Applying next-generation sequencing to enable marker-assisted breeding for adaptive traits in common bean (*Phaseolus vulgaris* L.)

by

Andrew James Tock

A thesis submitted to the University of Warwick
in partial fulfilment of the requirements for the degree of
Doctor of Philosophy in Life Sciences

School of Life Sciences
University of Warwick
April 2017

Table of Contents

List of Tables

List of Figures

Acknowledgements

This thesis is the culmination of the most rewarding four years of my adult life. I am indebted to Professor Eric Holub for his peerlessly generous mentoring and enthusiastic supervision, which have been instrumental in my doctoral training. I am proud of the scientist I am becoming and this is due in large part to him. I am also extremely grateful to Dr Guy Barker for his expert supervision and advice, which have given me the bioinformatics bug. His guidance on developing and implementing a research strategy is something that I will continue to try to emulate.

It has been a humbling honour to work with researchers and support staff based at Warwick Crop Centre. I am thankful to Drs Joana Vicente and John Taylor, who have provided me with invaluable training in, and insights into, plant pathology, microbiology and molecular genetics. Professor John Walsh helped me to appreciate the merit of being a "doubting Thomas" early in my training, from which this work benefits. I feel privileged to have benefited from the exceptionally generous training in computational biology and statistical genetics offered by Dr Peter Walley, Siva Samavedam and Dr Laura Baxter. Their infectious enthusiasm has inspired me to follow in their footsteps. Thanks go to Dr Sajjad Awan for his detailed guidance on water-use efficiency experiments. Thanks are due to Peter Watson, Trevor Harold, Sally Mann, Rebecca Jones, Colin Jones and John Baker for consistently ensuring that glasshouse and field experiments ran as smoothly as possible. I thank my advisory panel, Dr Jose Gutierrez-Marcos and Professor Orkun Soyer, for constructive feedback throughout the project.

I am grateful to the Midlands Integrative Biosciences Training Partnership (MIBTP), the Biotechnology and Biological Sciences Research Council (BBSRC) and the Medical and Life Sciences Research Fund, whose studentship and financial support enabled me to undertake this training. The studentship and the accommodating staff of CABI allowed me to embark on a fantastic three-month placement in Uganda, where I evaluated the operations of plant clinics.

Lindsay and Ross MacLennan, and Bess and Jack Tock have helped me to keep my sense of humour and perspective throughout. Most of all, I am thankful to my mum and dad for their encouragement and support throughout my education, and to my partner, Giulia, for her patience, humour and support.

Declaration

This thesis is submitted to the University of Warwick in support of my application for the degree of Doctor of Philosophy. It has been composed by myself and has not been submitted in any previous application for any degree. Due to the nature of the research aims, sections of text in the **General introduction** and **Chapter 2** (**2.2. MATERIALS AND METHODS**) are similar to parts of the thesis submitted in September 2012 to the University of Warwick in support of the author's application for the degree of Master of Science in Sustainable Crop Production: Agronomy for the 21st Century, entitled "Reviving Phaseolus bean germplasm to enable navy bean genetics for UK production".

Parts of this PhD thesis in the **General introduction**, **Chapter 2** and the **General discussion** (on halo blight) have been used in the preparation of a manuscript for publication provisionally entitled "Genome-wide linkage and association mapping of halo blight resistance in common bean to Race 6 of the globally important bacterial pathogen". Authors of the manuscript are Andrew J. Tock (first author), Deidré Fourie, Peter G. Walley, Eric B. Holub, Alvaro Soler, Karen A. Cichy, Marcial A. Pastor-Corrales, Qijian Song, Timothy G. Porch, John P. Hart, Renato C. C. Vasconcellos, Joana G. Vicente, Guy C. Barker and Phillip N. Miklas (corresponding author).

Contributors

The work presented (including data generated and data analysis) was carried out by the author, except in the cases outlined below:

Siva Samavedam contributed bioinformatics analysis and training in quality processing, reference-based alignment, variant detection and genetic marker development using plant transcriptome datasets.

Lesley Ward and Jeanette Selby performed library preparations for all plant transcriptomic and bacterial genomic nucleic acid samples. Transcriptome libraries were sequenced by Lesley Ward and Jeanette Selby at the Genomics Facility, School of Life Sciences, University of Warwick. Bacterial genomic libraries were sequenced by Illumina Cambridge Ltd. at Chesterford Research Park, Saffron Walden, UK.

Staff at the Institute of Biotechnology Genomic Diversity Facility, Cornell University, USA, performed library preparations and sequencing of plant DNA samples submitted for genotyping-by-sequencing.

Photographs in Figure 1.1 were used with permission from Professor Howard F. Schwartz (Colorado State University). Figure 1.2 was taken from CABI (2016a) with permission. Table 1.1 was taken and adapted from Miklas *et al*. (2014) and Taylor *et al*. (1996a) with permission from Dr Phillip N. Miklas (USDA Agricultural Research Service) and Dr John D. Taylor. The variant-calling pipeline depicted in Figure 2.2 was used with permission from Dr Dan Bolser (European Bioinformatics Institute).

Summary

This research establishes molecular breeding capability for adapting common bean (*Phaseolus vulgaris* L.) to UK growing conditions. A high-resolution linkage map was constructed for a bi-parental recombinant inbred population (large brown × small white haricot) using genotyping-by-sequencing data. Pre-breeding material was exploited to enable genetic mapping and marker-assisted selection of essential adaptive traits, including (**1**) resistance to halo blight, caused by *Pseudomonas syringae* pathovar *phaseolicola* (*Psph*), (**2**) root architecture related to abiotic stress tolerance and nutrient acquisition, (**3**) earliness of maturity, (**4**) plant architecture amenable to mechanical harvest, and (**5**) seed coat colour of consumer interest.

A 500-kb mapping interval was defined for quantitative resistance to the broadly virulent *Psph* race 6, a devastating bacterial pathogen that threatens global bean production with losses from halo blight. Complementary research generating high-quality draft genomes for 32 pathogenically and geographically diverse isolates of *Psph* identified five high-probability candidate determinants of the broad virulence of *Psph* race 6, including avirulence protein AvrD. Pathogenicity effectors that are highly conserved within the pathovar were identified as candidate targets for potential race-nonspecific resistance to halo blight.

Putative QTL for root architecture traits associated with water and nutrient acquisition were detected on chromosome Pv07. A useful breeding strategy may be to select for larger taproot diameter in view of the comparatively high heritability of this trait. Potentially desirable alleles on Pv07 are linked in coupling phase with the dominant allele of seed coat pigmentation factor *P*. Identification of lines recombinant for these alleles may prove useful for the introgression of genes governing physiological resilience into white-seeded varieties adapted to UK growing conditions. Provisional QTL for morphological and reproductive traits of agronomic importance, including plant architecture, growth stage and yield, were identified using phenotypic data obtained from pilot field and polytunnel evaluations of the recombinant inbred population.

Abbreviations

A – net photosynthetic rate

ADF – Assay Development Format

ADP – Andean Diversity Panel of *Phaseolus vulgaris* lines

ANOVA – analysis of variance

ATACX[n] – *Arabidopsis thaliana* ACYL-CoA OXIDASE

A. thaliana – *Arabidopsis thaliana* (Linnaeus) Heynhold

avr gene/Avr protein – pathogen avirulence gene/protein

BAM – Binary Alignment/Map file format

BCMV – *Bean common mosaic virus*

BGYMV – *Bean golden yellow mosaic virus*

BLAST – Basic Local Alignment Search Tool

BLASTN – nucleotide BLAST

BLASTP – protein BLAST

B[*n*] – *Phaseolus vulgaris* genetic map linkage group number

bp – base pair

BRGA – basal root growth angle

BRIG – BLAST Ring Image Generator

BRN – basal root number

BRWN – basal root whorl number

BWA – Burrows-Wheeler Aligner

BYMV – *Bean yellow mosaic virus*

CABI – Centre for Agriculture and Biosciences International

CAPS – cleaved amplified polymorphic sequence

Cas-9 – CRISPR-associated protein-9 nuclease

CbiX – cobalamin (vitamin B12) biosynthesis protein

CC – coiled-coil N-terminal domain of plant disease resistance protein

CDC – cell division cycle protein

CDK – cyclin-dependent protein kinase

CDS – coding DNA sequence(s)

cfu – colony-forming units

CIAT – International Center for Tropical Agriculture

CLP1 – pre-mRNA CLEAVAGE AND POLYADENYLATION FACTOR 1

cM – centiMorgan

Contig – contiguous sequence

CRISPR – clustered, regularly interspaced, short palindromic repeat

CRK – cysteine-rich receptor-like protein kinase

CTAB – cetyltrimethyl ammonium bromide

Cultivar – cultivated variety

DAS – days after sowing

DEFL – defensin-like gene

d.f. – degrees of freedom

EST – expressed sequence tag

ETI – effector-triggered immunity

FAO – Food and Agriculture Organization of the United Nations

FC – field capacity

$F_{[n]}$ – the *n*th filial generation

FRET – fluorescence resonance energy transfer

GBS – genotyping-by-sequencing

GN#1 Sel 27 – Great Northern Nebraska Number 1 Selection 27

GRP – glycine-rich RNA-binding protein

g_s – stomatal conductance

GWA – genome-wide association

h – hour(s)

ha – hectare(s)

HMM – hidden Markov model

Hop – hypersensitive response and pathogenicity outer protein

Hrc – Hrp conserved

Hrp – hypersensitive response and pathogenicity

HSD – Tukey's honest significant difference

IDH/cICDH/IMDH – (cytosolic) isocitrate/isopropylmalate dehydrogenase

INDEL – nucleotide sequence insertion or deletion

IUPAC – International Union of Pure and Applied Chemistry

IWUE – intrinsic water-use efficiency

JA – jasmonic acid

JAZ – JASMONATE ZIM DOMAIN

JDT population – haricot-type recombinant inbred lines selected by Dr John D. Taylor and colleagues for resistance to halo blight and physiological traits, whose pedigree is described in Dodd and Taylor (1991, unpublished data)

KASP – competitive allele-specific PCR

kb – kilobase

kDa – kilodalton

LBD – lateral organ boundaries domain-containing protein

LOD – logarithm (base 10) of the odds

LOV1 – LOCUS ORCHESTRATING VICTORIN EFFECTS 1

LRR – leucine-rich repeat protein/domain

LSD – least significant difference

MAF – minor allele frequency

MAPK/MPK – mitogen-activated protein kinase

MAPKK/MKK – mitogen-activated protein kinase kinase

Mb – megabase

MPa – Megapascal

mRNA – messenger RNA

mtRRL – multi-tier reduced-representation library

NADP – nicotinamide adenine dinucleotide phosphate

NB-ARC – nucleotide-binding adapter shared by APAF-1, Resistance proteins, and CED-4 domain

NBS – nucleotide-binding site

NCBI – National Center for Biotechnology Information

NDR1 – NON RACE-SPECIFIC DISEASE RESISTANCE 1

NGS – next-generation sequencing

NLR – nucleotide-binding site–leucine-rich repeat protein

NVRS – National Vegetable Research Station

OD – optical density

P. acutifolius – *Phaseolus acutifolius* Gray (tepary bean)

PAMP – pathogen-associated molecular pattern

P. angustissimus – *Phaseolus angustissimus* Linnaeus

PBS1 – AVRPPHB SUSCEPTIBLE 1

PCR – polymerase chain reaction

P. fluorescens / *Pfl* – *Pseudomonas fluorescens*

PI – plant introduction (gene bank accession)

PNPT – polyribonucleotide nucleotidyltransferase

PPFS – Predict Phenotypes From SNPs package for statistical association analysis

PPR – Pentatricopeptide repeat-containing protein

PPV – positive predictive value

PRP-BP – PvPRP1-BINDING PROTEIN

PRR – pattern recognition receptor

Pse-[*n*] – gene symbol for dominant race-specific resistance to *Pseudomonas syringae* pathovar *phaseolicola* (causing halo blight) in *Phaseolus vulgaris*

Psg – *Pseudomonas syringae* pathovar *glycinea*

Psm – *Pseudomonas syringae* pathovar *maculicola*

Psph – *Pseudomonas syringae* pathovar *phaseolicola*

Pspi – *Pseudomonas syringae* pathovar *pisi*

Pss – *Pseudomonas syringae* pathovar *syringae*

Pst – *Pseudomonas syringae* pathovar *tomato*

P. syringae – *Pseudomonas syringae*

PTI – PAMP-triggered immunity

pv. – pathovar (pathogen variant)

Pv[*n*] – *Phaseolus vulgaris* reference-genome chromosome number (Schmutz *et al*., 2014)

PvPRP1 – *Phaseolus vulgaris* PROLINE-RICH PROTEIN 1

P. vulgaris – *Phaseolus vulgaris* Linnaeus (common bean)

QTL – quantitative trait locus/loci

R – Pearson product-moment correlation coefficient

R^2 – proportion of phenotypic variation explained by locus

RBP – RNA-binding protein

RCBD – randomised complete block design

RefSeq – the Reference Sequence non-redundant protein database of the National Center for Biotechnology Information

REML – restricted maximum likelihood

R gene/R protein – plant disease resistance gene/protein

RIL – recombinant inbred line

RO-DI – reverse-osmosis-fed deionised (water)

ROS – reactive oxygen species

RPF1 – RNA PROCESSING FACTOR 1

rpm – revolutions per minute

RPM[n] – disease resistance protein conferring resistance to *Pseudomonas syringae* pathovars in *Arabidopsis thaliana*

RPP[n] – disease resistance protein conferring resistance to *Peronospora parasitica* (causing downy mildew) in *Arabidopsis thaliana*

RPS[n] – disease resistance protein conferring resistance to *Pseudomonas syringae* pathovars in *Arabidopsis thaliana*

RRM – RNA recognition motif

SA – salicylic acid

SAM – Sequence Alignment/Map file format

SARF – sum of adjacent recombination fractions

SD – sample standard deviation

SE RIL – SOA-BN × Edmund recombinant inbred line

SIF – Simple Interaction Format

SIRB – SIROHYDROCHLORIN FERROCHELATASE B

SNP – single-nucleotide polymorphism

SNV – single-nucleotide variant

sp. – a singular unspecified species within a given genus

spp. – multiple unspecified species within a given genus

SSCP – single-strand conformational polymorphism

SSH – subtraction suppression hybridisation

SSR – simple sequence repeats (also known as microsatellites)

Syn. – synonymous with a previously assigned gene or protein name

t – tonnes

T3E – type III-secreted effector

T3SS – type III secretion system

T4SS – type IV secretion system

TAIR – The Arabidopsis Information Resource

TBLASTN – translated nucleotide BLAST using protein query sequence

T-DNA – transfer DNA

TILLING – Targeting Induced Local Lesions IN Genomes

TIR – Toll/interleukin-1 receptor-like N-terminal domain of plant disease resistance protein

T_m – melting temperature

UPL – ubiquitin-protein ligase

USDA – United States Department of Agriculture

VCF – Variant Call Format

VICTL – VARIATION IN COMPOUND TRIGGERED ROOT GROWTH RESPONSE-LIKE (TIR–NLR class disease resistance protein)

VICTR – VARIATION IN COMPOUND TRIGGERED ROOT GROWTH RESPONSE (TIR–NLR class disease resistance protein)

Warwick HRI – Warwick Horticultural Research Institute

WRR[n] – disease resistance protein conferring resistance to *Albugo candida* (white rust) in *Arabidopsis thaliana*

WUE – water-use efficiency

xa[*n*] – recessive allele conferring resistance to the rice bacterial blight pathogen *Xanthomonas oryzae* pathovar *oryzae*

Chapter 1

General introduction

1.1. A NUTRITIOUS FOOD OF GLOBAL IMPORTANCE

Background

Common bean (*Phaseolus vulgaris* L., $2n = 2x = 22$) is a staple source of gluten-free protein, carbohydrates, fibre, vitamins and micronutrients for direct consumption in the human diet (Miklas *et al*., 2006). It is the main ingredient in a British favourite comfort food, beans-on-toast, and complements carbohydrate and amino acid nutrition derived from cereal grains (Bressani, 1983). Globally, we eat more common beans than any other food legume (Broughton *et al.*, 2003). They are a major source of dietary protein for people in developing countries in the face of scarce and unaffordable animal protein (Broughton *et al*., 2003; Beebe *et al*., 2013).

The higher protein content of legumes compared with other plants is primarily attributable to their ability to fix atmospheric nitrogen (N) in root nodules by forming symbiotic relationships with soil rhizobia (Broughton *et al*., 2003). Grain legumes like dry beans are regularly intercropped or rotated with cereals by smallholders in sub-Saharan Africa (Dakora and Keya, 1997). They are adapted to conditions in widely distributed ecological zones throughout the tropics and subtropics, including in semi-arid regions where soil fertility is poor (Shisanya and Gitonga, 2007). Coupled with their capacity for biological N fixation, this makes them useful crops for restoring nutrient-depleted soils, providing resource-poor farmers with a valuable alternative to relying principally on costly synthetic fertilisers.

Despite its nutritional value and small genome (~ 587 Mb; Schmutz *et al.*, 2014), progress in common bean improvement by means of genome-based genetic tools has been hampered by its comparatively low economic value, with its cultivation and consumption as a staple crop occurring primarily in developing countries (Galeano *et al.*, 2009a,b). Nonetheless, the release of the first chromosome-scale version of the assembled common bean reference genome (*Phaseolus vulgaris* V1.0; Schmutz *et al*., 2014),[1] and breakthrough technologies for high-throughput genotyping, are enabling major achievements in linkage and association mapping of economically important disease resistance, market class and

[1] Sequence data are available from Phytozome at http://www.phytozome.net/commonbean (Goodstein *et al*., 2012).

agronomic traits (Moghaddam *et al*., 2014; Cichy *et al*., 2015; Kamfwa *et al*., 2015; Song *et al.*, 2015; Perseguini *et al*., 2016; Zuiderveen *et al*., 2016).

