650-654. doi:10.1590/S1516-89132008000500002

Young RA, Melotto M, Nodari RO, Kelly JD (1998) Marker assisted dissection of the oligogenic anthracnose resistance in common bean cultivar G2333. Theor Appl Genet 96:87–94. doi:10.1007/s001220050713

Wolf S, Hématy K, Höfte H (2012) Growth control and cell wall signaling in plants. Ann Rev Plant Biol 63:381–407. doi:10.1146/annurev-arplant-042811-105449

Zipfel C, Robatzek S, Navarro L, Oakeley EJ, Jones JD, Felix G, Boller T (2004) Bacterial disease resistance in Arabidopsis through flagellin perception. Nature 428:764–767. doi:10.1038/nature02485

Zmasek CM, Godzik A (2011) Strong functional patterns in the evolution of eukaryotic genomes revealed by the reconstruction of ancestral protein domain repertoires. Genome Biol 12(1):R4. doi:10.1186/gb-2011-12-1-r4.

Figure 1
Click here to download Figure: Figure 1_COK-4 manuscrip_Final.pptx

Fig. 1

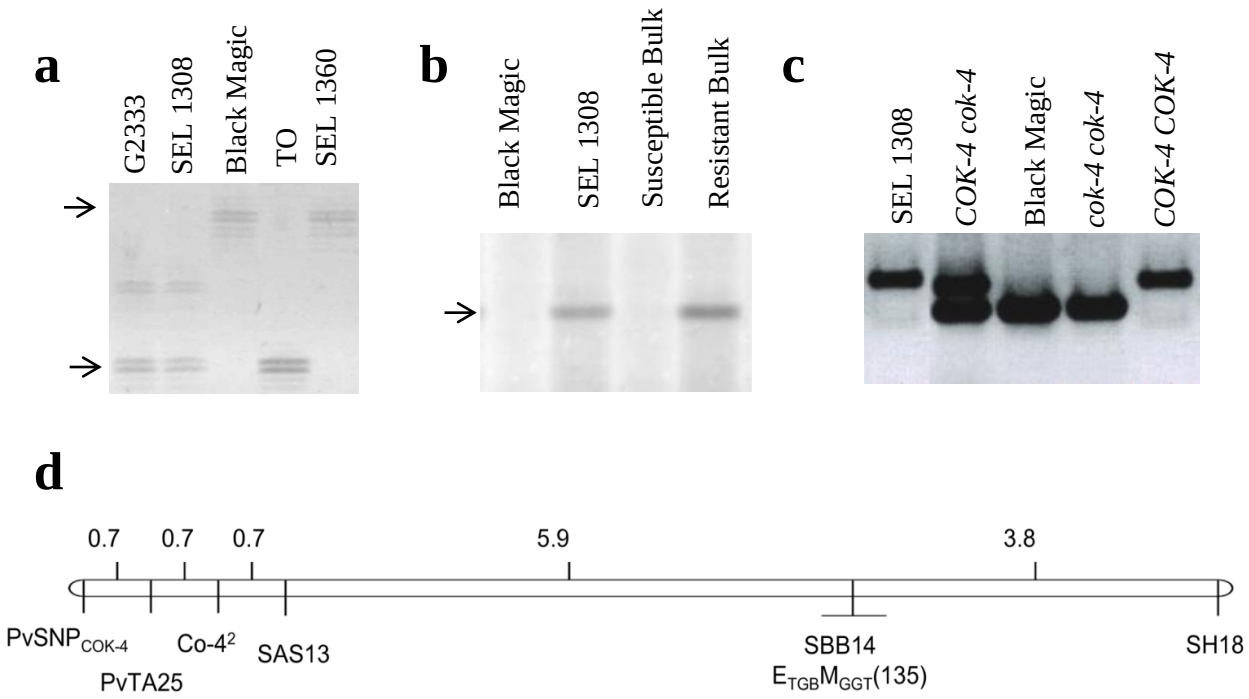


Fig. 1 Molecular marker banding pattern in resistant and susceptible genotypes of common bean. G2333, SEL 1308, and TO are homozygous anthracnose resistant genotypes, Black Magic and SEL 1360 are homozygous anthracnose susceptible genotypes. **a** and **b** Silver-stained polyacrylamide gel showing polymorphisms (scored bands indicated by the arrows) detected with the PvTA25 SSR primers **(a)** and the $E_{TGC}M_{GGT}$(135) AFLP **(b)** markers. **c** Ethidium bromide-stained agarose gel showing PCR-amplified DNA fragments with the $PvSNP_{COK-4}$ marker (700 bp lower band and 1000 bp upper band); legends on top of the lanes indicate the genotype of F_2 individuals from the cross between SEL1308 and Black Magic. **d** Predicted genetic distance among the markers linked to the *Co-4* locus. The numbers on the top are estimated distances between molecular markers in centiMorgans (cM) calculated with the MapMaker software considering LOD score > 3 as threshold. The linkage map diagram was created with the MapChart 2.2 program (Voorrips 2002) and the scale was set as 10 mm/cM.