Common bean is an important ingredient of the British diet, with people in the United Kingdom opening more than one million tins of haricot beans per day (Thring, 2011; H. J. Heinz Co. Ltd., c2016). However, common bean is not grown on a commercial scale in this country (Food and Agriculture Organization of the United Nations, c2015). Consequently, UK consumers depend entirely upon imports of dry common beans of all market classes, largely from Canadian production (FAO, c2013).

A successfully adapted common bean cultivar can be developed with the application of modern genetics and breeding tools, and would provide UK farmers with a viable new rotational crop. A UK-adapted cultivar would combine several vital characteristics that define the crop ideotype. In particular, it would exhibit cold tolerance for seedling establishment in spring, resistance to diseases occurring over summer, and earliness of maturity such that it is harvestable in early September, before the onset of autumnal damp weather that can cause seed spoilage due to fungal infection and discolouration (Austin and MacLean, 1972a; Conway *et al*., 1982; Hardwick, 1988; Dodd and Taylor, 1991, unpublished data). It would also incorporate a number of essential physiological, morphological and reproductive traits, including efficient acquisition and use of water and nutrients to maximise yield and minimise farm inputs, an upright plant architecture conducive to mechanical harvest, and seeds of the colour, size, shape and quality demanded by local markets (Conway *et al*., 1982).

Agronomic gains achievable through marker-assisted selection for disease resistance, physiological resilience and seed quality traits have the potential to extend beyond UK production. Approximately 50% and 25% of global common bean production occurs in Latin America and Africa (Beebe *et al*., 2013). Most smallholding farmers in these regions lack adequate access to agronomic advice and affordable synthetic treatments for biotic and abiotic stresses that severely constrain yield (Beebe *et al*., 2009; Danielsen and Kelly, 2010). Crop genetic improvements are therefore imperative for improving household food security and sustainable production.

1.2. HALO BLIGHT

In the UK, halo blight (caused by *Pseudomonas syringae* pathovar *phaseolicola*), *Bean common mosaic virus* (BCMV), anthracnose (*Colletotrichum lindemuthianum*), and white mould (*Sclerotinia sclerotiorum*) are among the serious potential diseases of common bean (Conway *et al*., 1982; Hardwick, 1988). Genes conferring resistance to these diseases have been incorporated into bean lines bred for UK conditions (Conway *et al*., 1982; Hardwick, 1988). One of the parental lines being exploited in the current research is haricot (navy bean) cultivar Edmund, which combines resistances to halo blight, BCMV and anthracnose. This registered variety was developed at the National Vegetable Research Station (NVRS) near Wellesbourne as part of a UK navy bean breeding programme (Conway *et al.*, 1982), and incorporates good yield and canning characteristics.

Pathogen biology

Halo blight is an economically important disease of common bean, caused by the Gram-negative bacterium *Pseudomonas syringae* pv. *phaseolicola* (*Psph*) (Burkholder, 1926; Young *et al*., 1978). This seed-borne bacterium actively reproduces following germination of bean seeds. After seedling emergence, contaminated plant exudates are spread from plant to plant by water droplets carried on the wind during wet weather or irrigation, or by physical contact between adjacent plants. *Psph* enters cultivated and weedy hosts through stomata and wounds (Arnold *et al*., 2011). As a hemibiotroph, the pathogen can sequester nutrients from both living and, later in pathogenesis, dead or dying host tissue cells (Lindeberg *et al*., 2012). It survives and grows parasitically in intracellular spaces of living plant tissues, and subsequently causes host cell death (Arnold *et al*., 2011; Lindeberg *et al*., 2012). *P. syringae* delivers determinants of pathogenicity (disease-causing ability) and virulence (degree of pathogenic capability) known as effectors into host cells via a specialised organelle, the type III secretion system (T3SS). These effectors allow the pathogen to subvert immune responses and cause disease in susceptible hosts, but can trigger defence responses that arrest pathogen proliferation in resistant plants encoding cognate receptor proteins (Arnold *et al*., 2009, 2011).

Symptoms of halo blight infection on field-grown common beans include small water-soaked and/or necrotic lesions on leaves, pods, petioles or stems. At temperatures below 23 °C, *Psph* produces more phaseolotoxin, which causes lesions to develop broad yellow-green haloes (Figure 1.1; Harveson, 2009; Arnold *et al*., 2011). Systemic infection is characterised by generalised chlorosis, and inhibition and distortion of vegetative growth (Harveson, 2009; Arnold *et al*., 2011).

Figure 1.1. Halo blight symptoms observed in the field on *Phaseolus vulgaris* leaves and pod. Photographs are courtesy of Howard F. Schwartz, Colorado State University.

Prevalence and prevention

Halo blight is globally distributed and is most damaging in regions with temperate climates and at higher altitudes in the tropics (Figure 1.2; Taylor *et al*., 1996a; Harveson, 2009; CABI, 2016a,b). *Psph* is a devastating crop pathogen and has been reported to cause yield losses of up to 45% (Prosen *et al*., 1993; Asensio-S.-Manzanera *et al*., 2006; Singh and Schwartz, 2010; Félix-Gastélum *et al*., 2016). Genetic resistance provides the most effective method of halo blight control and is critical to the production of disease-free seed (Taylor *et al*., 1996a; Asensio-S.-Manzanera *et al*., 2006; Arnold *et al*., 2011; Miklas *et al*., 2014). In developing countries, many farmers save seed from a previous harvest, and crop rotation is not practised as widely as in developed countries. Consequently, the production of

pathogen-free seed within this context is often impracticable for farmers growing susceptible varieties (Taylor *et al*., 1996a). Alternative hosts present in the tropics, coupled with the commonly short periods between consecutive crops, would further limit the long-term feasibility of maintaining clean seed of susceptible varieties (Taylor *et al*., 1996a).

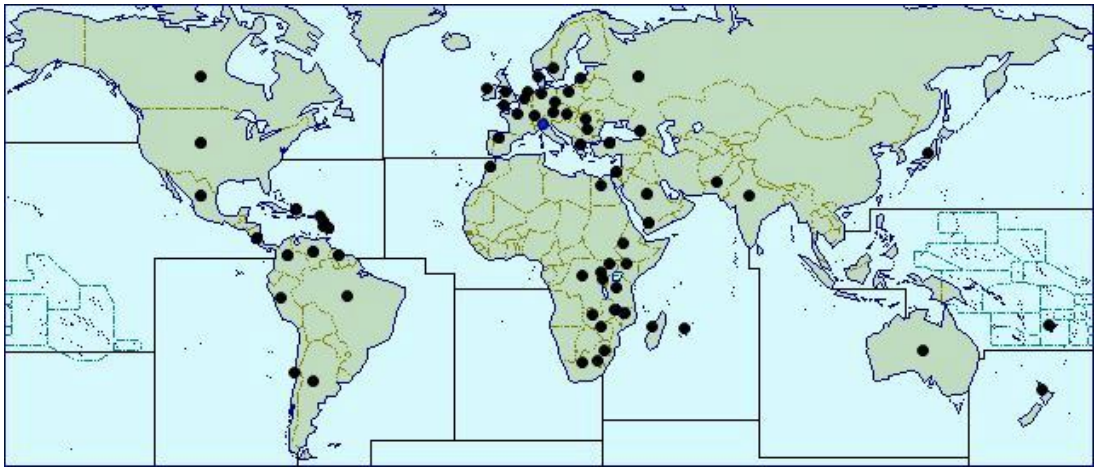


Figure 1.2. Global distribution of *Pseudomonas syringae* pv. *phaseolicola* (taken from CABI, 2016a, with permission).
● = present, no further details; ● = widespread.

Race structure

Seminal research in the 1990s established Phaseolus–Pseudomonas as an experimental pathosystem for investigating the molecular basis of plant–pathogen interactions (Teverson, 1991; Jenner *et al*., 1991; Mansfield *et al*., 1994; Taylor *et al*., 1996a,b; Stevens *et al*., 1998; Tsiamis *et al*., 2000). Taylor *et al*. (1996a) identified nine pathotypic races of *Psph* from an extensive collection of isolates by distinct patterns of resistance and susceptibility across eight differential bean cultivars and lines. Genetic evidence indicated a gene-for-gene relationship whereby race-specific plant disease resistance (*R*) genes trigger defence responses upon recognition of pathogen avirulence proteins. Race designations were thus informed by the predicted presence or absence of avirulence proteins based on the phenotypic attributes of *Psph* isolates.

This work applied and built upon the formative host–parasite gene-for-gene model developed by Flor (1942), who investigated the genetic basis of resistance and pathogenicity in a plant pathosystem using the flax rust fungus *Melampsora lini*. Under this hypothesis, plant–pathogen interaction phenotypes exhibited by resistant

plants are governed by genes in the host that correspond to genes controlling pathogenicity in the pathogen in a pairwise relationship (reviewed in Flor, 1971). Flor (1971) cited studies of 18 different pathosystems for which gene-for-gene relationships have been proposed. The model enables prediction of host and pathogen genotypes based on qualitative interaction phenotypes (compatible or incompatible) elicited from a series of host differential lines following challenge with a series of pathogen variants (Flor, 1971).

Interaction phenotypes observed by Taylor *et al*. (1996a) indicated that five major-effect *R* genes in Phaseolus provide resistance against all of the *Psph* races except race 6 (Table 1.1, taken and adapted from Miklas *et al*., 2014 and Taylor *et al*., 1996a,b, with permission). These *R* genes govern race-specific interactions with pathogen avirulence (*avr*) genes. Four of these *R* genes have since been localised on different common bean chromosomes by linkage mapping using various recombinant inbred populations: *Pse-1* and *Pse-2* were mapped to different locations on *P. vulgaris* chromosome 10 (Pv10) by Miklas *et al*. (2009, 2011); *Pse-3* is tightly linked with the *I* gene conferring resistance to BCMV at the telomeric end of the long arm of Pv02 (Miklas *et al*., 2011, 2014); and a preliminary mapping of *Pse-4* to Pv10 by Miklas *et al*. (2014) has since been revised based on more conclusive evidence indicating a different location (P. N. Miklas, personal communication). Subsequently obtained data suggest that the gene from common bean line BelNeb-RR-1 mapped to Pv10 is distinct from *Pse-4* (P. N. Miklas, personal communication; Table 1.1). Miklas *et al*. (2014) also suggested that multiple genes in common bean may confer resistance to race 5 only, based on segregation ratios not consistent with single-gene inheritance.

Similarly to the *Pse-1* locus (Miklas *et al*., 2009), *Pse-6* from BelNeb-RR-1 confers independent dominant resistance to races 1, 5, 7 and 9, and is located at the telomeric end of the short arm of chromosome Pv04 (Miklas *et al*., 2014). Miklas *et al*. (2014) reported an apparent recombination event at the *Pse-6* locus, as one BelNeb-RR-1 × A55 RIL was resistant to race 7 and susceptible to races 1 and 9. Using different RIL populations, Miklas *et al*. (2014) placed provisionally named *R* genes *Pse-(Race 1)* and *Pse-(Race 7)* (which confer race-specific resistances) near *Pse-6* on Pv04. The *Pse-6* locus on Pv04 was thus resolved as a putative cluster of *R* genes that separately confer different resistance specificities (Miklas *et al*., 2014). Miklas *et al*. (2011) also determined a broader effect for *Pse-2* (Pv10) than was

previously reported (Teverson, 1991; Taylor *et al*., 1996a,b) based on complete cosegregation of resistance to *Psph* races 2, 3, 4, 5, 7, 8 and 9 in recombinant inbred lines (RILs) derived from a cross between *P. vulgaris* differential lines ZAA 12 and Canadian Wonder (Table 1.1). These genetically characterised genes confer largely dominant resistance, although recessive resistances to race 5 and to race 8 have been predicted (Teverson, 1991; Miklas *et al*., 2011, 2014).

Table 1.1. Interaction phenotypes exhibited by *Phaseolus vulgaris* (common bean) and *P. acutifolius* (tepary bean) host differential lines following inoculation with isolates of nine races of *Pseudomonas syringae* pv. *phaseolicola*, with predicted race-specific resistance (*R*) genes indicated. Taken and adapted from Miklas *et al*. (2014: 2107) and Taylor *et al*. (1996a,b: 474, 482) with permission

			Race								
Host differential	***R* genes**	**Chromosome**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Canadian Wonder	None		+	+	+	+	+	+	+	+	+
ZAA 54 (A52)	*Pse-4*	Unmapped	+	+	+	+	−	+	+	+	+
Tendergreen	*Pse-3*	Unmapped	+	+	−HR	−HR	+	+	+	+	+
A55	*Pse-3*	Pv02	+	+	−HR	−HR	+	+	+	+	+
Red Mexican UI-3	*Pse-1*	Pv10	−	+	+	+	−	+	−	+	−
	Pse-4	Unmapped	+	+	+	+	−	+	+	+	+
1072 (tepary)	?‡	Unmapped	+	−	+	−	−	+	−	+	+
ZAA 55 (A53)	*Pse-3*, *Pse-4*	Unmapped	+	+	−HR	−HR	−	+	+	+	+
ZAA 12 (A43)	*Pse-2*	Pv10	+	−	−HR	−HR	−	+	−	−	−
	Pse-3	Pv02	+	+	−HR	−HR	+	+	+	+	+
	Pse-4	Unmapped	+	+	+	+	−	+	+	+	+
Guatemala 196-B	*Pse-1*, *Pse-3*, *Pse-4*	Unmapped	−	+	−HR	−HR	−	+	−	+	−
BelNeb-RR-1	*Pse-4* or *Pse-?*†	Pv10†	+	+	+	+	−	+	+	+	+
	Pse-6	Pv04	−	+	+	+	−	+	−	+	−
Minuette	*Pse-3*, *Pse-4*	Unmapped	+	+	−HR	−HR	−	+	+	+	NT
	Pse-(Race 7)	Pv04	+	+	+	+	+	+	−	+	+
I9365-31	*Pse-3*, *Pse-4*	Unmapped	+	+	−HR	−HR	−	+	+	+	+
	Pse-(Race 1)	Pv04	−	+	+	+	+	+	+	(−)	+

Pse-1 was placed on *Phaseolus vulgaris* chromosome 10 (Pv10) by Miklas *et al*. (2009). *Pse-2* was also mapped to Pv10 by Miklas *et al*. (2011). *Pse-3* was mapped to Pv02 and found to be tightly linked with the *I* gene for *Bean common mosaic virus* resistance by Miklas *et al*. (2011, 2014). †*Pse-4* was provisionally positioned on Pv10 by Miklas *et al*. (2014), but a putative different location has been determined based on subsequently obtained data (P. N. Miklas, personal communication). This suggests that the gene on Pv10 conferring resistance solely to race 5 in BelNeb-RR-1 may not be *Pse-4*. *Pse-6*, *Pse-(Race 7)* and *Pse-(Race 1)* were mapped to the telomeric end of the short arm of Pv04 by Miklas *et al.* (2014). *Pse-(Race 7)* and *Pse-(Race 1)* are provisional gene symbols. ‡*Pse-2* was formerly proposed to be present in the *P. acutifolius* (tepary bean) differential line 1072 (Teverson, 1991; Taylor *et al.*, 1996a,b). Miklas *et al.* (2011) showed that the broader effect of *Pse-2* in ZAA 12 (A43) indicates that this inference is no longer valid.

+: apparent susceptible (compatible) reaction. −: apparent resistant (incompatible) reaction. −HR: apparent resistant reaction with severe hypersensitive cell death. (−): partially resistant. NT: not tested.

Psph race 6 is undetected by previously mapped major *R* genes

Progress achieved by recent efforts to localise major-effect, race-specific *R* genes will enable their deployment in cultivars through marker-assisted breeding. However, evolution of pathogen virulence continues to hamper effective halo blight control. None of the major *R* genes that have been mapped to date is effective against *Psph* race 6 (Miklas *et al*., 2014), which defines the most broadly virulent pathotype. Race 6 has a global distribution and was found to be the most frequently occurring race (particularly throughout Africa, Latin America and Europe) in the race differentiation survey of 175 isolates conducted by Taylor *et al*. (1996a). The global predominance of this race may reflect the absence of a widely deployed race-specific *R* gene effective against race 6 (Taylor *et al*., 1996b). As a broadly virulent pathogen on Phaseolus beans, *Ppsh* race 6 continues to threaten bean production worldwide with losses from halo blight (Lamppa *et al*., 2002; Rico *et al*., 2003; Félix-Gastélum *et al*., 2016).

Taylor *et al*. (1996b) evaluated a wide diversity collection of Phaseolus (1,048 accessions from the Americas and Africa) with six *Psph* races, and identified a low frequency of quantitative, potentially race-nonspecific, resistance that includes resistance to race 6. Quantitative resistance is partial and expressed as a continuum of variation between the highest and lowest levels of disease resistance. It is distinct from qualitative or race-specific resistance, which is expressed as qualitatively discrete phenotypic categories or in large increments (Federation of British Plant Pathologists, 1973, via Teverson, 1991). Qualitative resistance is generally controlled by single, race-specific *R* genes that induce a localised hypersensitive cell death response to halt pathogen growth. Quantitative resistance typically has polygenic inheritance, with quantitative trait loci (QTL) contributing partial effects that slow or limit the extent of disease development (Hulbert *et al.*, 2001; Boyd *et al*., 2013; González *et al.*, 2016). In some cases, these QTL can confer broad-spectrum resistance to multiple pathogens and/or races, such as the wheat leaf rust (*Puccinia triticina*), yellow/stripe rust (*Puccinia striiformis* f. sp. *tritici*) and powdery mildew (*Blumeria graminis* f. sp. *tritici*) resistance factors *Lr34/Yr18/Pm38* and *Lr46/Yr29/Pm39*, and the rice blast (*Magnaporthe oryzae*) race-nonspecific resistance factor *PI21* (Leach *et al.*, 2001; Boyd *et al.*, 2013).