Figure 2

Click here to download Figure: Figure 2_COK-4 manuscrip_Final.pptx

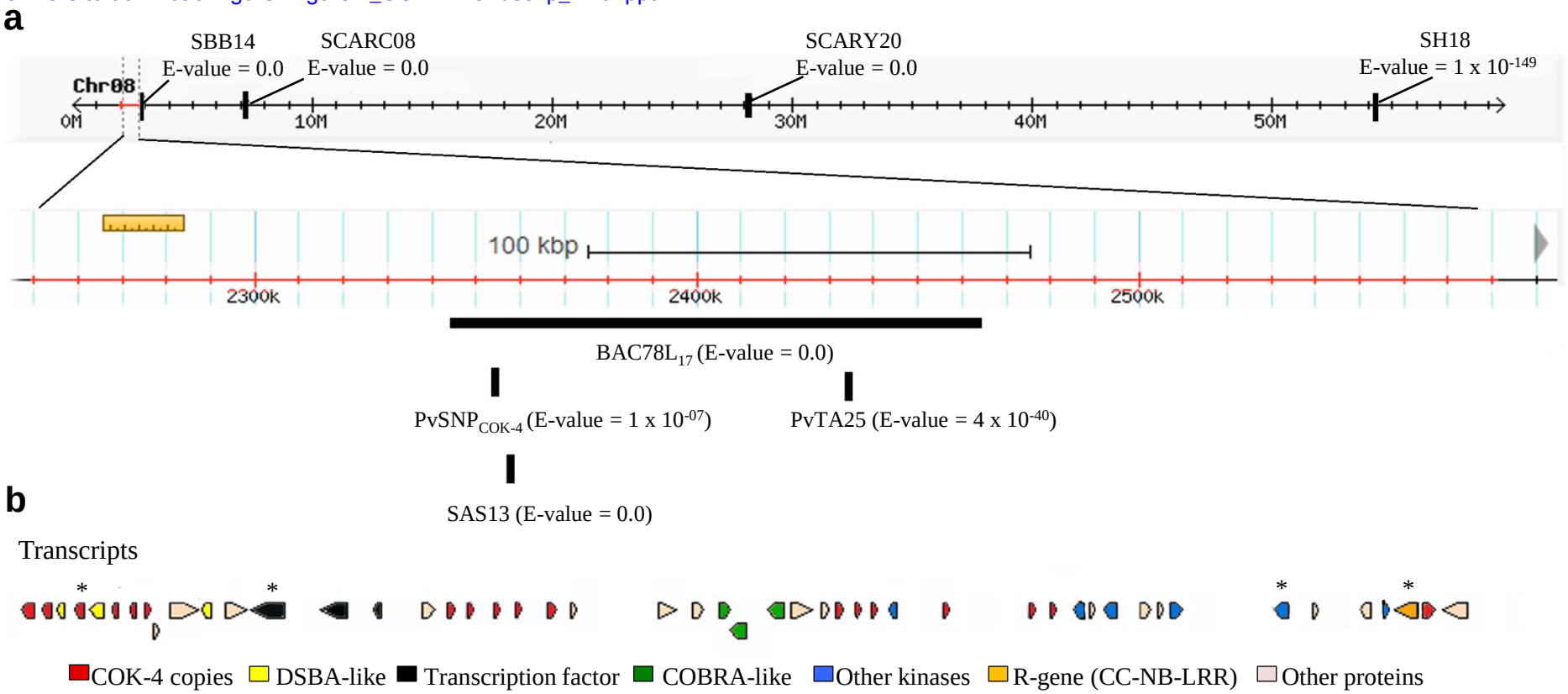


Fig. 2 Genomic boundaries and structure of the bean anthracnose resistance locus *Co-4*. **a** Modified genome browser representation of the 325 Kbp in the chromosome 8 (Chr08) spanning the *Co-4* locus that contains the markers SAS13, PvTA25, $PvSNP_{COK-4}$, as well as the BAC $78L_{17}$ sequences. Physical location was determined by BLASTN analysis (threshold E-value $\leq 1 \times 10^{-5}$) of the marker sequences against the common bean genome v.1.0 (Schmutz et al. 2014; http://www.phytozome.net/). **b** Predicted transcripts in *Co-4* region are shown below its genomic positions in (**a**). Color codes, as indicated in the legend below the figure, represent predicted gene functions. Asterisks above the transcript indicate genes that were analyzed by RT-qPCR.

Figure 3
Click here to download Figure: Figure 3_COK-4 manuscrip_Final.pptx

Fig. 3

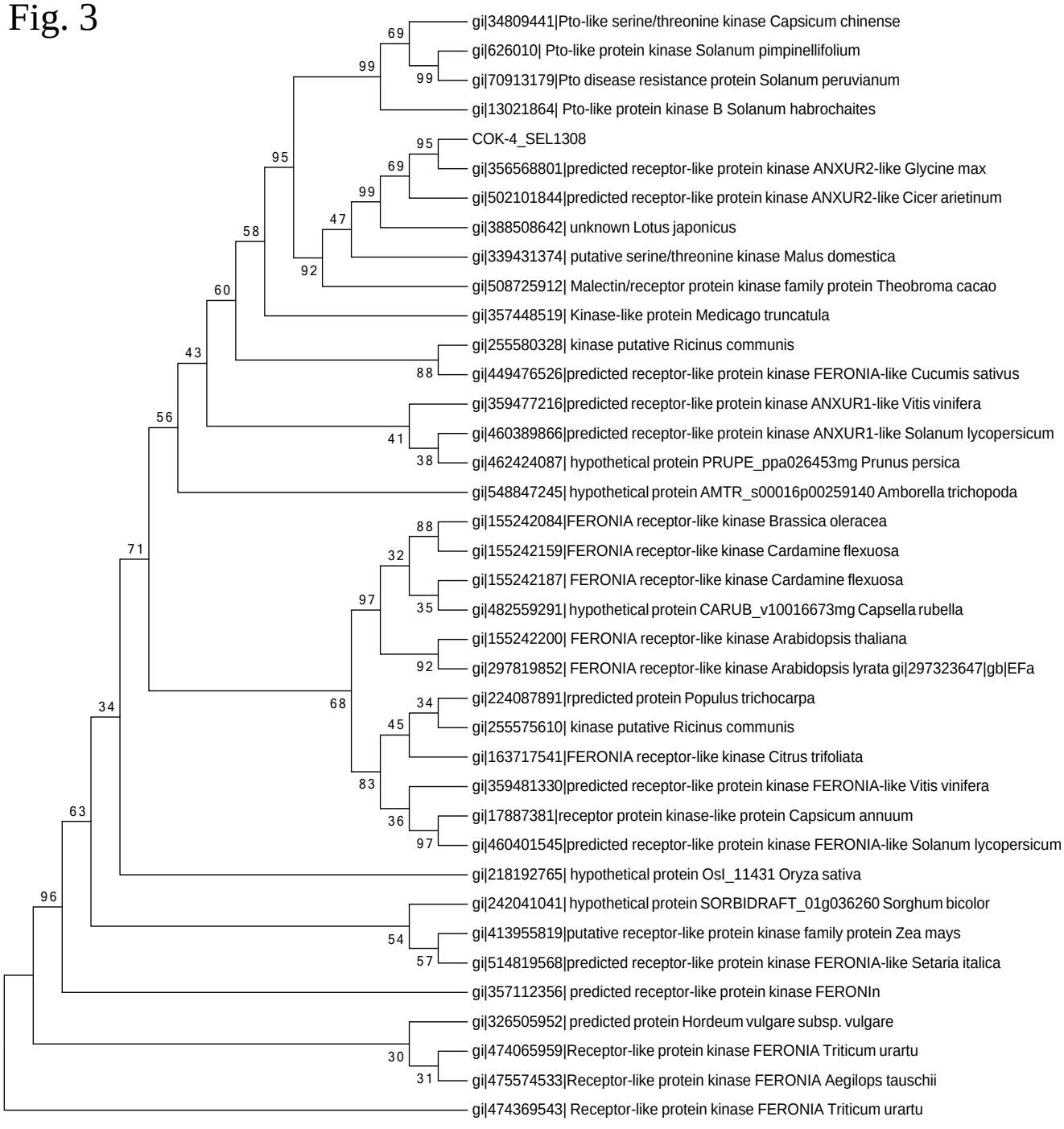


Fig. 3 COK-4_SEL 1308 predicted protein clustered with members of the *Catharanthus roseus* RLK1 (CrRLK1) protein family (mainly FERONIA-like and ANXUR-like) from diverse plant species. Phylogenetic analysis was performed with the maximum parsimony method using the MEGA 5.05 software (Tamura et al. 2011). Bootstrap support values are adjacent to the tree nodes.

Figure 4

Click here to download Figure: Figure 4_COK-4 manuscrip_Final.pptx

Fig. 4

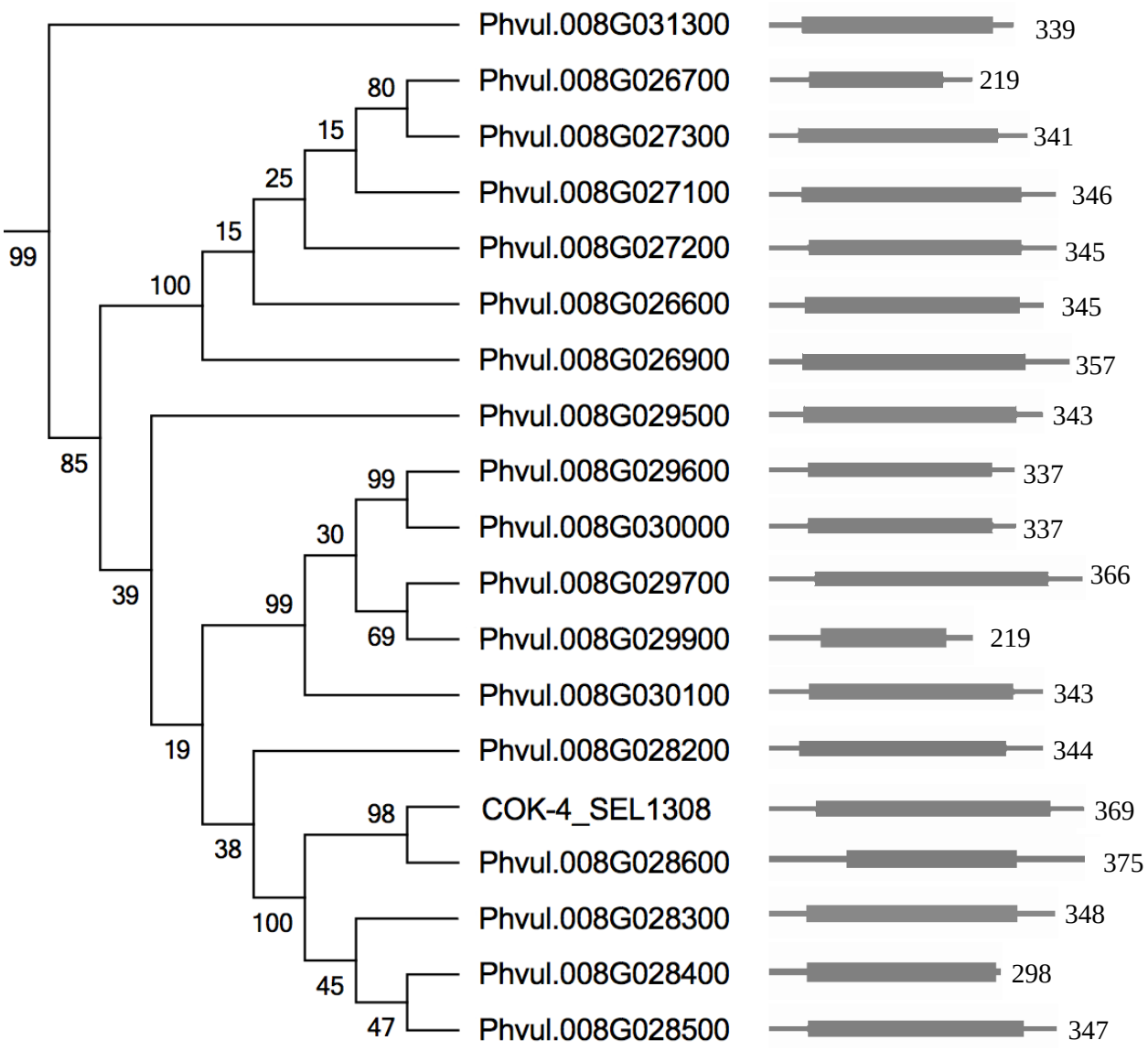


Fig. 4 COK-4_SEL 1308 predicted protein clustered with common bean kinases located in the *Co-4* genomic region of Pv08. Phylogenetic analysis of predicted amino acid sequence was performed with the maximum parsimony method using the MEGA 5.05 software (Tamura et al. 2011). Bootstrap support values are provided adjacent to nodes. The diagram shown in front of the transcript name represents the single kinase domain (grey rectangles) within the protein. The numbers indicate the total amino acids of each protein. Diagram was adapted from the protein domain view of Phytozome. Only the clade containing the COK-4 is shown (refer to Fig. S1 for the entire tree with the top 100 kinases most similar to COK-4_SEL1308 in the G19833 reference genome).