Quantitative resistance from PI 150414 (‘Frijol Rojo’, a red dry bean landrace from El Salvador; Patel and Walker, 1965) and Great Northern Nebraska No. 1

Selection 27 (GN#1 Sel 27, a selection from a great northern landrace; Miklas *et al*., 2003) has been deployed in various cultivars and lines. These include Edmund, Wis HBR 40, Wis HBR 72, Kranskop-HR 1 and RH13 with resistance from PI 150414 (Hagedorn *et al*., 1974; Fouilloux and Bannerot, 1977, via Taylor *et al*., 1978; Taylor *et al*., 1996b), and cultivars Jules and Harris with resistance from GN#1 Sel 27 (P. N. Miklas, personal communication). Definitive evidence for a shared or different allele(s) conferring quantitative resistance in PI 150414 and GN#1 Sel 27 is lacking (Coyne *et al*., 1971; Hill *et al*., 1978; Taylor *et al*., 1978). Delayed flowering in GN#1 Sel 27 and delayed pod formation in PI 150414 render these two lines unsuitable for UK production (Taylor *et al*., 1978).

The quantitative resistance from PI 150414 has been deployed widely in dry and snap bean breeding programmes for at least two decades (J. D. Taylor, personal communication; also see Silbernagel and Hannan, 1992), helping to reduce bean crop losses. However, the extent to which this resistance is durable under field conditions that would otherwise favour disease progression is unknown, although no resistance-breaking *Psph* strains have been reported (J. D. Taylor, personal communication). Genetic experiments indicated simple inheritance of quantitative resistance in PI 150414 (Taylor *et al*., 1978, and references therein). It appeared to be recessively inherited in separate crosses with two susceptible cultivars (also see Patel and Walker, 1966; Dickson and Natti, 1966), but was incompletely dominant in crosses with another cultivar lacking resistance (Taylor *et al*., 1978; also see Hill *et al*., 1972). This globally useful and potentially durable resistance is the primary target for genetic mapping in the current research, as described in Chapter 2.

Two further examples of quantitative, potentially race-nonspecific, *Psph* resistance were identified in components of farmers' mixtures from Tanzania (Small Masusu; Teverson, 2003) and Rwanda (2702/2; Taylor *et al*., 1996b). Preliminary evidence from allelism tests indicated that these resistances are genetically distinct from each other and from that in PI 150414 (J. D. Taylor, personal communication). These putative additional sources might therefore provide opportunities for *R* gene 'pyramiding' to bolster quantitative resistance against evolving pathogen populations (Vicente *et al*., 2012).

CAL 143 (a CIAT-bred red-speckled calima dry bean cultivar popular in East Africa; Chataike *et al*., 2011; R. Chirwa, personal communication) combines useful levels of quantitative halo blight resistance effective against *Psph* race 6 (Miklas *et*

al., 2006; P. N. Miklas, personal communication) with resistances to bean rust (*Uromyces appendiculatus*; Murillo *et al*., 2006) and angular leaf spot (*Phaeoisariopsis griseola*; Chataika *et al*., 2010; Oblessuc *et al*., 2012). CAL 143 exhibited monogenic dominant resistance to uncharacterised field isolates of *Psph* endemic to Malawi (Chataika *et al*., 2011).

Ghising *et al*. (2016) mapped resistance to *Psph* race 6 to the same region of chromosome Pv04 to which *Pse-6* (conferring dominant, race-specific resistance to races 1, 5, 7 and 9), *Pse-(Race 1)* and *Pse-(Race 7)* were mapped by Miklas *et al*. (2014). This locus on Pv04 and another on Pv05 were found to be associated with resistance to race 6 in a genome-wide association study of the United States Department of Agriculture (USDA) core collection of Mesoamerican and Andean bean accessions (Ghising *et al*., 2016). Among the 383 accessions examined, ten were highly resistant and were thus identified as potentially novel sources of race 6 resistance (PI 313531, PI 313727, PI 325684, PI 311940, PI 319677, PI 355419, PI 511767, PI 531862, PI 207193 and PI 290995). However, the markers identified on Pv04 and Pv05 accounted for low percentages of phenotypic variation for race 6 resistance observed within the collection (each ≤ 8%; Ghising *et al*., 2016). None of the accessions was evaluated for resistance to other races.

Trabanco *et al*. (2014) detected minor-effect QTL on Pv04 (distant from the *Pse-6* locus at the telomeric end of the short arm) and Pv06 for resistance to *Psph* race 6, which explained 11% and 12% of phenotypic variation in a Xana × Cornell 49-242 recombinant inbred population. They also identified QTL for resistance to race 7 at the same Pv06 locus and at another location on Pv06, accounting for 16% and 27% of phenotypic variation. However, the proposed source of resistance in this population, Cornell 49-242, was not observed by Taylor *et al*. (1978) to be resistant to either *Psph* race 1 (which has since been subdivided into races 1, 5, 7 and 9; Taylor *et al*., 1996a) or race 2 (subsequently subdivided into races 2, 6 and 8) in leaf and pod inoculation experiments.

Duncan *et al*. (2008, 2014a,b) identified a novel source of high-level resistance effective against *Psph* race 6, governed by two independently inherited recessive genes in common bean pinto US14. They released the breeding line US14HBR6, which was developed by selection of single-plant progenies scored as highly resistant over three generations (Duncan *et al.*, 2014a). Genetic mapping of

this resistance has not yet been reported, and its effectiveness against other races was not tested.

González *et al.* (2016) conducted a multi-environment study investigating the genetic basis of quantitative resistance to the nine *Psph* races in primary and trifoliolate leaf, stem and pod tissues using a segregating Andean common bean RIL population (PMB0225 × PHA1037). Resistances from previously identified quantitative halo blight resistance sources were not reported to be present in either of the Andean parents, which were suggested to each possess different quantitative resistance alleles. For reaction to race 6, González *et al.* (2016) identified 11 minor-effect epistatic QTL without detectable additive effects when considered alone (with each interaction explaining ≤ 8.52% of phenotypic variation), and one minor-effect QTL with both epistatic and individual additive effects (explaining 2.64% and 2.04% of phenotypic variation, respectively). These QTL were positioned on linkage groups 1 (three epistatic), 3 (two epistatic), 5 (one epistatic), 6 (one epistatic), 7 (two epistatic, and one epistatic and additive), 9 (one epistatic) and 10 (one epistatic) (see Figure 2 and Tables 1 and 3 in González *et al.*, 2016). QTL with individual additive effects against *Psph* races 3, 4 and 5 in all organs colocalised at the end of linkage group 2, tightly linked with the *I* gene marker SW13. The resistance alleles for most of these QTL were derived from *I*-bearing parent PMB0225. However, resistance alleles at QTL on linkage group 2 for stem resistance to races 3 and 5 were inherited from PHA1037, which lacks the resistance *I* allele that is tightly linked with *Pse-3* conferring race-specific resistance to *Psph* races 3 and 4. González *et al.* (2016) were unable to determine whether organ-nonspecific resistance mapped in this region of linkage group 2 is conferred by *Pse-3* as mapped by Miklas *et al.* (2011) or by other tightly linked genes or QTL. None of the QTL detected by González *et al.* (2016) conferred resistance to all *Psph* races.

Plant disease resistance (R) proteins

The plant immune system confers resistance to pathogens via two main layers of defence response mechanisms. The first layer is mediated by transmembrane pattern-recognition receptors (PRRs), which recognise extracellular microbial elicitors called pathogen-associated molecular patterns (PAMPs) that are expressed by pathogens as they colonise their hosts (Dangl and Jones 2001; Dodds and Rathjen, 2010; Dangl *et al.*, 2013). Cell-surface perception by PRRs of PAMPs, such as

bacterial flagellin or fungal chitin, induces PAMP-triggered immunity (PTI). PTI initiates plant defences (e.g., through intracellular signalling, transcriptional reprogramming and regulation of biosynthetic pathways) that can arrest pathogen proliferation (Dangl *et al*., 2013). Transmembrane PRRs thus play essential defence response roles in plant and animal innate immunity. Many of these PRRs are receptor-like kinases composed of an extracellular leucine-rich repeat (LRR) domain, a transmembrane domain and a cytoplasmic protein kinase domain (Jones and Jones, 1997; Dangl and Jones, 2001; Diévart and Clark, 2004).

While not invariant, PAMPs are generally highly conserved amongst diverse microbial species (Boyd *et al*., 2013). They confer essential functional attributes to pathogens and so evolve more slowly than pathogen virulence effectors (such as those delivered into the host cell via the T3SS), which often are expendable or more variable (Jones and Dangl, 2006). Transmembrane PRRs are similarly highly conserved within plant species. Consequently, they mediate resistance mechanisms that are less vulnerable to pathogen evolution than those governed by intracellular R proteins involved in effector-triggered immunity (ETI) (Dodds and Rathjen, 2010).

Correlations between PTI-mediated defence responses and QTL governing multi-pathogen resistance have been reported in soybean (Valdés-López *et al*., 2011) and wheat (Miedaner *et al*., 2012). Additionally, two QTL governing PTI-based non-host resistance to *Psph* have been reported in *Arabidopsis thaliana*, including *FLS2* encoding the flagellin PRR (Forsyth *et al*., 2010; Ahmad *et al*., 2011). In common bean, LRR-kinases on chromosomes Pv08 (Burt *et al*., 2015) and Pv01 (Zuiderveen *et al*., 2016) have been identified as candidates for potential multi-race resistance to anthracnose. A genome-wide association study of anthracnose resistance in the common bean Andean Diversity Panel (ADP) by Zuiderveen *et al*. (2016), for instance, found the Pv01 locus to be significantly associated with quantitative resistance to multiple races of *Colletotrichum lindemuthianum*.

The second layer of the plant immune system is located inside the cell and is governed by nucleotide-binding site–leucine-rich repeat (NLR) receptors, which represent a highly polymorphic superfamily of intracellular proteins (Dangl *et al.*, 2013). NLRs are characterised by a central nucleotide-binding domain (NB-ARC; nucleotide-binding adapter shared by APAF-1, Resistance proteins, and CED-4), which is often flanked by an amino-terminal coiled coil (CC) or Toll/interleukin-1 receptor (TIR)-like domain and carboxyl-terminal leucine-rich repeats (Sarris *et al*.,

2016). However, variability in NLR architecture (e.g., lacking LRRs or possessing additional functionally important domains) has been proposed to potentially enable recognition of diverse pathogen effectors (Sarris *et al.*, 2016).

The majority of molecularly characterised plant *R* genes encode NLRs (Dangl *et al*., 2013). Activation of NLRs can precipitate ETI through direct or indirect recognition of pathogen avirulence effectors that have been delivered into the host cytoplasm to subvert extracellular PRR-mediated immunity (Dodds and Rathjen, 2010). The biochemical mechanisms by which NLRs recognise effectors are not fully understood, although several instances of direct receptor–ligand interactions between NLRs and avirulence effectors have been reported (Jia *et al*., 2000; Deslandes *et al*., 2003; Kang *et al*., 2005; Dodds *et al*., 2006; Ueda *et al.*, 2006; Catanzariti *et al.*, 2010; Krasileva *et al.*, 2010; Chen *et al.*, 2012; Kanzaki *et al*., 2012; Cesari *et al.*, 2013; Maqbool *et al.*, 2015). Alternatively, the 'guard hypothesis' proposes that some R proteins monitor and detect changes in host proteins targeted by pathogen effectors ('guardees'), or in molecular decoys of these targets (McHale *et al*., 2006; van der Hoorn and Kamoun, 2008; Dangl *et al.*, 2013). Upon detection by an NLR receptor, this surveillance mechanism is thought to activate defence signal transduction pathways that lead to a hypersensitive cell death response, characterised by the accumulation of salicylic acid and reactive oxygen species, that attempts to arrest pathogen infection and proliferation (Woloshen *et al*., 2011).

Several NLR receptors confer resistance to *P. syringae* pathovars expressing specific avirulence effectors. In common bean, for example, *Rpsar-1* and *Rpsar-2* independently confer specific resistance to *P. syringae* pv. *maculicola* strains expressing *avrRpm1* (Chen *et al.*, 2010). The soybean alleles or tightly linked genes *Rpg1-b* and *Rpg1-r* confer specific resistance to *P. syringae* pv. *glycinea* expressing *avrB* and *avrRpm1*, respectively (Ashfield *et al.*, 2003). In Arabidopsis, *RPM1* encodes a cognate NLR receptor with dual resistance specificity against *P. syringae* strains expressing *avrB* or *avrRpm1* (Grant *et al.*, 1995), whereas *RPS2* and *RPS5* confer resistance specificities against strains expressing *avrRpt2* and *avrPphB*, respectively (Mindrinos *et al.*, 1994; Shao *et al.*, 2003). However, the effector target protein RIN4 is monitored and guarded by both RPM1 and RPS2 independently, which detect phosphorylation or cleavage of the guardee by corresponding *P. syringae* effectors (Mackey *et al*., 2002; Axtell *et al.*, 2003; Andersson *et al*., 2006).

Sarris *et al.* (2016) found domains homologous to RIN4 to be fused with predicted NLR proteins in barley, rice and apple, suggesting that they may have been recruited to function as integrated sensors enabling effector perception. Similarly, type III effectors AvrRps4 from *P. syringae* pv. *pisi* and PopP2 from *Ralstonia solanacearum* target numerous WRKY transcription factors with defence roles in Arabidopsis (Sarris *et al*., 2016). Integration of a decoy WRKY DNA-binding carboxyl-terminal domain within the NLR receptor RRS1, operating in a complex with NLR RPS4 partly mediated by their TIR domains, enables pathogen perception and defence activation (Sohn *et al.*, 2014; Saucet *et al.*, 2015; Sarris *et al*., 2015).

Intracellular R proteins often confer qualitative, race-specific resistance that provides effective short-term resistance against rapidly evolving pathogen virulence effectors (Boyd *et al*., 2013). However, the accumulation of resistance QTL and *R* genes conferring quantitative and qualitative effects in cultivars through marker-assisted selection can deliver effective, broad-spectrum and durable resistance, as has been deployed against the rice blast pathogen *Magnaporthe oryzae* (formerly *Magnaporthe grisea*) (Liu *et al*., 2004; Wu *et al*., 2004) and the wheat leaf rust pathogen *Puccinia triticina* (Leach *et al.*, 2001; Boyd *et al.*, 2013).

Psph pathogenicity and effector repertoires

The pathogenic capability of *Psph* on its hosts is determined by the translocation of effector proteins into the cytoplasm of host cells through the hypersensitive response and pathogenicity (Hrp) type III secretion system (T3SS) (Vencato *et al*., 2006). The HrpL alternative sigma factor is an essential regulator of transcription of many of the genes critical for *Psph* pathogenicity (Arnold *et al*., 2009, 2011). HrpL induces multiple genes encoding virulence effectors designated Hrp outer proteins (Hop) and avirulence (Avr) proteins, and those that encode the secretory apparatus (Thwaites *et al.*, 2004; Vencato *et al*., 2006; Arnold *et al.*, 2009, 2011). A *hrp* and *hrp* conserved (*hrc*) gene cluster is conserved amongst several other *P. syringae* pathovars, forming a pathogenicity island containing at least six operons (Alfano *et al.*, 2000). Ortiz-Martín *et al.* (2010a,b) elucidated the positive regulatory roles of HrpL, HrpA and GacA with regard to the T3SS in *Psph*, and the negative regulatory roles of Lon, HrpV and HrpT. They found that HrpL also induces non-T3SS-dependent virulence proteins and represses flagellar function.

Psph avirulence effector genes *hopF1* (eliciting a resistant reaction from bean plants bearing *R* gene *Pse-1*), *hopX1* (corresponding to *Pse-2*), and *hopAR1* (corresponding to *Pse-3*) have been cloned (Tsiamis *et al*., 2000; Stevens *et al*., 1998; and Jenner *et al*., 1991). Previous investigations to identify genes with potential functions in virulence and pathogenicity within *Psph* genomes have examined one isolate (race 6 isolate 1448A collected from *P. vulgaris*; Joardar *et al*., 2005) or two isolates (1448A, and race 1 isolate 1644R collected from *Vigna radiata*; Baltrus *et al*., 2012). Vencato *et al*. (2006) identified 27 candidate effector genes within the genome of *Psph* 1448A by *in silico* searches for common effector motifs. Zumaquero *et al*. (2010) validated the presence of 27 effectors in *Psph* 1448A by analysis of their translocation and/or expression. They revealed quantitative contributions of selected effectors to virulence using single- and double-knockout mutants in competitive index assays.

In a comparative analysis of the effector repertoires of *Psph* 1644R and 1448A, Baltrus *et al*. (2012) reported that eight of the identified effectors were uniquely present within *Psph* 1448A, while 17 were shared. In addition, a gene predicted to encode effector HopV1 was truncated in *Psph* 1644R compared with *Psph* 1448A. Baltrus *et al*. (2012) transferred effectors present in *Psph* 1448A and absent from *Psph* 1644R into the latter to investigate their ability to affect the growth of transformed *Psph* 1644R cultures on the universal halo-blight-susceptible differential cultivar Canadian Wonder. None of the seven effectors tested was found to consistently increase growth.

The current research describes high-quality draft genomes generated for 32 pathogenically and geographically diverse isolates of *Psph* (Chapter 3). Computational genome-wide association genetics approaches are applied to enable identification of (**1**) candidate targets in *Psph* for potential race-nonspecific host resistance, and (**2**) candidate determinants of the broad virulence of *Psph* race 6.

1.3. PHYSIOLOGICAL RESILIENCE

Cold tolerance

Tolerance to the low temperatures (< 15 °C) typical of the period within which common bean can be field-sown in northwestern Europe (mid-May to mid-June) is required for adaptation to the British climate (Austin and MacLean, 1972a; Drijfhout *et al*., 1991). Efforts to identify common bean lines with tolerance to low temperatures for use in the development of UK-adapted cultivars were initiated at the NVRS in the 1970s (Austin *et al*., 1970; Austin and MacLean, 1972a,b; Hardwick *et al*., 1978). However, Hardwick and Andrews (1981) reported that different methods developed for the evaluation of cold tolerance (e.g., Austin and MacLean, 1972b; Silbernagel, 1977) were premised on an incorrect assumption that all growth traits of each genotype are affected equally by reductions in temperature. Consequently, these methods provided inconsistent assessments of lines identified by one method to be 'tolerant' (Hardwick and Andrews, 1981).

Austin and MacLean (1972b) described methods for the identification of promising genotypes amongst material of diverse geographic heritage. They estimated the relative growth rates of 11 varieties in controlled-environment experiments by regression of the logarithm of varietal mean plant weight on that of varietal mean seed weight under different constant temperatures (12.5 °C and 20 °C). The hypothesis underlying this method is that varietal differences in relative growth rates will be indicated by the size of residuals from regression. Varieties with consistently large positive residuals were considered tolerant to the suboptimal temperature. Hardwick and Andrews (1981: 206) highlighted that this method for evaluating cold tolerance "rests on the assumption that seed weight is a true measure of plant weight at the start of growth". However, only moderate goodness of fits were obtained when the analysis was applied to measurements collected from plants grown under day/night temperature regimes of 25/20 °C (R = 0.421) and 20/15 °C (R = 0.3) (Hardwick and Andrews, 1981). Improved fits were obtained when embryonic axis weight, rather than seed weight, was used as a regression variate. This modified method thus accounts for possible genotypic differences in the efficiency with which cotyledonary reserves are mobilised (Drijfhout *et al*., 1991), which otherwise could result in the misclassification of genotypes as tolerant.

Hardwick and Andrews (1981) concluded that cold-tolerant genotypes could be identified by evaluation of the effect of temperature on the distribution of dry weight over the stem and leaves; that is, promising genotypes are those which exhibit a lesser reduction in stem weight relative to leaf weight under suboptimal temperatures.

Silbernagel (1977) identified three cold-'tolerant' common bean genotypes, which were selected for their rapid seedling emergence under low temperature. However, these lines performed poorly using the modified methodology proposed by Hardwick and Andrews (1980, 1981), who found that this phenotype did not ensure improved performance at later stages of growth.

Drijfhout *et al*. (1991) described a simpler method than that of Hardwick and Andrews (1980, 1981) for the identification of genotypes with a lesser growth reduction under a suboptimal temperature, which also corrects for the effects of genotypic differences in seed weight. However, reliable methods such as these require evaluation under controlled-environment conditions, and under both suboptimal and optimal temperatures with replication (Drijfhout *et al*., 1991). This poses a considerable challenge to the evaluation of large plant populations to enable genetic mapping.

Lines classified as cold-'tolerant' using the unmodified method of Austin and MacLean (1972b) were evaluated under field conditions by Hardwick *et al*. (1978). Their findings demonstrated that fluctuating temperatures and other conditions in the field preclude reliable and reproducible assessments of genotypic differences in response to low temperatures. However, a subsequently evaluated landrace of northern European heritage, SOA-BN, exhibited promising levels of cold tolerance and potential drought tolerance in controlled-environment and/or field experiments (R. C. Hardwick, personal communication; Dodd and Taylor, 1991, unpublished data). Unfortunately, this line was introduced to the NVRS Phaseolus collection towards the end of the period within which funding was available for the UK haricot bean breeding programme. However, Dr John Taylor and colleagues crossed SOA-BN with multiple-disease-resistant cultivar Edmund to initiate development of a recombinant inbred population. The population has been advanced from generations F_3 and F_4 to generations F_7 and F_8 in the current research (APPENDIX 1). Chapter 4 describes characterisation and provisional genetic mapping of morphological and

developmental traits related to physiological resilience using this population at F_6 and F_7.

In research to dissect the genetic basis of cold tolerance in a wild relative of common bean, Woronuk *et al*. (2010) identified transcripts that were differentially expressed in both *P. vulgaris* and *Phaseolus angustissimus* under exposure to low temperatures (7 °C day/5 °C night for 72 h) compared with non-stressed controls. They found a *CYP72A14*-like gene belonging to the cytochrome P450 superfamily of monooxygenases to be down-regulated in *P. vulgaris* and up-regulated in *P. angustissimus* under cold stress. Members of this superfamily have been implicated in plant responses to abiotic stress based on up-regulation triggered by reactive oxygen species (ROS) and lipid metabolism in *Arabidopsis* spp. (Narusaka *et al*., 2004). Woronuk *et al*. (2010: 102) suggested that down-regulation in common bean "may indicate either inappropriate regulation in response to ROS or early damage to the gene regulatory mechanism and degradation of transcripts of this gene." Similarly, Vijayan *et al*. (2011) identified transcripts encoding cytochrome P450s, heat shock proteins and DNA-binding proteins that were unique to *P. angustissimus* expressed sequence tags (ESTs) compared with *P. vulgaris* ESTs in a subtraction suppression hybridisation (SSH) experiment involving optimal- and low-temperature treatments.

Drought tolerance

Numerous studies have revealed intersecting regulatory mechanisms for plant responses to cold and drought stresses (e.g., Seki *et al*., 2001, 2002, 2003; Narusaka *et al*., 2004; and reviewed in Guy, 2003; and Shinozaki and Yamaguchi-Shinozaki, 2006). Is has been suggested that plant adaptations to cold stress may be pleiotropic effects of drought tolerance mechanisms that evolved earlier in plants (Guy, 2003).

Drought severely constrains global common bean yield. Much of the world's production occurs in regions where water is limiting, such as in the rain-fed production zones of Latin America and Africa (Beebe *et al.*, 2013); 7 million metric tonnes of the more than 23 million metric tonnes of dry beans produced globally are grown in Latin America and Africa (Broughton *et al.*, 2003). The semi-arid highlands of Mexico, for instance, constitute the largest drought-prone production zone, comprising over 1 million hectares of cultivated common beans (Beebe *et al.*, 2013). Yields of below 0.4 t ha^{-1} in dry years in this zone are between 20% and 25%

of those achieved in countries with large-scale and intensive agricultural systems, such as in the USA, Canada and parts of Brazil (Beebe *et al.*, 2013). Similarly, in sub-Saharan Africa, water deficit has been estimated to cause annual dry bean seed yield losses of 396,000 tonnes (Wortmann *et al.*, c2004).

Unlike their wild relatives, cultivated varieties of common bean are generally susceptible to drought stress and few genetic sources of tolerance have been identified (Cortés *et al.*, 2012b, 2013). The development of cultivars that are sufficiently and consistently high-yielding under drought stress is therefore critical for improving food security. Varieties that require less water can also help to mitigate the adverse environmental impacts of intensive irrigation.

Compared with other predominantly self-pollinating plant species, relatively high levels of genetic diversity and phenotypic variability exist within the common bean species (Gepts *et al.*, 2008). This is thought to be attributable to its subdivision into two discrete genepools as a result of dual domestication events in the Andes and Mesoamerica, its multiple secondary centres of diversity in Africa and Asia, and the attendant evolution of different subpopulations to adapt to specific environmental conditions (Cortés *et al.*, 2011; Cortés *et al.*, 2013; Blair *et al.*, 2013).

Collections of wild common bean genotypes adapted to drought-prone environments have been identified as potentially valuable sources of allelic variation for drought tolerance traits (Cortés *et al.*, 2012b, 2013). In an ecological diversity study to assess the drought tolerance of 104 wild common bean accessions in their natural habitats, Cortés *et al.* (2013) observed that wild genotypes occupy more broadly distributed habitats, including more regions where drought is endemic. They suggested that variation in drought tolerance between wild and cultivated genotypes could be exploited to dissect the genetic basis of this trait and underlying physiological traits (e.g., photosynthetic rate, leaf water potential, water-use efficiency, moisture retention capacity, osmotic potential), and to enable QTL analysis and marker-assisted breeding for drought tolerance. However, the necessity to avoid the yield reduction characteristic of wild genotypes has been highlighted.

Genetic studies by White *et al.* (1994a,b) showed that a genetic background adapted to local environmental conditions is critical for the expression of drought tolerance in common bean. Parental material selected independently in dry highland Mexico and in Colombia that exhibited drought tolerance in their native habitats performed poorly in terms of yield and combining ability in the other environment.

The value of parental lines selected as sources of drought tolerance therefore depended on their yield in the given environment. Accordingly, even if alleles or QTL conditioning drought tolerance are present, their phenotypic expression may be suppressed where their effect is outweighed by the component of local adaptation (Miklas *et al.*, 2006; Beebe *et al.*, 2013).

Genetic transformation to introduce alleles governing drought tolerance into locally adapted cultivars might provide an expedient approach. However, given that drought tolerance is a quantitative trait governed by complex gene networks, transgenic approaches focusing on individual alleles may be limited in their utility. Molecular breeding for enhanced tolerance, through the accumulation of marker-linked functional alleles in locally adapted lines, could therefore offer the most effective means of achieving stable expression across different environments (Miklas *et al.*, 2006).

1.4. GENETIC MARKER DEVELOPMENT IN COMMON BEAN

The development and use of genetic markers for common bean improvement is not as advanced as for economically and traditionally more important crops, such as cereals (Galeano *et al.*, 2009a, and references therein). Several types of polymerase chain reaction (PCR)-based multi- and single-locus markers have nonetheless been developed for common bean (Becerra-Velásquez and Gepts, 1994; Beebe *et al.*, 2000, 2001; Blair *et al.*, 2012).

Efforts to increase the number of well-distributed single-locus markers, such as microsatellites (simple sequence repeats, SSRs; e.g., Blair *et al.*, 2011, 2012) and single-nucleotide polymorphisms (SNPs; e.g., Ramírez *et al.*, 2005; Hyten *et al.*, 2010; Schmutz *et al*., 2014; Hart and Griffiths, 2015), have recently intensified. These markers have been used for linkage map construction and saturation (Blair *et al.*, 2003, 2012; Galeano *et al.*, 2009a,b, 2011, 2012; Song *et al*., 2015), association mapping (Galeano *et al.*, 2012; Cichy *et al*., 2015; Kamfwa *et al*., 2015; Perseguini *et al*., 2016; Zuiderveen *et al*., 2016), as well as for synteny (Hougaard *et al.*, 2008; Galeano *et al.*, 2009b, 2011; McConnell *et al.*, 2010) and diversity analyses (Blair *et al.*, 2006, 2013; Cortés *et al.*, 2011).

‘Useful’ genetic markers

An ideal genetic marker is highly polymorphic (allowing the geneticist or breeder to exploit natural variation), multi-allelic, evenly distributed throughout the genome, codominant and non-epistatic (de Vienne, 2003a). It should also be amenable to high-throughput and cost-efficient genotyping, providing accurate and reproducible data with a high precision of allele calls (Cortés *et al.*, 2011).

Single-locus markers meet most of the above criteria, although higher costs are associated with SSRs than SNP markers because they are less tractable for high-throughput genotyping (Cortés *et al.*, 2011). SSRs are multi-allelic, however, and therefore more polymorphic than generally bi-allelic SNPs (Galeano *et al.*, 2012). In a study comparing SSR and SNP marker technologies for genotyping in maize (*Zea mays* L.), Jones *et al.* (2007) reported that SNP technologies can nonetheless provide markers that are superior in quality and quantity, and that lower polymorphism can be compensated for by increasing marker number and by detecting SNP haplotype blocks.

SNPs are the most abundant source of sequence variation for polymorphic marker development in any genome (Galeano *et al.*, 2012). Due to their abundance and the increasing accessibility to automated technologies for high-throughput, low-cost SNP genotyping with high precision, SNPs have become a preferred source of markers for several research applications (Cortés *et al.*, 2011). These include high-resolution linkage and association mapping, marker-assisted selection, diversity studies, and positional cloning of target genes (Hyten *et al.*, 2010; Galeano *et al.*, 2012; Blair *et al.*, 2013).

Non-genic SNP discovery and genotyping

Hyten *et al.* (2010) reported an efficient and cost-effective system for high-throughput SNP discovery and genotyping in common bean. In response to the challenges and costs associated with SNP discovery using next-generation sequencing (NGS) in the absence of a published whole genome sequence for common bean, they constructed a multi-tier reduced-representation library (mtRRL) using a series of restriction digestions to reduce genome complexity. Sequencing of the mtRRL exploited the differential advantages of two NGS platforms; sequences

generated by the Roche 454-FLX pyrosequencing system (one run)—for their greater read-length and utility for assembling reference sequences—were coupled with those obtained from the Illumina Genome Analyzer II (one run), which produces shorter reads but with much greater read depth at considerably lower cost. These shorter reads were aligned with the assembled 454-derived reference sequences to detect 3,487 predicted SNPs. A total of 827 of the 1,050 candidate non-genic SNPs assayed produced successful Illumina GoldenGate [2] (Oliphant *et al.*, 2002) genotyping assays (79%).

Hyten *et al.* (2010) also highlighted the disadvantages of alternative SNP discovery methods based on transcriptome sequencing data for the development of genome-wide SNP markers at loci of unknown function for use in linkage mapping. Namely, fewer SNPs will be identified in coding DNA sequences (CDS) because they are more conserved than non-CDS. Moreover, the success rates of SNP genotyping assays will be reduced due to the hybridisation of primers or probes both to the SNP-containing gene sequence and to any conserved paralogues. They suggested that SNP assays based on cDNA sequences would also suffer from decreased success rates because of the interference of intronic regions with primer hybridisation.

Genic and expressed sequence tag-based SNP detection

Gene- and expressed sequence tag (EST)-based SNP discovery, however, has been an important source of markers in plant genetics due to their potential causative role in governing traits of interest (Cortés *et al.*, 2011; Galeano *et al.*, 2012); the saturation of genetic maps with functional markers is necessary for their useful application in marker-assisted selection.

[2] This PCR-based SNP detection platform provides for highly multiplexed genotyping through the use of two differentially dye-labelled, allele-specific forward primers and one locus-specific reverse primer that contains a unique address sequence that targets a correspondingly coded microbead. Following hybridisation of dye-labelled PCR products to microbeads with complementary unique address codes, fluorescence signal detection enables allele calling for each SNP.

Using the competitive allele-specific PCR (KASP) SNP detection platform (Cuppen, 2007; LGC Genomics),[3] Cortés *et al.* (2011) evaluated SNP marker diversity at 84 gene-based (including genes involved in drought response) and ten non-genic loci in a diversity panel of 70 genotypes representing the Andean and Mesoamerican genepools and previously used as parents of mapping populations. Although these markers were useful for inter-genepool comparisons, the potential application of these markers to breeding was limited by low polymorphism within each genepool. The diversity analysis found that SSRs were more powerful for within-genepool race differentiation. In view of the species' complex evolutionary history and subdivision, Cortés *et al.* (2011) proposed that an "arsenal" of complementary genetic markers, combining both SNPs and SSRs, should be deployed to investigate genetic diversity for the purposes of linkage and association mapping in common bean.

Blair *et al.* (2013) developed a high-throughput SNP marker system for common bean using a 768-feature Illumina GoldenGate assay based on tentative orthologous legume gene sequences to evaluate parental polymorphisms and to perform a diversity analysis, drawing upon diverse common bean genotypes. A total of 736 of the 768 genic SNPs assayed produced successful GoldenGate assays (96%). These are comparable in number but considerably higher in proportion to those successfully scored in the GoldenGate assay of Hyten *et al.* (2010), possibly resulting from the different sources of SNPs exploited by these two studies.

Consistent with the findings of Cortés *et al.* (2011), however, the 736 SNP markers scored by Blair *et al.* (2013) were most useful for distinguishing between genepools, owing to lower intra-genepool polymorphism, and therefore for inter-genepool genetic map saturation. Both the GoldenGate analysis of Blair *et al.* (2013) and the KASP assay of Cortés *et al.* (2011) had comparable missing data rates (2.7% and 2.5%, respectively). Cost per data point was lower with the GoldenGate assay,

[3] This PCR-based technology uses competitive allele-specific oligonucleotide extension, which occurs in the absence of a single-base mismatch in the SNP position at the 3′ end of one of two competing allele-specific forward primers, enabling the Taq polymerase to elongate the strand. This is combined with signal generation by fluorescence resonance energy transfer (FRET) that occurs when one of two dye-labelled oligonucleotide reporter probes—each of which corresponds to the tail sequence of one of the two allele-specific primers—is removed from its complementary quencher and binds to the complementary sequence generated by the common reverse primer following the second round of PCR. This system has the advantage that the allele-specific forward primers do not require labelling; only the reporter probes, which form part of the universal master mix that can be produced in bulk, are fluorescently labelled.

although this platform requires evaluation of a full 768-feature array of SNPs per assay. The KASP platform offers more flexibility in the number of SNPs assayed, making the cost per genotype lower. Accordingly, Blair *et al.* (2013) suggested that, following SNP validation using either platform, KASP is most appropriate and cost-effective for more targeted SNP genotyping, using a small subset of markers, for the purpose of genetic mapping. The GoldenGate platform, by contrast, is better suited to the genotyping of large collections of diverse germplasm with large numbers of SNP markers, to be applied in association studies. Both technologies have been shown to deliver robust and efficient SNP validation and genotyping with reproducible data generation (Chen *et al.*, 2010; Cortés *et al.*, 2011; Blair *et al.*, 2013).

Intronic SNP assays using single-strand conformational polymorphism

To capture higher levels of gene-based polymorphism, Galeano *et al.* (2012) explored intronic variation in common bean ESTs to develop 313 new intron-based SNP markers at target genes. They used the single-strand conformational polymorphism (SSCP) technique and Sequenom MassARRAY platform to assess marker polymorphism between parental lines and in the inter-genepool DOR364 (Mesoamerican parent) × G19833 (Andean parent) RIL population. Fifty-three of these markers were mapped on an integrated linkage map for the RIL population. To overcome the problem of non-overlapping sets of QTL that is associated with data collected from populations derived from different parental crosses and grown in different experimental environments, Galeano *et al.* (2012) contrasted marker–trait correlations across two environments, drought and irrigation. They highlighted the utility of this type of experimental design for marker-assisted selection, an important goal of which is to select for the stable expression of desired traits across different environments.