Figure 5
Click here to download Figure: Figure 5_COK-4 manuscrip_Final.pptx

Fig. 5

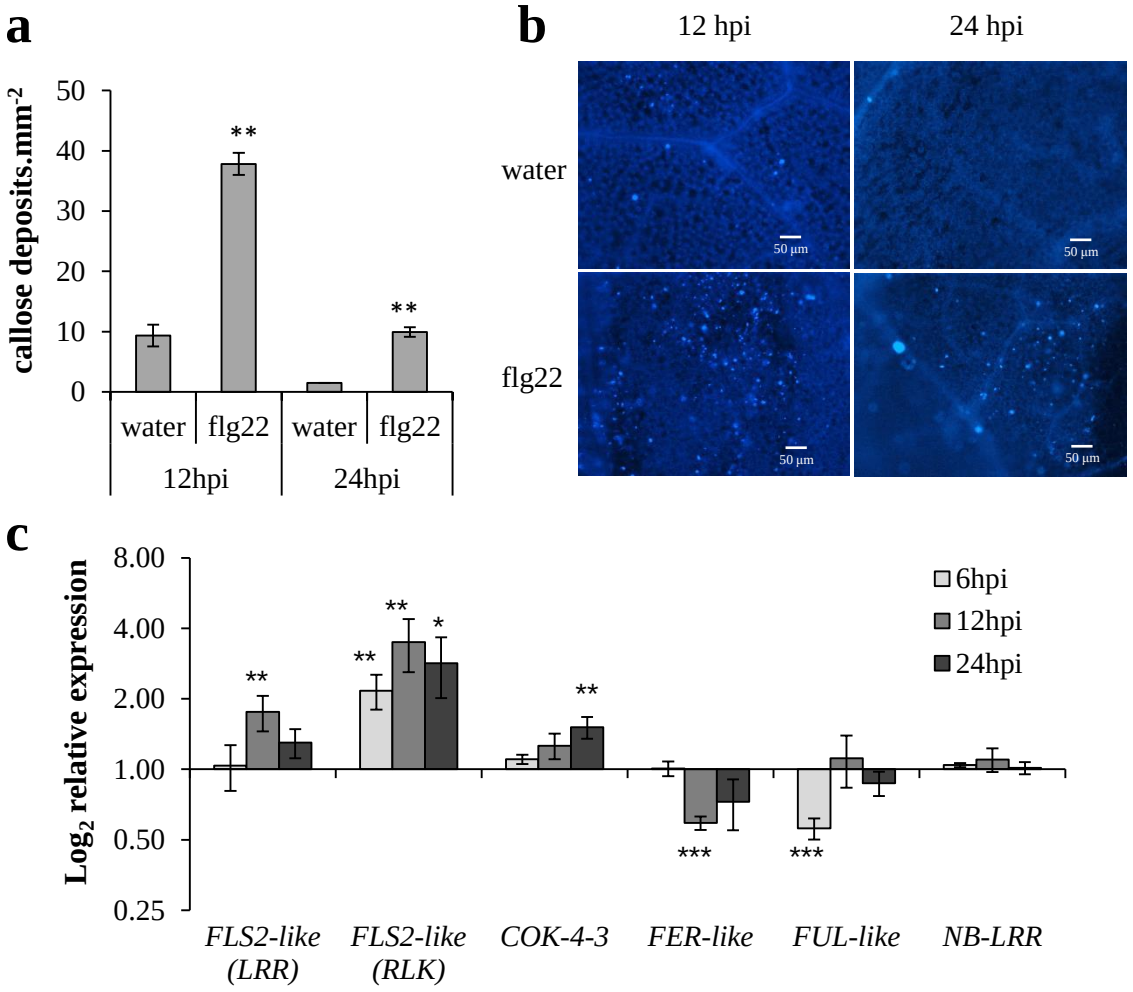


Fig. 5 G2333 responses to the PAMP flg22. **a** Graph shows the average number of callose deposits per mm^{-2} of G2333 leaf tissue infiltrated with 1 µM flg22 or water. Results are shown as average of 108 to 135 images in three independent biological replicates ± standard error. **b** Representative images (100 x magnification) of aniline blue stained G2333 leaves 12 h or 24 h post incubation (hpi) with flg22 or water. **c** Expression of the indicated genes (x-axis) in G2333 leaves immersed in 5 µM flg22 at 6, 12, and 24 hpi relative to the their expression in water-immersed leaves (control) considered as 1. Data points are average of at least two biological replicates ($n \geq 6$ ± standard error). Asterisk above the bars of all graphs indicate statistical significance calculated with Student's *t* test (**p<0.01, ***p<0.001).