SNP analysis using the SSCP technique was identified as a simple and inexpensive gel-based alternative to other approaches used for EST-derived SNP marker conversion in common bean, which involve specialised and/or expensive equipment and reagents (Galeano *et al.*, 2009b). The SSCP method is based on differences of DNA conformation that result from sequence variations in single-stranded DNA fragments, which can be visualised and detected by migration in non-denaturing polyacrylamide gel electrophoresis. Although this approach hinges on

SNPs that cause conformational changes, it is estimated that close to 100% of sequence variations are detectable in PCR amplicons of < 200 bp (de Vienne, 2003b). This detection rate decreases as the fragment size increases. As Galeano *et al.* (2009b) highlight, however, NGS technologies can be utilised to identify SNPs between parental lines in advance, and SNP-flanking PCR primers subsequently designed to amplify < 200-bp fragments for subsequent SSCP analysis.

Enzyme-based SNP marker development and mapping

Two studies investigating syntenic relationships among legumes used cleaved amplified polymorphic sequence techniques (CAPS and derived CAPS) to convert common bean EST-based polymorphisms into SNP markers (Hougaard *et al.*, 2008; McConnell *et al.*, 2010). These methods require enzyme-specific restriction digestions, however, and can be expensive for evaluating large SNP resources (Cortés *et al.*, 2011).

Galeano *et al.* (2009a) reported an alternative enzyme-based method for EST-derived SNP discovery and marker development in common bean based on EcoTILLING, an adaptation of the mutation detection approach used in Targeting Induced Local Lesions IN Genomes, for detecting polymorphisms of target genes in natural populations (Comai *et al.*, 2004). This method involved cleavage of single-base-pair heteroduplex mismatches by digestion with the single-strand-specific endonuclease CEL I to evaluate and map SNP markers that were developed based on an *in silico* analysis of common bean ESTs for SNP discovery by Ramírez *et al.* (2005).

One notable SNP locus placed on the inter-genepool linkage map of DOR364 × G19833 by Galeano *et al.* (2009a) (BSNP 28, located on linkage group B3), following CEL I-based segregation analysis, shows homology to a drought tolerance-related gene coding for *S*-adenosylmethionine decarboxylase in lima bean (*Phaseolus lunatus* L.). This enzyme contributes to transcriptional and post-transcriptional regulation of gene expression (Hu *et al.*, 2005), and functions in plant sensing of changes in the abiotic environment, including increased water deficit stress (Micheletto *et al.*, 2007).

The EcoTILLING approach has a number of advantages, including that SNPs occurring singly within, and situated in the middle of, PCR amplicons were detectable on agarose gels following digestion. The agarose gel system was not

effective, however, for detecting polymorphism where two or more SNPs were contained within amplicons. This limitation can nonetheless be overcome by using NGS data from parental lines for appropriate SNP marker design.

Additionally, CEL I can be isolated from celery (*Apium graveolens* L.) stalks, producing a nuclease of equivalent digestion efficiency to a commercially supplied version of the enzyme (Galeano *et al.*, 2009a). Although this method does not provide for the high-multiplex and high-throughput assay of SNPs offered by oligonucleotide pool assay platforms (such as KASP and GoldenGate), SNP genotyping by CEL I digestion does not require proprietary technology or expensive, specialised equipment, resulting in a competitive cost per genotype of USD 1.26 (Galeano *et al.*, 2009a). Galeano *et al.* (2009a) also reported that the error rate of CEL I SNP detection assays is no higher than that of direct sequencing of parental alleles.

Genotyping-by-sequencing

The recent release of the whole-genome sequence of common bean (Schmutz *et al.*, 2014) makes genotyping-by-sequencing (GBS) a simple and cost-efficient alternative. GBS enables high-throughput, simultaneous SNP discovery and genotyping, and involves the construction of reduced-representation libraries using restriction digestions to reduce genome complexity (Elshire *et al.*, 2011). The main advantage of GBS over the method described by Hyten *et al.* (2010), which used Illumina GoldenGate assays, is that it generates tens to hundreds of thousands of genotyped SNP markers, which, in addition, are not confined to non-genic regions. Genotyping data generated by high-density SNP-marker GBS assays can be used for more accurate mapping of loci implicated in governing traits of interest (McClean *et al.*, 2011).

A modified GBS approach for use in soybean (*Glycine max*) was recently described by Sonah *et al.* (2013). This approach enabled detection of 10,120 high-quality SNPs, 39.5% of which were in genic regions and 52.5% of these were in CDS. They reported a 98% validation rate using Sanger sequencing at a set of randomly selected SNPs. Additionally, Sonah *et al.* (2013) found that SNP calls could be increased by nearly 40%, with a more than doubled depth of coverage, by the use of selective primers to further reduce genome complexity, resulting in higher throughput and lower cost per genotype. This approach therefore offers an attractive

option for efficient SNP calling in an entire collection of RILs in a bi-parental population.

Hart and Griffiths (2015) described a GBS protocol optimised for common bean, generating 7,530 high-quality informative SNPs. They defined a 974-kb mapping interval for the *By-2* allele conferring resistance to *Bean yellow mosaic virus* (BYMV) by association analysis. The utility of GBS for high-resolution linkage mapping of disease resistance, morphological, developmental and reproductive traits was therefore investigated in the current research using the SOA-BN × Edmund recombinant inbred population.

1.5. RESEARCH AIMS

In 2011, efforts began at the University of Warwick Crop Centre to revive advanced breeding material from a UK common bean genetics programme, which ran for nearly two decades at the same experimental research farm near Wellesbourne (National Vegetable Research Station or NVRS) until 1990 (Hardwick, 1988). The current project aims to advance the revived research by establishing molecular breeding capability for adapting common bean to the UK climate. It advances an NVRS-derived population of RILs to enable genetic mapping and marker-assisted breeding for adaptive traits that UK farmers would require of a home-grown common bean.

A long-term goal, beyond the scope of this project, is to provide UK farmers with a novel legume break crop that could improve soil health and aid grass weed control (Miller *et al*., 2002; Brisson *et al*., 2010). Impact would also be achieved by ultimately establishing a food production and supply chain for common beans in the UK, providing consumers with a nutritious source of home-grown vegetable protein. These broader impact aims are being pursued in recognition of the need to introduce more novel crops into UK farming to create more resilient farming systems that are less dependent on a small number of crops. The incorporation of new crops into rotations can also reduce the incidence of major pests and diseases currently occurring in the UK, and mitigate the environmental and economic costs of the repeated use of specific agrochemistry that accompanies limited crop rotations (Collier, 2014).

Three complementary aims were pursued in this research: **(1)** construct a high-resolution linkage map for a bi-parental RIL population to enable genetic mapping of quantitative resistance to halo blight, a locally and globally important disease constraint on common bean production; **(2)** apply genome-wide association genetics to identify candidate virulence determinants in race 6 of *Pseudomonas syringae* pathovar *phaseolicola*, which is undetected by the major Phaseolus bean *R* genes conferring qualitative resistance to halo blight; and **(3)** develop molecular breeding capacity to enable selection of physiological and morphological traits required for low-input production of common bean in the UK.

Chapter 2

High-resolution linkage mapping of resistance in common bean to *Pseudomonas syringae* pv. *phaseolicola* race 6 using genotyping-by-sequencing (GBS)

2.1. INTRODUCTION

Effective prevention of halo blight of common bean (*Phaseolus vulgaris* L.) caused by *Pseudomonas syringae* pathovar *phaseolicola* (*Psph*) requires the deployment of resistance sources that are appropriate for controlling regionally important races of the pathogen. The deployment of durable resistance to the globally prevalent *Psph* race 6 (which is undetected by any of the known race-specific resistance [*R*] genes; Taylor *et al*., 1996a,b; Miklas *et al*., 2009, 2011, 2014) is critical to controlling this broadly virulent pathogen, which continues to threaten common bean production worldwide.

Durable disease resistance is achieved when a pathogen population is unable to overcome host resistance through shifts in its range of virulence (Taylor *et al*., 1996b). Durable resistance does not 'break down', despite the deployment of *R* genes in a large plant population, for a protracted period, over a large area in which environmental conditions would otherwise favour disease development (Iyer-Pascuzzi and McCouch, 2007).

Taylor *et al*. (1996b) proposed the deployment of combinations of race-specific and quantitative, possibly race-nonspecific, resistance in common bean cultivars as a means of achieving potentially durable resistance to halo blight. The global pervasiveness of *Psph* is thought to be attributable to its ability to spontaneously mutate to virulence through mutations in pathogen effector genes, leading to the loss of avirulence (Avr) proteins. This has been observed to enable the pathogen to change from race 4 to race 2 such that it is not recognised by the race-specific *R* gene *Pse-3* (Teverson, 1991; Taylor *et al*., 1996a,b; Jackson *et al*., 2000; Pitman *et al*., 2005). Research enabling marker-assisted selection for quantitative, potentially race-nonspecific, resistance could therefore make an important contribution to breeding efforts to establish durable halo blight resistance.

In the current research, inoculation with *Psph* races 6 and 1 caused localised red-brown necrotic lesions at the point of maximum (infiltrative) inoculation on common bean plants with disease scores close to 1.0 or 2.0 ('highly resistant' or 'resistant'; Figure 2.1A–B). Vegetative growth of these plants appeared to be unaffected by the pathogen treatment. The red-brown necrotic reaction is symptomatic of localised cell death or the 'hypersensitive response', a biochemical

defence mechanism that arrests pathogen infection of, and development on, the host. Infiltrative inoculation is required to induce this confirmatory hypersensitive response in resistant lines. Plants with disease scores of 2.0 reportedly exhibit 'field resistance' and, in the absence of infiltrative inoculation, develop neither red-brown necrosis nor traces of water-soaking under field conditions where disease pressure is high (J. D. Taylor, personal communication). Plants with disease scores close to 4.0 or 5.0 ('susceptible' or 'fully susceptible') developed water-soaked lesions beyond the area of maximum inoculation and distributed at random over the leaf undersurface, indicative of bacterial infection (Figure 2.1D–E; Taylor *et al*., 1978). These symptoms were caused by the non-infiltrative spray method of inoculation described below, which simulates the dissemination of inoculum occurring during rainfall and wind in the field (Arnold *et al*., 2011). Susceptible and fully susceptible plants also exhibited severe inhibition and distortion of vegetative growth at the shoot apical meristem. Plants with disease scores close to 3.0 ('slightly susceptible') exhibited symptoms intermediate between those displayed by 'resistant' and 'susceptible' plants (Figure 2.1C), with more moderate inhibition of vegetative growth at the shoot apical meristem.

The objective of this research is to characterise and map the quantitative resistance to *Psph* race 6 derived from common bean accession PI 150414, which was deployed in haricot cultivar Edmund, by applying genotyping-by-sequencing (GBS) to a segregating recombinant inbred population. Phenotypic variation within the population in response to separate inoculations with races 6, 1 and 3 was evaluated under glasshouse conditions. These phenotyping and genotyping data were combined to enable localisation of quantitative trait loci (QTL) governing halo blight resistance derived from PI 150414.

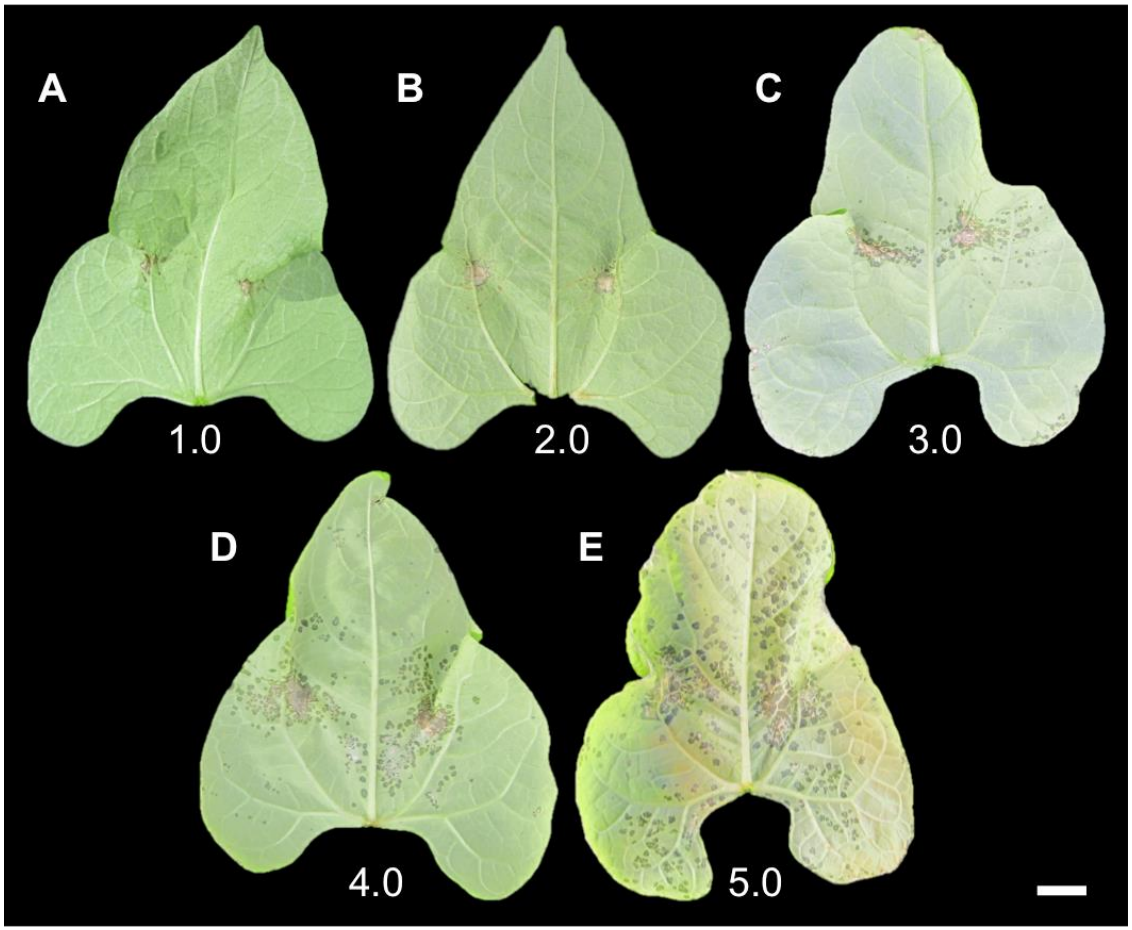


Figure 2.1. Interaction phenotypes exhibited by different *Phaseolus vulgaris* genotypes ten days after inoculation with *Pseudomonas syringae* pv. *phaseolicola* race 6 or race 1, including (**A**) Red Mexican UI-3 (disease score = 1.0; highly resistant), (**B**) Edmund (disease score = 2.0; resistant), (**C**) SOA-BN × Edmund recombinant inbred line 113 (disease score = 3.0; slightly susceptible), (**D**) SOA-BN × Edmund recombinant inbred line 22 (disease score = 4.0; susceptible), and (**E**) SOA-BN (disease score = 5.0; fully susceptible). Bar = 1 cm.

2.2. MATERIALS AND METHODS

Plant material

A bi-parental mapping population was developed for this study from a cross between SOA-BN and Edmund (hereafter referred to as the SE population; APPENDIX 1). Phenotype and genotype data were generated for 80 F_6 and F_7 SE recombinant inbred lines (RILs) generated by the single-seed descent method. The SE population was initiated and developed up to generations F_3 and F_4 by Dr John Taylor and colleagues at Warwick Horticultural Research Institute (HRI) near Wellesbourne. The author of this thesis subsequently advanced and curated the population over four growing seasons from 2012 to 2015; F_7 and F_8 seed are now available in the Phaseolus collection maintained at the University of Warwick Crop Centre.

SOA-BN is a brown-seeded accession of northern European origin and is susceptible to all of the *Psph* races described by Teverson (1991). Researchers at the National Vegetable Research Station (NVRS; subsequently Warwick HRI) identified this accession as a source of cold tolerance in laboratory and field experiments, and of earliness of maturity and potential drought tolerance over three seasons of field evaluations in the UK (R. C. Hardwick, personal communication; Dodd and Taylor, 1991, unpublished data). Edmund was registered as a commercial white haricot (navy bean) variety with multiple resistance to seed-transmitted diseases, including halo blight (possessing *Pse-3* and quantitative resistance from PI 150414), *Bean common mosaic virus* (BCMV) and anthracnose (caused by *Colletotrichum lindemuthianum*). This variety was developed at the NVRS as part of a UK breeding programme (Conway *et al*., 1982).

A second mapping population of 70 inbred lines derived from SOA-BN and Edmund in the NVRS breeding programme (hereafter referred to as the JDT lines) was also evaluated for resistance to *Psph* race 6 (two replicates) and race 1 (one replicate) using the methods described below. The JDT lines included 55 F_7 inbreds selected for halo blight resistance at F_2 and 15 F_{13} inbreds selected for resistance at F_8 (Dodd and Taylor, 1991, unpublished data). All plant material was obtained or developed from the Phaseolus collection held at Warwick Crop Centre, Wellesbourne, UK.

Inoculation

Seeds of each SE RIL were sown in 7-cm plastic pots (one seed per pot) containing Levington M2 compost with a vermiculite covering, arranged in a completely randomised design. Plants were grown under glasshouse conditions at Warwick Crop Centre, Wellesbourne, UK (52°12.54′N, 1°36.23′W), with a heating temperature of 18 °C and a venting temperature of 20 °C, and with supplementary lighting (400W high-pressure sodium lamps) to provide a 16-h photoperiod. Experiments were replicated in time to give up to five pseudo-replicates per RIL per bacterial isolate. Replication level varied across SE RILs due to seed limitations and variable germination; the SE population was in development up to October 2014 and so it was necessary to conserve seeds of some RILs to enable investigation of other traits within the three-year project.