Figure 6

Click here to download Figure: Figure 6_COK-4 manuscrip_Final.pptx

Fig. 6

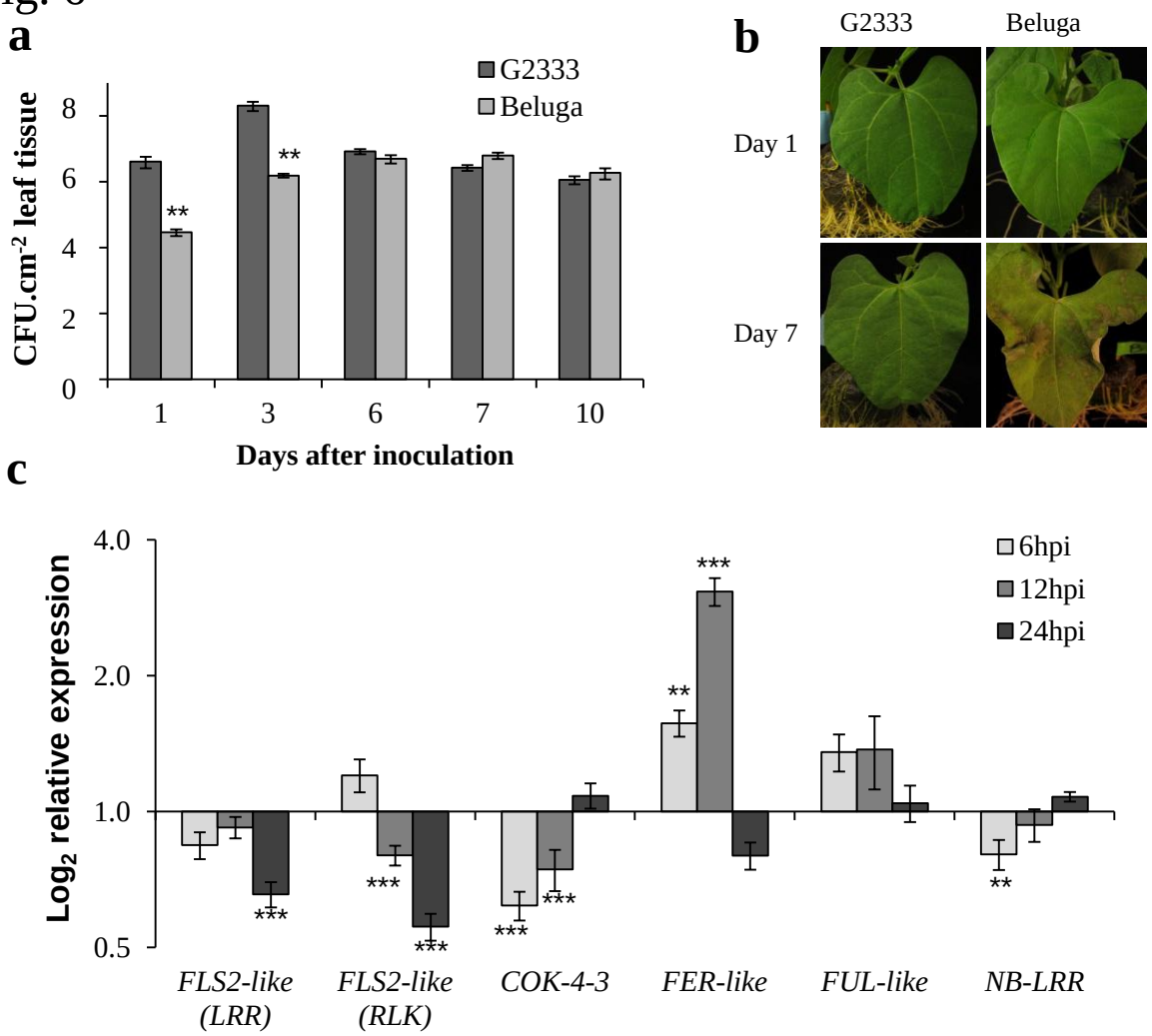


Fig. 6 G2333 responses to *Pseudomonas syringe* pv. *phaseolicola* (Pph). **a** G2333 showed tolerance to Pph (NPS3121), with no bacterial growth in the leaf apoplast of fully expended primary leaves dipped inoculated with 10^8 CFU/ml. **b** Halo blight symptoms were observed after 7 days of inoculation only for Beluga genotype. **c** Expression of the indicated genes (x-axis) in G2333 leaves dipped in 10^8 CFU/ml of Pph relative to the their expression in mock-inoculated leaves (control) considered as 1. Data points are average of at least two biological replicates ($n \geq 6 \pm$ standard error). Asterisk above the bars of all graphs indicate statistical significance calculated with Student's *t* test (**p<0.01, ***p<0.001).

Tables

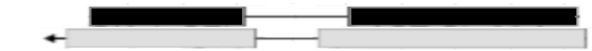
Click here to download Table: Tables_COK-4 manuscript_Final.docx

Table 1 Segregation analysis of molecular markers linked to the *Co-4* gene using an F_2 mapping population derived from the SEL 1308 x Black Magic genetic cross.

Locus	Expected segregation ratio	Observed frequency	*P*[a]	Marker type
PvTA25	1:2:1	19:45:24	0.74	SSR
$E_{TGC}M_{GGT}$(135)	3:1	69:28	0.38	AFLP
SAS13	3:1	70:27	0.52	SCAR
SBB14	1:2:1	17:52:28	0.22	SCAR
SH18	3:1	65:31	0.10	SCAR
$PvSNP_{COK\text{-}4}$	1:2:1	200:426:223	0.53	SNP
Co-4^2 / *co-4*^{2}	3:1	626:223	0.39	-

[a] *P* = statistical probability calculated with the chi-square test.

Table 2 Common bean genome regions similar to the SEL 1308 *COK-4* gene.

Common bean gene	*COK-4* copy[a]	E-value[b]	Identities	Gaps	Alignment scheme[c]
Phvul.008G026600	*COK-4-1*	1 x 10^{-149}	776/1098(71%)	75/1098(6%)	
Phvul.008G026700	*COK-4-2*	4 x 10^{-108}	433/581(75%)	11/581(1%)	
Phvul.008G026900	*COK-4-3*	7 x 10^{-104}	671/890(75%)	28/890(3%)	
Phvul.008G027100	*COK-4-4*	1 x 10^{-98}	666/887(75%)	25/887(3%)	
Phvul.008G027200	*COK-4-5*	1 x 10^{-142}	773/1097(70%)	70/1097(6%)	
Phvul.008G027300	*COK-4-6*	6 x 10^{-140}	757/1067(71%)	45/1067(4%)	
Phvul.008G028200	*COK-4-7*	7 x 10^{-128}	793/1049(76%)	12/1049(1%)	
Phvul.008G028300	*COK-4-8*	0.0	1038/1102(94%)	5/1102(0%)	
Phvul.008G028400	*COK-4-9*	0.0	1034/1100(94%)	14/1100(1%)	
Phvul.008G028500	*COK-4-10*	0.0	1049/1100(95%)	3/1100(0%)	
Phvul.008G028600	*COK-4-11*	0.0	1093/1111(98%)	1/1111(0%)	
Phvul.008G029500	*COK-4-12*	0.0	781/1061(74%)	38/1061(3%)	

Phvul.008G029600	*COK-4-13*	6 x 10^{-172}	754/1031(73%)	50/1031(4%)	
Phvul.008G029700	*COK-4-14*	2 x 10^{-165}	751/1035(73%)	63/1035(6%)	
Phvul.008G029900	*COK-4-15*	6 x 10^{-93}	403/550(73%)	25/550(4%)	
Phvul.008G030000	*COK-4-16*	3 x 10^{-169}	752/1031(73%)	50/1031(4%)	
Phvul.008G030100	*COK-4-17*	3 x 10^{-175}	756/1035(73%)	59/1035(5%)	
Phvul.008G031300	*COK-4-18*	5 x 10^{-64}	547/758(72%)	36/758(5%)	

Common bean region	E-value[b]	Identities	Gaps
Chr08:2438870..2439364	4 x 10^{-68}	358/495(72%)	12/495(2%)
Chr08:2450588..2451617	4 x 10^{-157}	748/1030(73%)	40/1030(4%)
Chr05:20356383..20356071	7 x 10^{-47}	231/313(74%)	1/313(0%)