Several bacterial isolates were used separately in SE population experiments, including *Psph* race 6 isolate 716B, race 1 isolate 725A, and race 3 isolate 1301A. All isolates were obtained from the collection held at Warwick Crop Centre, Wellesbourne, UK (APPENDIX 1). Original race designations were confirmed based on the interaction phenotypes exhibited by the halo blight host differential lines listed in Taylor *et al*. (1996a), except for *Phaseolus acutifolius* accession 1072, which was unavailable. In view of the virulence of isolates 716B and 725A on differential line ZAA 12 (A43) possessing *Pse-2* (conferring resistance to races 2, 3, 4, 5, 7, 8 and 9; Miklas *et al*., 2011), and the avirulence of isolate 1301A on lines possessing *Pse-3* (conferring severe hypersensitive resistance to races 3 and 4; Taylor *et al*., 1996a), accession 1072 was not required for race confirmation. Isolates for inoculation were cultured on King's Medium B (King *et al.*, 1954) for 24–48 h at 25 °C. Inocula were prepared by suspending bacteria in sterile high-purity water and adjusting the suspensions to 10^8–10^9 colony-forming units ml^{-1} (methods used to determine the inoculum concentration are described in APPENDIX 2).

Bean seedlings were inoculated as described by Taylor *et al*. (1996a). Specifically, ten-day-old seedlings with newly unfolded unifoliate leaves (first pair of true leaves) were inoculated by spraying the bacterial suspension onto both leaves (~ 5 ml per plant) with a DeVilbiss atomiser nozzle attached to a CEJN Series 208 air gun, powered by a motorised air compressor pump. Leaves were inoculated by holding the spray nozzle 1–2 cm from the undersurface and spraying the inoculum in

two ~ 0.5-cm-diameter areas either side of the leaf midrib, thereby infiltrating the tissue through the stomata. The whole of the leaf undersurface was then sprayed with the inoculum until completely wet, with the spray nozzle held 10–15 cm from the leaf (Teverson, 1991; Taylor *et al.*, 1996a). One seedling of each parental line was mock-inoculated with sterile high-purity water according to this method.

Inoculated and mock-inoculated seedlings were kept separately under conditions of high humidity by sealing them in 780-mm × 620-mm autoclavable polypropylene bags (40 μm thick) for 48 h (10–15 plants per bag), before being returned to glasshouse conditions. One seedling of each parental line was included in each polypropylene bag as resistant and susceptible controls. Seedlings of Red Mexican UI-3 (possessing *Pse-1*) were included as highly resistant controls in race 1 experiments. Seedlings of Guatemala 196-B, Tendergreen and ZAA 12 (A43) (possessing *Pse-3*) were included as severe hypersensitive resistant controls in race 3 experiments.

Phenotypic assessments were made ten days after inoculation according to the five-point infection scale defined by Innes *et al*. (1984: 308) and illustrated in Figure 2.1, with decimal values assigned to plants considered to be intermediate between two integer scores: "(**1**) red-brown necrotic reaction in the area of maximum inoculation either side of the leaf midrib (highly resistant); (**2**) red-brown necrotic reaction with a trace of water-soaking (resistant); (**3**) some necrosis but more extensive water-soaking largely confined to the area of maximum inoculation (slightly susceptible); (**4**) small water-soaked lesions (< 1 mm diameter) distributed at random over the leaf undersurface (susceptible); and (**5**) larger water-soaked lesions (1–3 mm diameter) distributed at random over the leaf undersurface (fully susceptible)." High-resolution photographs were taken to document the interaction phenotype exhibited by each RIL plant in each experiment.

Reaction to *Psph* race 3 was assessed qualitatively to enable mapping of *Pse-3*, conferring race-specific resistance to races 3 and 4. Inbred lines that exhibited a severe hypersensitive response were predicted to possess *Pse-3*. This severe form of hypersensitivity, triggered by races 3 and 4 in the presence of *Pse-3*, is phenotypically distinct from the localised necrotic reaction observed on plants predicted to possess resistance derived from PI 150414 only. It is characterised by the induction of "a more generalized hypersensitive response with red-brown necrotic lesions distributed over the entire undersurface of the inoculated leaves of

resistant cultivars" (Taylor *et al*., 1996a: 472). Lines which did not exhibit a severe hypersensitive response but which exhibited a resistant phenotype comparable to that derived from PI 150414 were noted.

All statistical analyses were performed in R version 3.3.0 (R Core Team, 2016). The mean reaction of each RIL to each race was calculated from the disease severity scores assigned to replicate plants. Fixed- and random-effects variance components were estimated by restricted maximum likelihood (REML) analysis using the mixed-effects modelling package lme4 version 1.1-12, and tested using lmerTest version 2.0-32.

Differences between the disease scores of parental genotypes were determined by one-way analysis of variance. The assumption of equal variances was assessed by *F*-tests. Means, 95% confidence intervals, and least significant differences (LSDs) were calculated and plotted. Pearson's product-moment correlation coefficients and simple linear models for the relationship between race 6 and race 1 mean disease scores were calculated for each population.

Genetic markers

The SE population was characterised for genome-wide parental SNP variation using GBS with ApeKI restriction enzyme-based complexity reduction (Elshire *et al*., 2011). High-quality genomic DNA of each SE inbred, parental line and JDT line was extracted from young trifoliolate leaves according to a CTAB protocol adapted from Afanador *et al*. (1993) (APPENDIX 3). Resistant and susceptible bulk samples for GBS consisted of equal quantities of DNA of each of 11 and five JDT lines that were resistant and susceptible, respectively, to both *Psph* race 6 and race 1. Double-stranded DNA was quantified using the Qubit dsDNA BR Assay Kit with a Qubit 2.0 Fluorometer (Invitrogen, Waltham, USA) to confirm that a per-sample concentration of $\geq$ 50 ng μl^{-1} was obtained. DNA quality was evaluated using a NanoDrop Spectrophotometer (NanoDrop, Wilmington, USA) and gel electrophoresis; samples were electrophoresed on 0.8% agarose gels stained with GelRed (Biotium, Fremont, USA) and visualised using an ultraviolet transilluminator to confirm that they were of high molecular weight (Figure A4.1, APPENDIX 4). A trial digestion of genomic DNA of the two parental lines and of eight randomly selected lines with 5 U per sample of the restriction enzyme EcoRI (Roche, Basel, Switzerland) for 2 h at 37 °C followed by 20 min at 65 °C was performed to confirm suitability for GBS.

Digested samples were electrophoresed and visualised on a 1% agarose gel (Figure A4.2, APPENDIX 4). Samples were submitted to the Institute of Biotechnology Genomic Diversity Facility, Cornell University, USA, for library construction and sequencing. A GBS library was prepared according to previously described protocols (Elshire *et al*., 2011), as optimised for common bean (Hart and Griffiths, 2015). The GBS library was sequenced twice (48-plex) by 101-cycle single-end sequencing on two lanes of an Illumina HiSeq 2500 instrument (Illumina, San Diego, USA), generating 101-bp single-end reads.

Sequencing reads were processed using the TASSEL 5.0 GBSv2 Discovery/Production Pipeline for species with a reference genome (Bradbury *et al*., 2007; Glaubitz *et al*., 2014; see full bibliographic reference for an overview of the pipeline used in the current research). High-quality sequence tags containing a single ApeKI cut site were identified and stored in a local database, specifying a minimum Phred-scaled quality score of 20 (which equates to a 1 in 100 probability of an incorrect base call) and a maximum number of stored Kmers of 10^9. Stored sequence tags were aligned to the first chromosome-scale version of the assembled common bean reference genome (*Phaseolus vulgaris* V1.0, genotype G19833; Schmutz *et al*., 2014) using the Burrows-Wheeler Aligner version 0.7.12 (BWA; Li and Durbin, 2009). SNPs within aligned tags were called according to default pipeline parameters. Qualifying SNPs were those with minor allele frequency (MAF) ≥ 0.01 and for which ≥ 0.8 of samples were genotyped. A comparatively relaxed minimum MAF was applied because a number of breeding lines of interest (which did not contribute to the pedigree of the SE population) were included in the GBS plate. SNPs with < 15-fold average read depth across all samples were discarded. The resultant Variant Call Format (VCF) file was imported into TASSEL 5.0 (Bradbury *et al*., 2007) for conversion into a matrix of homozygous and heterozygous genotypes. Monomorphic and uninformative markers were removed for a total of 1,248 SNPs for linkage map construction.

The *I*/*Pse-3*-linked cleaved amplified polymorphic sequence (CAPS) marker targeting SNP ss715641188 developed by Bello *et al*. (2014) was included within this finalised set of markers. PCR primers flanking the SNP were used to amplify a 311-bp fragment prior to digestion with restriction endonuclease TaqI. PCR reaction mixtures contained 10 mM of Tris-HCl (pH 8.3), 50 mM of KCl, 1.5 mM of $MgCl_2$, 0.001% gelatin, 0.2 mM of each dNTP, 0.4 μM of each primer, 20–100 ng of

genomic DNA, and 0.75 U of Taq DNA polymerase (1X REDTaq ReadyMix; Sigma Aldrich, Gillingham, UK), in a final volume of 25 µl. Amplifications were performed on a GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Foster City, USA) as described by Bello *et al*. (2014), consisting of a denaturation step at 94 °C for 5 min; followed by 35 cycles of 94 °C for 30 s, 58 °C for 30 s, and 72 °C for 30 s; and a final extension step at 72 °C for 5 min.

PCR amplicons were electrophoresed 1% agarose gels stained with GelRed (Biotium, Fremont, USA) and visualised using an ultraviolet transilluminator to confirm their presence and expected size. A 5-µl aliquot of each amplicon was digested with 1 µl of FastDigest TaqI enzyme and 1X FastDigest Green Buffer (Thermo Scientific, Loughborough, UK) in a final volume of 20 µl. Reactions were incubated at 65 °C for 15 min and electrophoresed and visualised on 2% agarose gels. TaqI cuts the resistant allele, yielding two bands of 201 and 110 bp (Bello *et al*., 2014).

In addition, RNA-Seq was performed to facilitate the identification of polymorphisms between the transcriptomes of the parental lines and a halo blight-resistant RIL (Capulet) derived from this parentage. Total RNA was isolated from young primary-leaf tissue of these three lines using the RNeasy Plant Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Libraries were prepared using the Illumina TruSeq RNA Library Preparation Kit (Illumina, San Diego, USA) according to the manufacturer's instructions and validated on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). Libraries were pooled at 10 nM and sequenced by paired-end sequencing on one lane of an Illumina Genome Analyzer IIx (Illumina, San Diego, USA), generating 70-bp paired-end reads.

The pipeline depicted in Figure 2.2 (taken from Bolser, 2016, with permission) was used to identify sequence variants between the common bean reference genome (Schmutz *et al*., 2014) and the transcriptomes of lines SOA-BN, Edmund and Capulet.

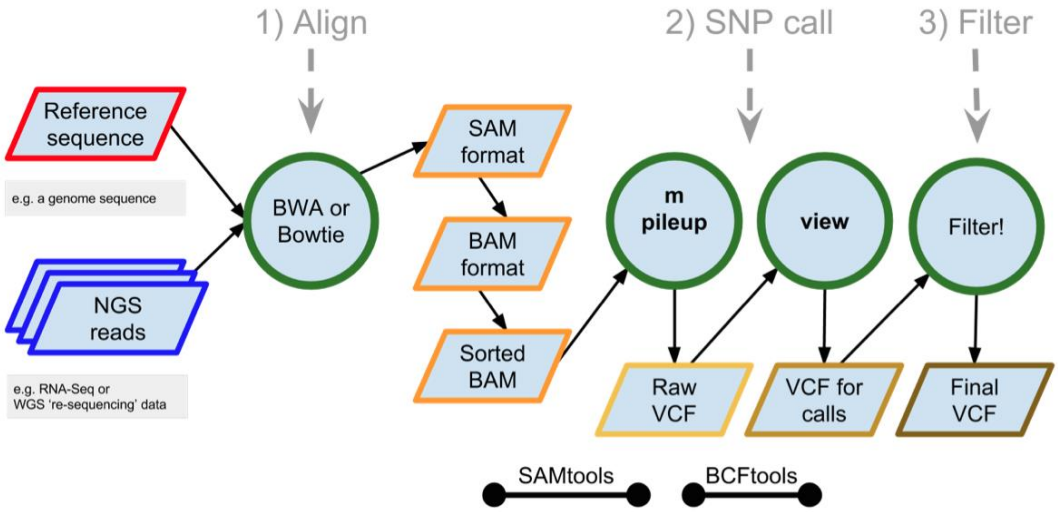


Figure 2.2. Flow diagram of the pipeline and programs used to call sequence variants between the *Phaseolus vulgaris* reference genome and the transcriptomes of *P. vulgaris* lines SOA-BN, Edmund and Capulet. BWA: Burrows-Wheeler Aligner. SAM: Sequence Alignment/Map format. BAM: Binary Alignment/Map format. VCF: Variant Call Format. Taken from Bolser (2016), with permission.

Specifically, sequence quality metrics for raw and cleaned reads were evaluated using FastQC version 0.11.2 (Andrews, 2010). Illumina adapter sequences, PCR primer fragments and low-quality bases were removed using Trimmomatic version 0.32 (Bolger *et al*., 2014) in paired-end and palindrome modes, which enable reliable detection of adapter read-through. Trimming settings informed by FastQC analysis were applied as follows: a maximum read–adapter sequence alignment mismatch count of 2 bases; a palindrome trimming accuracy score threshold of ≥ 30; a simple trimming accuracy score threshold of ≥ 10; removal of the first 13 bases from the beginning of reads; a sliding window size of four bases with an average Phred-scaled sequence quality threshold of ≥ 15; removal of low-quality bases from the beginning and end of reads, with a Phred-scaled quality threshold of ≥ 3; and a read length threshold of ≥ 36 bases.

Trimmed reads were aligned to the indexed common bean reference genome using Bowtie 2 version 2.2.4 (Langmead and Salzberg, 2012). Output files in the Sequence Alignment/Map (SAM) format were converted into the Binary Alignment/Map (BAM) format, and BAM files were sorted and indexed using SAMtools version 0.1.19 (Li *et al*., 2009) to enable efficient processing. Single-nucleotide variants (SNVs) and sequence insertions/deletions (INDELs) were predicted for each sample line against the reference genome, and filtered based on a Phred-scaled quality threshold of ≥ 20 and a depth threshold of ≥ 10-fold coverage

using the SAMtools command-line utilities BCFtools. Parental polymorphisms within candidate genes at which halo blight-resistant lines Edmund and Capulet share common alleles were identified by comparison of VCF files using custom Perl scripts. These variant alleles were annotated with information about their predicted effects on protein function (e.g., 'loss of function', 'nonsense-mediated decay', 'high', 'moderate', 'low' or 'modifier' putative effect impacts; silent, missense or nonsense mutations; amino acid and codon changes) using SnpEff version 4.1 (Cingolani *et al*., 2012).

Linkage map construction

Linkage map construction was performed with the 1,247 GBS-derived SNP markers and the *Pse-3*/*I*-linked CAPS marker using MapDisto version 2.0.b86 (Lorieux, 2012). Markers were assigned to linkage groups corresponding to the 11 common bean chromosomes using the 'Find linkage groups' command (LOD_{min} = 3.0; r_{max} = 0.24). Markers were ordered according to linkage using the seriation II algorithm and the SARF (sum of adjacent recombination fractions) locus-ordering criterion. The 'AutoRipple' and 'AutoCheckInversions' functions were applied repeatedly until each procedure could not find an improved order. Map distances were calculated using the Kosambi mapping function. Unlinked markers that could not be assigned were placed and ordered on a pseudo-linkage group, which includes markers from sequence tags aligning to reference genomic scaffolds.

QTL analysis

QTL analyses were conducted using R/qtl version 1.39-5 (Broman *et al*., 2003) in R version 3.3.0 (R Core Team, 2016). One-dimensional, single-QTL genome scans were performed using the multiple imputation mapping method with a scan interval of 1 cM (imputations = 100; error probability = 0.001). Although more computationally intensive, the multiple imputation method provides the most robust QTL analyses compared with alternative mapping methods implemented in R/qtl (Broman and Sen, 2009). Namely, standard interval mapping using maximum likelihood estimation, Haley-Knott regression and extended Haley-Knott regression perform poorly at markers with missing genotype data, potentially resulting in spurious QTL (Broman and Sen, 2009). Using the multiple imputation approach,

genotypes at markers with missing data are imputed based on the observed data for each genotyped individual, and a mean LOD score at each marker is derived from those obtained by the imputations. With 99.7% of genotype data present in the SE map, missing data was not a general concern. In view of the availability of sufficient processing power, however, multiple imputation was chosen as the preferred method for one-dimensional, single-QTL genome scans and for fitting and refining QTL models.

For high-resolution genetic maps, alternative methods are required to enable timely completion of the computationally expensive permutation tests that determine significance thresholds for two-dimensional, two-QTL genome scans (Broman and Sen, 2009). Accordingly, two-dimensional genome scans were performed by Haley-Knott regression with a scan interval of 1 cM (error probability = 0.001) to enable assessment of evidence for two-QTL models involving additive or interacting loci. Provided that the genetic map is not affected by substantial missing genotype data (resulting from a selective genotyping strategy, for example), Haley-Knott regression enables more robust QTL analyses than standard interval mapping by maximum likelihood estimation in R/qtl (Broman and Sen, 2009).

Permutation tests (1,000 permutation replicates) were performed for each one- and two-dimensional genome scan to identify loci with LOD scores greater than or approaching estimated significance thresholds. A LOD score is the $\log_{10}$ ratio of the likelihood that there is a QTL at a given location to the likelihood that a QTL is not present anywhere in the genetic map (Broman and Sen, 2009). Computationally expensive permutation tests for two-dimensional genome scans were performed on Ubuntu Server LTS 14.04 (64-bit, 384 GB of RAM, and two 16-core processors); each 1,000-permutation test was divided into five simultaneous tasks (200 permutation replicates per task) and the results were combined for evaluation of evidence for QTL with additive or interacting effects.

The QTL support interval was defined as the interval in which the LOD score is within 1.8 units of its maximum (Broman and Sen, 2009). The physical boundaries of each QTL were determined based on the positions of the closest markers flanking the 1.8-LOD support interval in the common bean reference genome. Interval size at major-effect loci was further defined by manual inspection of the haplotypes of RILs recombinant for the parental alleles near the QTL. QTL models were fitted and refined using the multiple imputation method (scan interval =

1 cM; imputations = 100; error probability = 0.001) to further test the significance of each locus and to derive the percentage of phenotypic variation explained.

Candidate resistance protein orthology and network prediction

Arabidopsis thaliana proteins with the highest amino acid identity to candidate *P. vulgaris* proteins potentially conferring quantitative resistance to halo blight were identified by BLASTP comparison using The Arabidopsis Information Resource (TAIR) submission form (Berardini *et al*., 2015). Potential orthologues identified in *A. thaliana* were submitted for Bayesian probabilistic functional gene network inference using the AraNet version 2 web server (Lee *et al*., 2015) to predict possible defence response pathways that may mediate the mechanism governing quantitative resistance to halo blight. Simple Interaction File (SIF) outputs generated for interactome visualisation were imported into Cytoscape version 3.4.0 (Shannon *et al*., 2003) for formatting, and the 'Prefuse Force-Directed Layout' with unweighted edges was applied.

2.3. RESULTS

Interaction phenotypes

The halo blight-resistant parent Edmund had a mean disease score of 2.1 following separate inoculations with *Psph* race 6 (sample standard deviation, SD = 0.12; n = 28) or race 1 (SD = 0.11; n = 31). In contrast, the susceptible parent SOA-BN had a mean disease score of 4.4 (SD = 0.44; n = 25) following inoculation with race 6, and a race 1 mean disease score of 4.2 (SD = 0.27; n = 34). Means, 95% confidence intervals and LSDs for the parental lines are summarised in Figure 2.3. All Edmund plants exhibited a severe hypersensitive cell death response following inoculation with race 3 (n = 13) due to presence of race-specific *R* gene *Pse-3*, whereas none of the SOA-BN plants exhibited this distinct host response (n = 14). The control variety for resistance conferred by *Pse-1*, Red Mexican UI-3, had a mean disease score of 1.0 following inoculation with race 1 (SD = 0.06; n = 14).

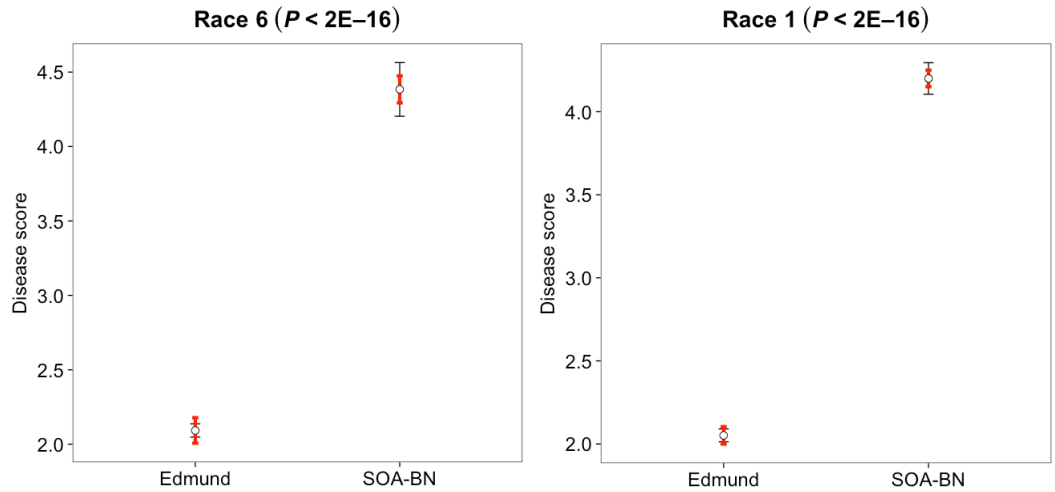


Figure 2.3. Summary of interaction phenotypes exhibited by *Phaseolus vulgaris* parental genotypes Edmund and SOA-BN following inoculation with *Pseudomonas syringae* pv. *phaseolicola* race 6 (left) or race 1 (right). Means (white circles), 95% confidence intervals (thin black lines) and least significant differences (thicker red lines) are shown. Significance values obtained by one-way analyses of variance are indicated above each plot.

In the SE population, halo blight resistance segregated in a bimodal distribution following separate inoculations with *Psph* race 6 and race 1 in glasshouse experiments (Figure 2.4A), with no replication effects detected (Table A5.1, APPENDIX 5). For both races, approximately 50% of the SE inbreds had a mean disease score < 3.0 ('highly resistant' to 'resistant'/'slightly susceptible'), and

approximately 50% had a mean disease score > 3.0 ('slightly susceptible' to 'fully susceptible'). Previously reported genetic evidence suggested that the resistance from PI 150414 (deployed in Edmund) is simply inherited (Taylor *et al*., 1978, and references therein). Taken together, these findings indicate that a single resistance factor with major effect is segregating (1:1) against a background of one or more minor-effect genes.

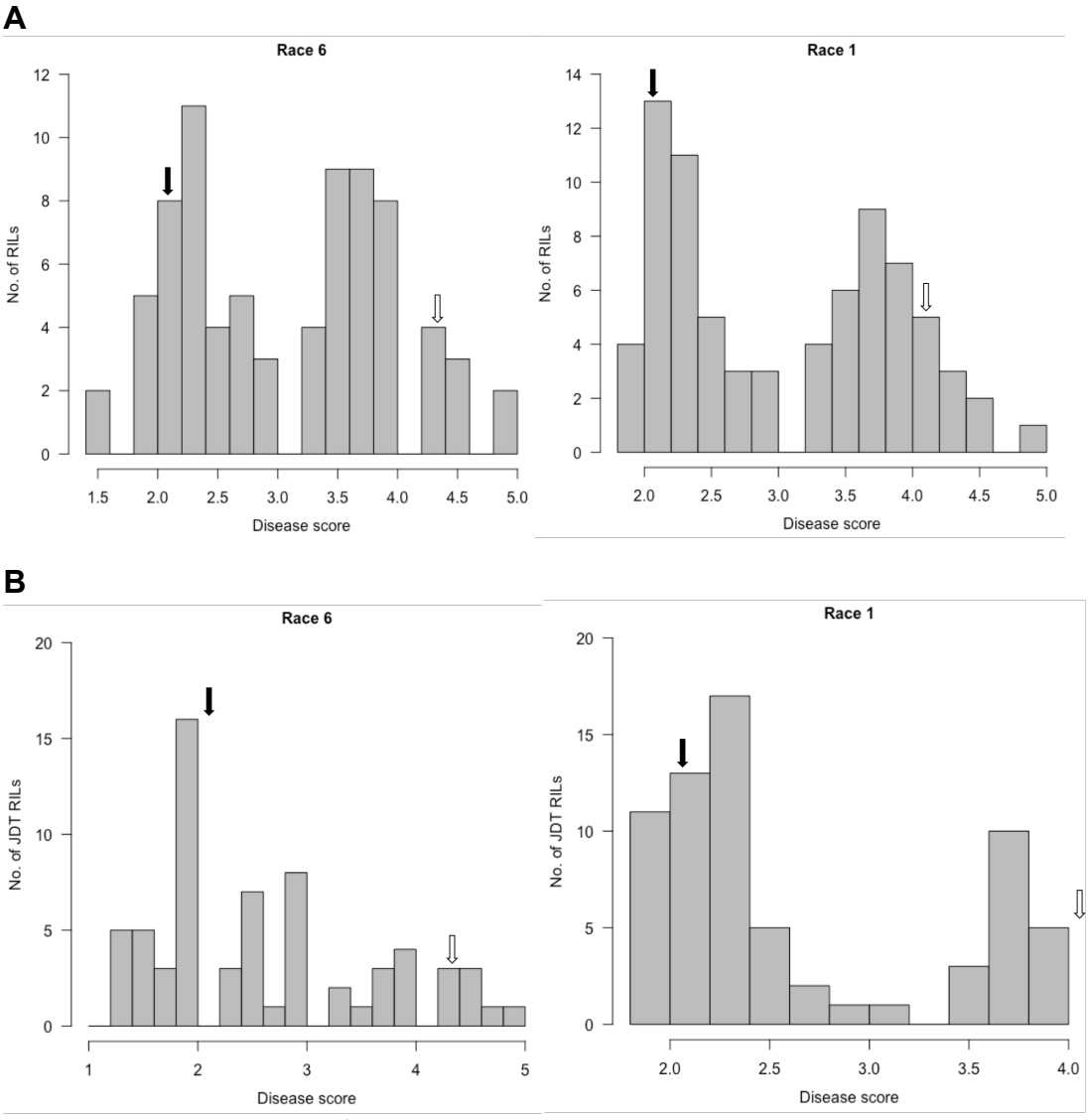


Figure 2.4. Distribution of interaction phenotypes in glasshouse experiments within two recombinant inbred mapping populations of *Phaseolus vulgaris* following separate inoculations with *Pseudomonas syringae* pv. *phaseolicola* race 6 (left) and race 1 (right), including (**A**) SOA-BN × Edmund lines and (**B**) previously developed JDT lines. The phenotype scale ranges from highly resistant (score 1.0) to fully susceptible (score 5.0). Black arrows denote the mean disease score for Edmund and white arrows denote the mean disease score for SOA-BN.

A higher proportion of JDT lines had mean disease scores < 3.0 (n = 40; 60.6%) than had scores > 3.0 (n = 18; 27.3%) following inoculation with race 6 (Figure 2.4B). Eight lines had race 6 mean disease scores of 3.0 (12.1%). Similarly, 48 JDT lines had disease scores < 3.0 (70.6%), one had a score of 3.0 (1.5%), and 19 had scores > 3.0 (27.9%) following inoculation with race 1. A bias towards resistance was expected for this population because it was intentionally selected for halo blight resistance during earlier generations of inbreeding (Dodd and Taylor, 1991, unpublished data). The apparent phenotypic segregation suggests that resistance derived from PI 150414 is not explained by a single recessive factor in this population.

A Pearson product-moment correlation coefficient showed a positive correlation between race 6 and race 1 mean disease scores in both the SE (R = 0.86; n = 75; P < 2.2E−16) and JDT (R = 0.81; n = 64; P = 7.2E−16) inbred populations (Figure 2.5). Low disease scores following inoculation with one race were correlated with low scores following inoculation with the other race. For instance, SE inbreds with race 6 mean disease scores ≤ 3.0 also had race 1 mean disease scores ≤ 3.0. These findings support previous reports that the resistance derived from PI 150414 confers quantitative, potentially race-nonspecific, resistance to halo blight (Taylor *et al*., 1996b).

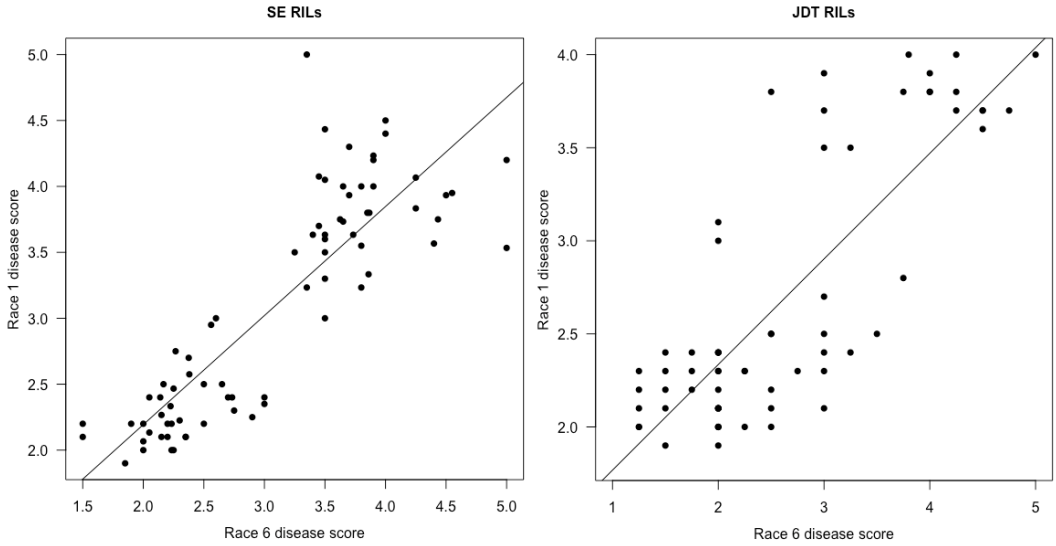


Figure 2.5. Scatterplots showing positive correlations between mean reactions to *Pseudomonas syringae* pv. *phaseolicola* race 6 and race 1 in both the SOA-BN × Edmund (SE; left; $R = 0.86$; $n = 75$; $P < 2.2E{-}16$) and JDT (right; $R = 0.81$; $n = 64$; $P = 7.2E{-}16$) populations of *Phaseolus vulgaris* recombinant inbred lines. SE population simple linear model: $y = 0.53953 + 0.82698x$. JDT population simple linear model: $y = 1.20047 + 0.56736x$.

QTL analysis

One-dimensional, single-QTL genome scans revealed a major-effect locus at the telomeric end of the short arm of chromosome Pv04 governing resistance to *Psph* race 6 (LOD = 29.20; $P = 0$) and to race 1 (LOD = 34.56; $P = 0$) in the SE population (Figure 2.6). SNP markers located at 0.6 Mb and 1.1 Mb on Pv04 define a 500-kb mapping interval for this locus. A two-dimensional genome scan detected individual additive minor-effect QTL on Pv08 at 58.4 Mb (LOD = 3.93; $P = 0.014$) and Pv03 at 43.6 Mb (LOD = 3.50; $P = 0.051$) for reaction to race 1 (Figures 2.7 and 2.8). That is, each of these minor-effect QTL contributes individually (non-epistatically) to the race 1 reaction phenotype, irrespective of the genotype at the major-effect QTL on Pv04. Neither additive nor epistatic minor-effect QTL were detected for reaction to race 6. QTL models for reactions to races 6, 1 and 3 are summarised in Table 2.1.

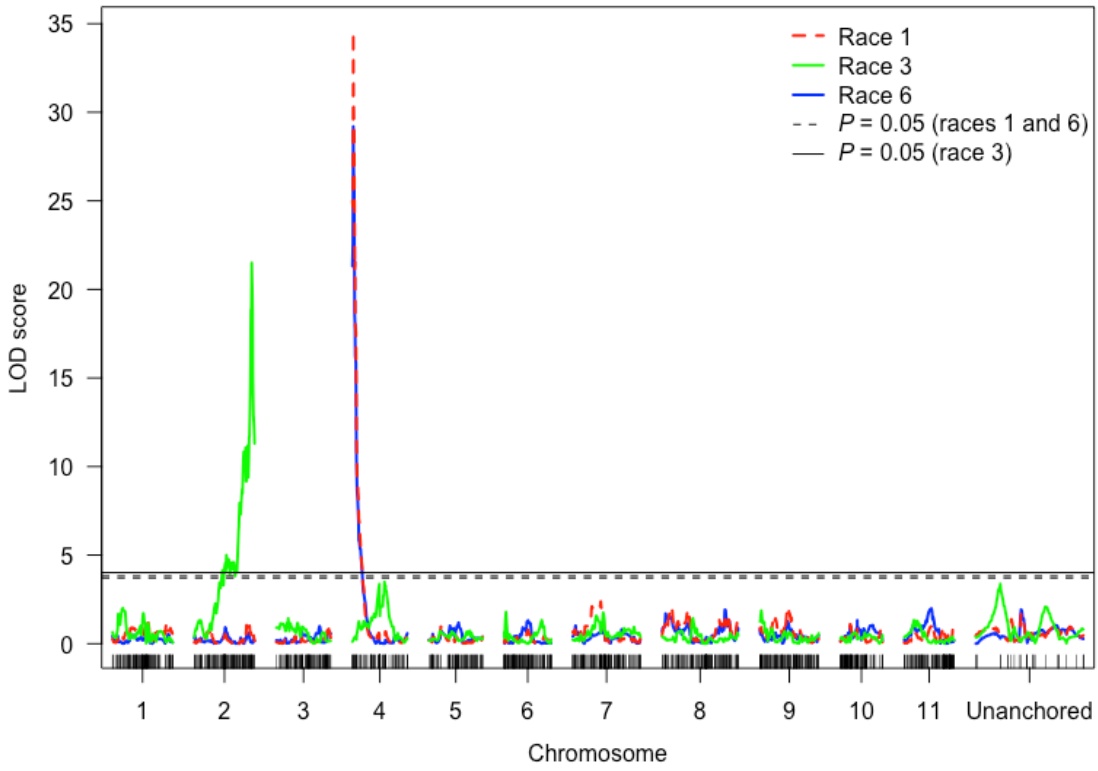


Figure 2.6. LOD profiles for reactions of *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred lines following inoculation with *Pseudomonas syringae* pv. *phaseolicola* races 6 and 1 (scored quantitatively) and race 3 (scored qualitatively) obtained by single-QTL genome scans using the multiple imputation method.

The severe hypersensitive response to race 3 (conferred by *Pse-3*) cosegregated with the *I* gene locus on Pv02 (LOD = 21.50; *P* = 0) (Figure 2.6; Table 2.1), which is consistent with previous observations (Teverson, 1991; Miklas *et al*., 2011, 2014). Linked markers show a pattern of segregation distortion in favour of the *I*/*Pse-3* allele conferring resistance. This distorted transmission has been observed in several common bean RIL populations (Miklas *et al*., 2011, 2014; Bello *et al*., 2014), with preferential selection of gametes possessing the allele conferring resistance proposed as the most likely cause (Miklas *et al*., 2014).

Similar results were obtained by one-dimensional, single-QTL standard interval mapping by maximum likelihood estimation, by Haley-Knott regression and by extended Haley-Knott regression. The mapping interval on chromosome Pv04 was further confirmed by manual inspection of the haplotypes of resistant and susceptible bulks at interval-defining markers. There are 38 predicted genes in the 500-kb mapping interval on chromosome Pv04 of the common bean reference genome (Table 2.2; Schmutz *et al*., 2014).