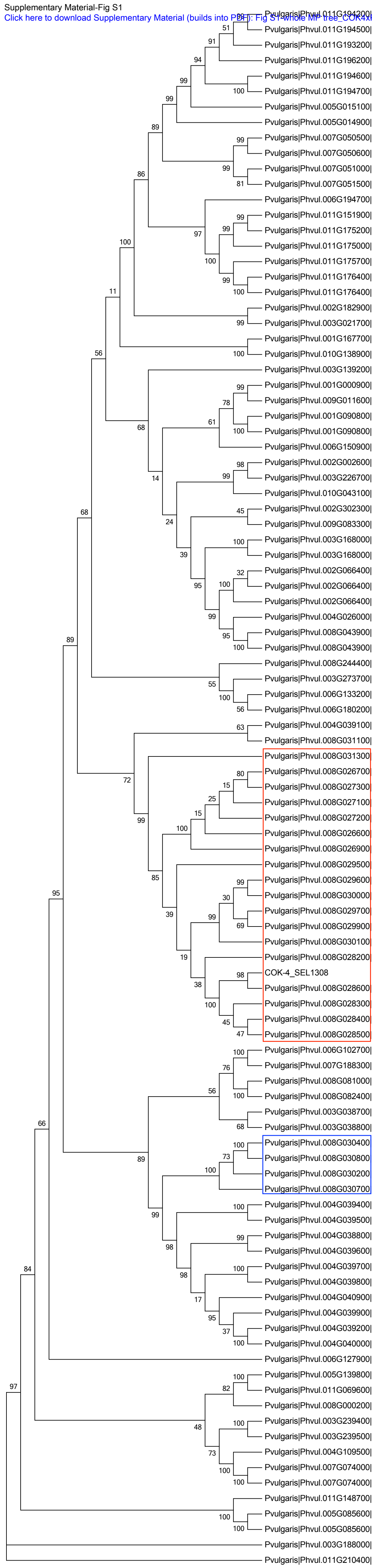
[a] Putative *COK-4* copies were named according to the gene order on Pv08, without implying functionality.

[b] BLASTN analysis was conducted using the *COK-4*_SEL1308 genomic sequence as query (1110 bp) against the common bean reference genome (Phytozome v1.0; http://phytozome.jgi.doe.gov/) to identify the genome location of the putative *COK-4* paralogs. After the regions with similarity to *COK-4* were identified, they were individually aligned with *COK-4*_SEL1308 (BLASTN pair-wise alignment bl2seq; http://www.ncbi.nlm.nih.gov/) to generate the diagrams in the table.

[c] Schematic representation of genes, adapted from the Phytozome genome view, showing 5'and 3'UTR regions (dark greys rectangles), introns (dark grey lines), and the predict coding regions (light grey rectangles). Black rectangles above the each gene diagram represent the *COK-4*_SEL1308 sequence that aligned with each gene.

Supplementary Material-Fig S1

Click here to download Supplementary Material (builds into PDF): Fig S1-whole MP tree_COK4x



Supplementary Material-Tables
Click here to download Supplementary Material (builds into PDF): Supplemental tables_COK-4 manuscript_Final.docx

Table S1 Simple sequence repeats (SSR) found in the BAC $78L_{17}$ insert sequence using the SSRIT software (http://www.gramene.org/gramene/searches/ssrtool).

SSR name	Motif	No. of repeats	SSR start position	SSR end position
PvTA16	TA	16	1677	1708
PvAT9	AT	9	3311	3328
PvAT5	AT	5	13227	13236
PvTA6	TA	6	14106	14117
PvTA25	TA	25	14315	14364
PvTG6	TG	6	14427	14438
PvCA5	CA	5	23552	23561
PvTA6-2	TA	6	27557	27568
PvTA5	TA	5	34893	34902
PvAT10	AT	10	35342	35361
PvAT8	AT	8	36177	36192
PvCT5	CT	5	38127	38136
PvCT8	CT	8	38149	38164
PvAT5-2	AT	5	52311	52320
PvAT6	AT	6	63691	63702
PvTA5-2	TA	5	64014	64023
PvAT5-3	AT	5	89859	89868
PvTA6-3	TA	6	98135	98146
PvAGA6	AGA	6	26742	26759
PvTAA6	TAA	6	62024	62041
PvAAT5	AAT	5	68888	68902

Table S2 Primer sequences for the newly developed SSR, AFLP, and $PvSNP_{COK-4}$ markers.

SSR	**Direction**	**Sequence (5'- 3')**
PvTA25	forward	CTTTATTATAGAGAATAAGACTCACC
	reverse	CGACAGAGAACTACCTAATCATTTGC
AFLP Adapter	**Primer**	**Sequence (5'- 3')**
*Eco*RI	1	CTCGTAGACTGCGTACC
	2	ATTTGGTACGCAGTCTAC
*Mse*I	1	GACGATGAGTCCTGAG
	2	TACTCAGGACTCAT
AFLP Pre-selective primers		**Sequence (5'- 3')**
*Eco*RI + T		GACTGCGTACCAATTCT
*Mse*I + G		GATGAGTCCTGAGTAAG
$PvSNP_{COK-4}$	**Direction**	**Sequence (5'- 3')**
SEL1308 and Black Magic[a]	forward	GTGTGGGCATGTGTTGTTCGAAGCCC
SEL1308	reverse	TCTCATCTGGTTCATACTTCAAGCAAC
Black Magic	reverse	GTCCGTAGCCGGGTAGCCAAAAGT

[a] Primer designed based on the consensus sequence of both genotypes.

Table S3 Primer sequences designed to assess gene expression analysis by RT-qPCR.

Gene code	Annotation	Direction	Sequence (5'- 3')
Phvul.001G133200	*PvIDE*	forward	GAGAGACTATGAGGTTGAAGC
		reverse	CCATGAACTCGTACACTTAAAG
Phvul.005G149200	*FLS2-like (LRR)*	forward	GCCTCACGGTGCTGAACAT
		reverse	CGGAGAGGAGGTTGTTACGGA
Phvul.002G196200	*FLS2-like (RLK)*	forward	CTCCAGAATTTGCCTACACGAG
		reverse	GAGTCCTGTCGGCCTTCTTT
Phvul.008G026900	*COK-4-3*	forward	CACGTTTTTCGCTCTACTCAC
		reverse	TGCCGCCACAACTTTCAGTA
Phvul.008G030800	*FER-like*	forward	GGCCAGATAGCAGCTCATTG
		reverse	CCCAACCACGTCGTTCATAG
Phvul.008G031200	*NB-LRR*	forward	GATGACCCAGAATCCACGACTTC
		reverse	GCCTCCTAGAACGATGATAGCC
Phvul.008G027800	*FUL-like*	forward	CTGAGGGCAATTGGTCTTTCGA
		reverse	GGATCAAGCTCATTTCCCAAGA

Table S4 Predicted transcripts of the *Co-4* genomic region in chromosome 8 (325kb; Chr08:2,245,000..2,570,000) based on RNA-seq data for the bean genome (Phytozome). Transcripts in bold letters code for predicted protein kinases. Transcripts in bold underlined letters were identified as copies of *COK-4*. Annotation is based on the Phytozome database (http://www.phytozome.net/).

***P. vulgaris* transcripts**	**Annotation**
Phvul.008G026600.1	Protein tyrosine kinase
Phvul.008G026700.1	Protein tyrosine kinase
Phvul.008G026800.1	DSBA-like thioredoxin domain
Phvul.008G026900.1	Protein tyrosine kinase
Phvul.008G027000.1	DSBA-like thioredoxin domain
Phvul.008G027100.1	Protein tyrosine kinase
Phvul.008G027200.1	Protein tyrosine kinase
Phvul.008G027300.1	Protein tyrosine kinase
Phvul.008G027400.1	No functional annotation
Phvul.008G027500.1	Zinc finger, C3HC4 type (RING finger)
Phvul.008G027600.1	DSBA-like thioredoxin domain
Phvul.008G027700.1	Regulator of chromosome condensation (RCC1) repeat
Phvul.008G027800.1	SRF-type transcription factor (DNA-binding and dimerization domain)
Phvul.008G027900.1	SRF-type transcription factor (DNA-binding and dimerization domain)
Phvul.008G028000.1	Myb-like DNA-binding domain
Phvul.008G028100.1	2OG-Fe(II) oxygenase superfamily
Phvul.008G028200.1	Protein tyrosine kinase
Phvul.008G028300.1	Protein tyrosine kinase
Phvul.008G028400.1	Protein tyrosine kinase
Phvul.008G028500.1	Protein tyrosine kinase
Phvul.008G028600.1	Protein tyrosine kinase
Phvul.008G028700.1	B12D protein
Phvul.008G028800.1	Cytochrome P450
Phvul.008G028900.1	Reversibly glycosylated polypeptide
Phvul.008G029000.1	COBRA-like protein
Phvul.008G029100.1	COBRA-like protein
Phvul.008G029200.1	COBRA-like protein
Phvul.008G029300.1	Mitochondrial carrier protein

Phvul.008G029400.1	Transferase family
Phvul.008G029500.1	Protein tyrosine kinase
Phvul.008G029600.1	Protein tyrosine kinase
Phvul.008G029700.1	Protein tyrosine kinase
Phvul.008G029800.1	Protein tyrosine kinase
Phvul.008G029900.1	Protein tyrosine kinase
Phvul.008G030000.1	Protein tyrosine kinase
Phvul.008G030100.1	Protein tyrosine kinase
Phvul.008G030200.1	Protein tyrosine kinase
Phvul.008G030300.1	No functional annotation
Phvul.008G030400.1	Protein tyrosine kinase
Phvul.008G030500.1	Magnesium transporters: CorA family
Phvul.008G030600.1	Core-2/I-Branching enzyme
Phvul.008G030700.1	Protein tyrosine kinase
Phvul.008G030800.1	Protein tyrosine kinase
Phvul.008G030900.1	No functional annotation
Phvul.008G031000.1	No functional annotation
Phvul.008G031100.1	Protein tyrosine kinase
Phvul.008G031200.1	NB-LRR domain
Phvul.008G031300.1	Protein tyrosine kinase
Phvul.008G031400.1	Alpha/beta hydrolase fold

Table S5 Predicted common bean (Pv) proteins in the 325 Kbp surrounding the *Co-4* genomic region (Pv08, 2,245,000..2,570,000) with significant similarity (BLASTP) to the predicted COK-4_SEL 1308 protein and their putative Arabidopsis (At) orthologs.

COK-4 vs. bean proteome[b]		**COK-4 vs. TAIR10[c]**		
Pv gene code[a]	E-value	At ortholog	E-value	At Annotation
Phvul.008G026600.1	4×10^{-124}	AT5G28680	5×10^{-52}	ANXUR2
Phvul.008G026700.1	9×10^{-86}	AT5G39000	5×10^{-38}	Malectin/receptor-like protein kinase
Phvul.008G026900.1	2×10^{-130}	AT3G51550	3×10^{-49}	FERONIA
Phvul.008G027100.1	9×10^{-125}	AT3G51550	7×10^{-55}	FERONIA
Phvul.008G027200.1	2×10^{-112}	AT3G51550	4×10^{-53}	FERONIA
Phvul.008G027300.1	4×10^{-117}	AT3G51550	5 x 10-56	FERONIA
Phvul.008G028200.1	1×10^{-153}	AT5G28680	9×10^{-63}	ANXUR2
Phvul.008G028300.1	0.0	AT3G51550	1×10^{-57}	FERONIA
Phvul.008G028400.1	0.0	AT3G51550	7×10^{-45}	FERONIA
Phvul.008G028500.1	0.0	AT5G28680	2×10^{-55}	ANXUR2
Phvul.008G028600.1	0.0	AT3G51550	2×10^{-52}	FERONIA
Phvul.008G029500.1	6×10^{-136}	AT5G28680	8×10^{-60}	ANXUR2
Phvul.008G029600.1	1×10^{-129}	AT3G51550	6×10^{-62}	FERONIA
Phvul.008G029700.1	1×10^{-138}	AT5G28680	1×10^{-61}	ANXUR2
Phvul.008G029800.1	2×10^{-31}	AT3G51550	6×10^{-82}	FERONIA
Phvul.008G029900.1	3×10^{-84}	AT3G51550	3×10^{-39}	FERONIA
Phvul.008G030000.1	2×10^{-130}	AT3G51550	7×10^{-62}	FERONIA
Phvul.008G030100.1	7×10^{-134}	AT5G28680	4×10^{-61}	ANXUR2
Phvul.008G030200.1	1×10^{-58}	AT5G39000	1×10^{-172}	Malectin/receptor-like protein kinase
Phvul.008G030400.1	8×10^{-55}	AT3G51550	0.0	FERONIA
Phvul.008G030700.1	9×10^{-60}	AT3G51550	0.0	FERONIA
Phvul.008G030800.1	3×10^{-54}	AT3G51550	0.0	FERONIA
Phvul.008G031100.1	7×10^{-70}	AT3G51550	7×10^{-63}	FERONIA

Phvul.008G031300.1	5×10^{-113}	AT5G28680	1×10^{-63}	ANXUR2

[a] Underlined gene codes are predicted to code for proteins containing both kinase and malectin.

[b] *Phaseolus vulgaris* Phytozome proteome database (threshold E value $\leq 1 \times 10^{-20}$ and identity > 35%).

[c] *Arabidopsis thaliana* TAIR10 protein database (threshold E value $\leq 1 \times 10^{-20}$ and identity > 35%).

Table S6 Putative COK-4 protein orthologs in 35 different plant species. The SEL 1308 COK-4 predicted protein was used as query for BLASTP analysis using the non-redundant (nr) protein database from NCBI (threshold E-values ≤ 1 x 10^{-20} and identity > 30%). Protein domain superfamily was inferred based on the NCBI conserved domain database (http://www.ncbi.nlm.nih.gov/cdd).

Species	GI number	E-value	Identity	Protein domain superfamily	Annotation
Glycine max	gi|356568801|	7 x 10^{-160}	65%	Kinase	ANXUR2 receptor-like kinase
Cicer arietinum	gi|502101844|	6 x 10^{-96}	47%	Kinase	ANXUR2 receptor-like kinase
Malus domestica	gi|339431374|	4 x 10^{-69}	39%	Kinase	Putative serine/threonine kinase
Theobroma cacao	gi|508725912|	4 x10^{-65}	40%	Kinase	Malectin/receptor protein kinase family
Hordeum vulgare	gi|326505952|	6 x 10^{-62}	37%	Kinase	Predicted protein
Vitis vinifera	gi|359477216|	4 x 10^{-61}	36%	Malectin and Kinase	ANXUR1 receptor-like kinase
Cucumis sativus	gi|449476526|	2 x 10^{-60}	40%	Malectin and Kinase	FERONIA receptor-like kinase
Citrus trifoliata	gi|163717541|	2 x 10^{-60}	385	Kinase	FERONIA receptor-like kinase
Aegilops tauschii	gi|475574533|	2 x 10^{-59}	38%	Malectin and Kinase	FERONIA receptor-like kinase
Triticum urartu	gi|474369543|	5 x 10^{-59}	37%	Malectin and Kinase	FERONIA receptor-like kinase
Oryza sativa	gi|218192765|	6 x 10^{-59}	37%	Malectin and Kinase	Hypothetical protein OsI_11431
Zea mays	gi|413955819|	9 x 10^{-59}	37%	Malectin and Kinase	Receptor protein kinase-like
Solanum lycopersicum	gi|460389866|	2 x 10^{-58}	39%	Malectin and Kinase	ANXUR1 receptor-like kinase
Setaria italica	gi|514819568|	2 x 10^{-58}	37%	Malectin and Kinase	FERONIA receptor-like kinase
Capsicum chinense	gi|34809441|	2 x 10^{-58}	36%	Kinase	Pto-like serine/threonine kinase
Triticum urartu	gi|474065959|	3 x 10^{-58}	37%	Malectin and Kinase	FERONIA receptor-like kinase
Ricinus communis	gi|255580328|	4 x 10^{-58}	38%	Malectin and Kinase	Putative kinase

Cardamine flexuosa	gi|155242159|	7×10^{-58}	37%	Malectin and Kinase	FERONIA receptor-like kinase
Brachypodium distachyon	gi|357112356|	1×10^{-57}	38%	Malectin and Kinase	FERONIA receptor-like kinase
Sorghum bicolor	gi|242041041|	1×10^{-57}	37%	Malectin and Kinase	Hypothetical protein SORBIDRAFT_01g036260
Arabidopsis thaliana	gi|155242200|	2×10^{-57}	38%	Malectin and Kinase	FERONIA receptor-like kinase
Populus trichocarpa	gi|224087891|	2×10^{-57}	38%	Malectin and Kinase	Predicted protein
Arabidopsis lyrata subsp. lyrata	gi|297819852|	3×10^{-57}	38%	Malectin and Kinase	Hypothetical protein ARALYDRAFT_485507
Vitis vinifera	gi|359481330|	3×10^{-57}	38%	Kinase	FERONIA receptor-like kinase
Lotus japonicus	gi|388508642|	4×10^{-57}	56%	Kinase	Unknown
Capsella rubella	gi|482559291|	5×10^{-57}	38%	Malectin and Kinase	Hypothetical protein CARUB_v10016673mg
Ricinus communis	gi|255575610|	7×10^{-57}	38%	Malectin and Kinase	Putative kinase
Cardamine flexuosa	gi|155242187|	8×10^{-57}	38%	Malectin and Kinase	FERONIA receptor-like kinase
Prunus persica	gi|462424087|	9×10^{-57}	37%	Malectin and Kinase	Hypothetical protein PRUPE_ppa026453mg
Brassica oleracea	gi|155242084|	1×10^{-56}	38%	Malectin and Kinase	FERONIA receptor-like kinase
Capsicum annuum	gi|17887381|	3×10^{-56}	37%	Malectin and Kinase	Receptor protein kinase-like
Amborella trichopoda	gi|548847245|	8×10^{-55}	36%	Malectin and Kinase	Hypothetical protein AMTR_s00016p00259140
Solanum peruvianum	gi|70913179|	3×10^{-54}	37%	Kinase	Pto disease resistance protein
Solanum habrochaites	gi|13021864|	6×10^{-54}	37%	Kinase	Pto-like protein kinase B
Medicago truncatula	gi|357448519|	2×10^{-53}	38%	Malectin and Kinase	Receptor protein kinase-like