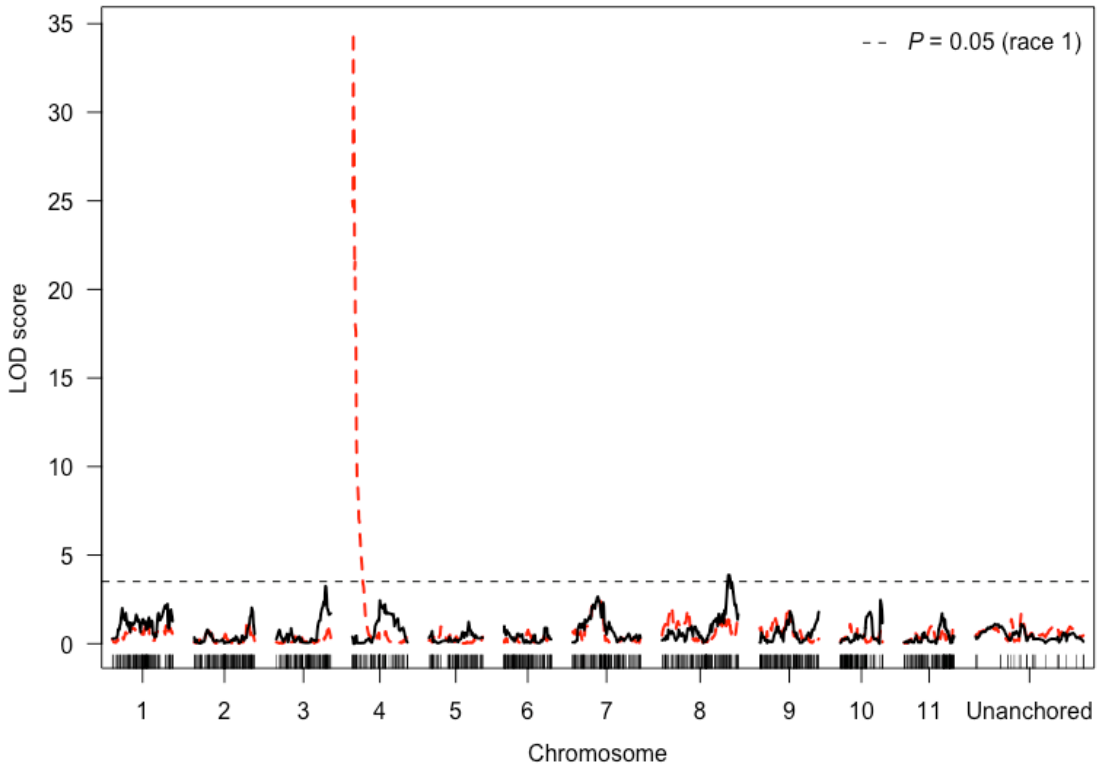


Figure 2.7. LOD profiles for reactions of *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred lines following inoculation with *Pseudomonas syringae* pv. *phaseolicola* race 1 with (black solid line) and without (red dashed line) the marker closest to the major-effect QTL on chromosome 4 included as a covariate.

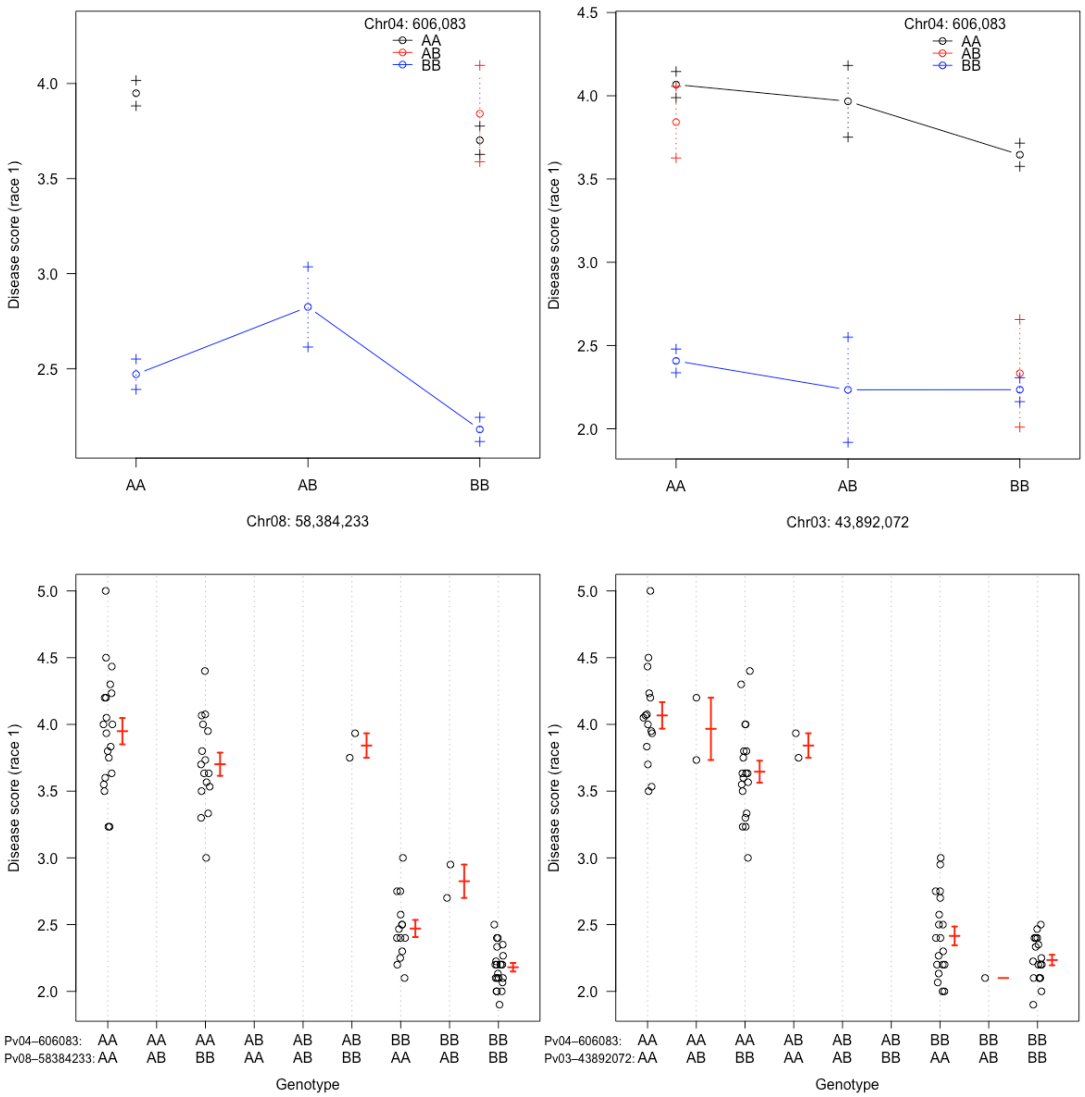


Figure 2.8. Additive relationships between the major-effect QTL on chromosome 4 and the minor-effect QTL on chromosome 8 (left) and chromosome 3 (right) conditioning resistance to *Pseudomonas syringae* pv. *phaseolicola* race 1 in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population. AA: SOA-BN (susceptible) parental genotype; BB: Edmund (resistant) parental genotype.

Table 2.1. Single- and multi-QTL models fitted and refined by multiple imputation for reactions to *Pseudomonas syringae* pv. *phaseolicola* races 6, 1 and 3 in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population

Race	Chromosome	Start	End	Size (Mb)	d.f.	LOD	R^2 (%)	Probability (*F*)	Additive effect	Dominance deviation
6	4	599,801	1,104,545	0.50	2	29.16	82.52	0***	−0.77707	0.53997
1	4	599,801	1,104,545	0.50	2	40.27	82.17	< 2E−16***	−0.74757	0.96039
	8	58,092,745	58,844,117	0.75	2	4.07	2.19	2.02E−04***	−0.14516	0.50921
	3	2,163,905	49,869,462	47.71	2	3.40	1.80	8.20E−04***	0.06932	0.29013
3	2	48,289,722	48,533,537	0.24	10	116.61	86.23	< 2E−16***	0.452129	0.215263
	9	32,928,931	33,468,765	0.54	6	73.55	5.44	< 2E−16***	0.008893	−0.178774
	Unanchored	N/A	N/A	N/A	6	71.42	4.75	< 2E−16***	−0.009418	0.169311
	8	1,643,838	2,757,636	1.11	2	57.22	1.88	< 2E−16***	0.013349	−1.143553
	2 × 9	N/A	N/A	N/A	4	59.73	2.22	< 2E−16***	N/A	N/A
	2 × Unanc.	N/A	N/A	N/A	4	70.00	4.33	< 2E−16***	N/A	N/A

Start and end coordinates define the physical boundaries of each QTL, which were determined based on the positions of the closest markers flanking the 1.8-LOD support interval in the common bean reference genome (Schmutz *et al*., 2014). Interval size at major-effect loci was further defined by manual inspection of the haplotypes of RILs recombinant for the parental alleles near the QTL.

Mb: megabase. d.f.: degrees of freedom. LOD: logarithm (base 10) of the odds score. R^2 (%): percentage of phenotypic variation explained by each locus.

***: significant at the 0.001 probability level.

Estimated additive effect sizes equate to half of the difference between the mean phenotype of RILs homozygous for the allele derived from parental line Edmund (allele 'B') and the mean phenotype of RILs homozygous for the allele derived from SOA-BN (allele 'A') ([BB−AA]/2). Negative values indicate a positive allelic contribution from SOA-BN to the phenotype (i.e., an increased disease score), and positive values indicate a positive allelic contribution from Edmund. The direction of the effect is inverted at the major-effect locus on chromosome 2 (*Pse-3*) due to the qualitative scoring method used to evaluate RILs for reaction to *Psph* race 3; a disease score of 1 corresponds to severe hypersensitive resistance (derived from Edmund) and a disease score of 0 corresponds to the absence of severe hypersensitivity (derived from SOA-BN).

Dominance deviation refers to a deviation from additive allelic action resulting from interaction between alleles of a locus.

All statistics were derived from analyses of variance by which loci were omitted one at a time from each QTL model to estimate the support for individual loci and their interactions. These values account for all estimated additive and interactive effects at each locus (Broman and Sen, 2009).

Candidate resistance protein orthology and network prediction

There are 19 potential *A. thaliana* orthologues of the 38 candidate genes located within the Pv04 mapping interval (Table 2.2). Probabilistic functional gene network inference using AraNet version 2 (Lee *et al*., 2015) predicted 13 of these potential orthologues to be indirectly connected to one another by varying degrees of separation via direct connections to other functionally associated genes in *A. thaliana* (Figure 2.9). This may suggest involvement of a cluster of genes in *P. vulgaris* in the same defence response pathway that mediates the quantitative resistance mechanism derived from PI 150414. A possible orthologous gene encoding an RNA-binding protein (RBP; AT3G27700) and three other possible orthologues were not connected to this network (Figure 2.10). Two valid query genes (AT3G27670 and AT4G16195) were not found in AraNet.

Table 2.2. Thirty-eight candidate genes located within the 500-kb mapping interval defined for quantitative resistance to halo blight on chromosome 4 of the *Phaseolus vulgaris* reference genome

Annotated gene ID	Location (Phytozome)	Predicted functional annotation[a]	*Arabidopsis thaliana* homologue (BLASTP)	Polymorphisms[b]
Phvul.004G007500	Chr04: 597278–601006	Protein/domain of unknown function (DUF1421)	AT5G14540 – Protein/domain of unknown function (DUF1421)	
Phvul.004G007600	Chr04: 605953–613920	RNA recognition motif (RRM)-containing protein	AT3G27700 (< 50%) – Zinc finger (CCCH-type) family protein / RNA recognition motif (RRM)-containing protein	6 moderate (missense); 1 low
Phvul.004G007700	Chr04: 623268–626839	Disease resistance protein (nucleotide-binding site–leucine-rich repeat protein with putative coiled-coil N-terminal domain, CC–NLR)	AT3G14470 (< 50%) – Nucleotide-binding adapter shared by APAF-1, Resistance proteins, and CED-4 (NB-ARC) domain-containing disease resistance proteins	
Phvul.004G007800	Chr04: 643296–647667	No annotations	AT5G46260 (< 50%) – Disease resistance protein with Toll/interleukin-1 receptor-like N-terminal domain (TIR–NLR)	
Phvul.004G007900	Chr04: 672127–676700	Disease resistance protein (CC–NLR); pseudogene in reference genome	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008000	Chr04: 678195–680885	Disease resistance protein (CC–NLR); pseudogene in reference genome	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008100	Chr04: 688650–693324	Disease resistance protein (CC–NLR); pseudogene in reference genome	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008200	Chr04: 707064–709398	Disease resistance protein (CC–NLR); pseudogene in reference genome	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008300	Chr04: 734268–740298	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008400	Chr04: 750857–752692	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008500	Chr04: 759676–761499	No annotations	AT5G40400 – Pentatricopeptide repeat (PPR)-containing protein	
Phvul.004G008600	Chr04: 765416–768590	Phospholipid-translocating ATPase	AT3G27870 – ATPase E1-E2 type family	
Phvul.004G008700	Chr04: 792116–795536	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008800	Chr04: 824480–828899	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G008900	Chr04: 838998–842360	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G009000	Chr04: 852372–853227	No annotations	AT3G27670 – Armadillo/β-catenin-like repeat protein	
Phvul.004G009100	Chr04: 853293–853832	Leucine-rich repeat (LRR)-containing protein	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G009200	Chr04: 856940–868944	Protein/domain of unknown function (DUF3730)	AT3G27670 – Armadillo/β-catenin-like repeat protein	
Phvul.004G009300	Chr04: 873737–877733	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G009400	Chr04: 892792–894806	E3 UBIQUITIN-PROTEIN LIGASE 6 (UPL6)	AT3G17205 – E3 UBIQUITIN-PROTEIN LIGASE 6 (UPL6)	

Table 2.2. Continued

Annotated gene ID	Location (Phytozome)	Predicted functional annotation[a]	*Arabidopsis thaliana* homologue (BLASTP)	Polymorphisms[b]
Phvul.004G009500	Chr04: 896498–901075	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G009600	Chr04: 906782–909895	No annotations	AT3G27870 – ATPase E1-E2 type family	
Phvul.004G009700	Chr04: 945326–955745	No annotations	AT3G27670 – Armadillo/β-catenin-like repeat protein	1 moderate (missense); 1 low
Phvul.004G009800	Chr04: 966792–971111	Disease resistance protein (CC–NLR)	AT3G14470 (< 50%) – NB-ARC domain-containing disease resistance protein	
Phvul.004G009900	Chr04: 976122–979478	PPR-containing protein	AT5G40400 – PPR-containing protein	
Phvul.004G010000	Chr04: 982581–992206	Protein/domain of unknown function (DUF3730)	AT3G27670 – Armadillo/β-catenin-like repeat protein	
Phvul.004G010100	Chr04: 999365–1001208	LATERAL ORGAN BOUNDARIES DOMAIN-CONTAINING PROTEIN 25 (LBD25)	AT3G27650 – LATERAL ORGAN BOUNDARIES DOMAIN-CONTAINING PROTEIN 25 (LBD25)	
Phvul.004G010200	Chr04: 1006379–1006744	Plant self-incompatibility protein S1	AT4G16195 (< 50%) – Plant self-incompatibility protein S1	
Phvul.004G010300	Chr04: 1010279–1010683	Plant self-incompatibility protein S1	AT4G16195 (< 50%) – Plant self-incompatibility protein S1	
Phvul.004G010400	Chr04: 1015240–1019248	Mitogen-activated protein kinase kinase (MAPKK); Serine/threonine-protein kinase STE7	AT5G40440 – Mitogen-activated protein kinase kinase 3 (MKK3)	1 low
Phvul.004G010500	Chr04: 1019873–1024420	WD40 REPEAT-CONTAINING PROTEIN L2DTL	AT3G27640 (< 50%) – Transducin/WD40 repeat-like superfamily protein	2 modifier
Phvul.004G010600	Chr04: 1025421–1029236	CLEAVAGE/POLYADENYLATION FACTOR 1 (CLP1)	AT5G11010 – Pre-mRNA cleavage complex II protein family	1 modifier
Phvul.004G010700	Chr04: 1035453–1036874	Cyclin-dependent protein kinase inhibitor (SMR2-related)	AT5G40460 – Protein/domain of unknown function	
Phvul.004G010800	Chr04: 1053217–1056560	Cobalamin (vitamin B12) biosynthesis (CbiX) protein	AT1G50170 – SIROHYDROCHLORIN FERROCHELATASE B	1 modifier
Phvul.004G010900	Chr04: 1057410–1060763	No annotations	AT1G12700 (< 50%) – RNA PROCESSING FACTOR 1 (RPF1, a PPR); ATP binding; nucleic acid binding; helicases	1 modifier
Phvul.004G011000	Chr04: 1084516–1089721	CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE 42 (CRK42); salt stress response/antifungal; protein tyrosine kinase; serine/threonine protein kinase	AT5G40380 – CYSTEINE-RICH RECEPTOR-LIKE PROTEIN KINASE 42 (CRK42)	
Phvul.004G011100	Chr04: 1093516–1103190	Polyribonucleotide nucleotidyltransferase	AT5G14580 – Polyribonucleotide nucleotidyltransferase	
Phvul.004G011200	Chr04: 1104370–1113109	Nicotinamide adenine dinucleotide phosphate (NADP)-specific isocitrate dehydrogenase (IDH)	AT5G14590 – Isocitrate/isopropylmalate dehydrogenase (IDH/IMDH) family protein	

[a]Predicted protein functional annotation in the *Phaseolus vulgaris* reference genome (genotype G19833) as identified by Phytozome version 11.0 (PANTHER/Pfam/EuKaryotic Orthologous Groups/National Center for Biotechnology Information) (Goodstein *et al*., 2012).
[b]SnpEff-predicted impacts on protein function of quality-filtered single-nucleotide variants (SNVs) shared by *Phaseolus vulgaris* lines Edmund and Capulet and different from SOA-BN and reference genotype G19833.

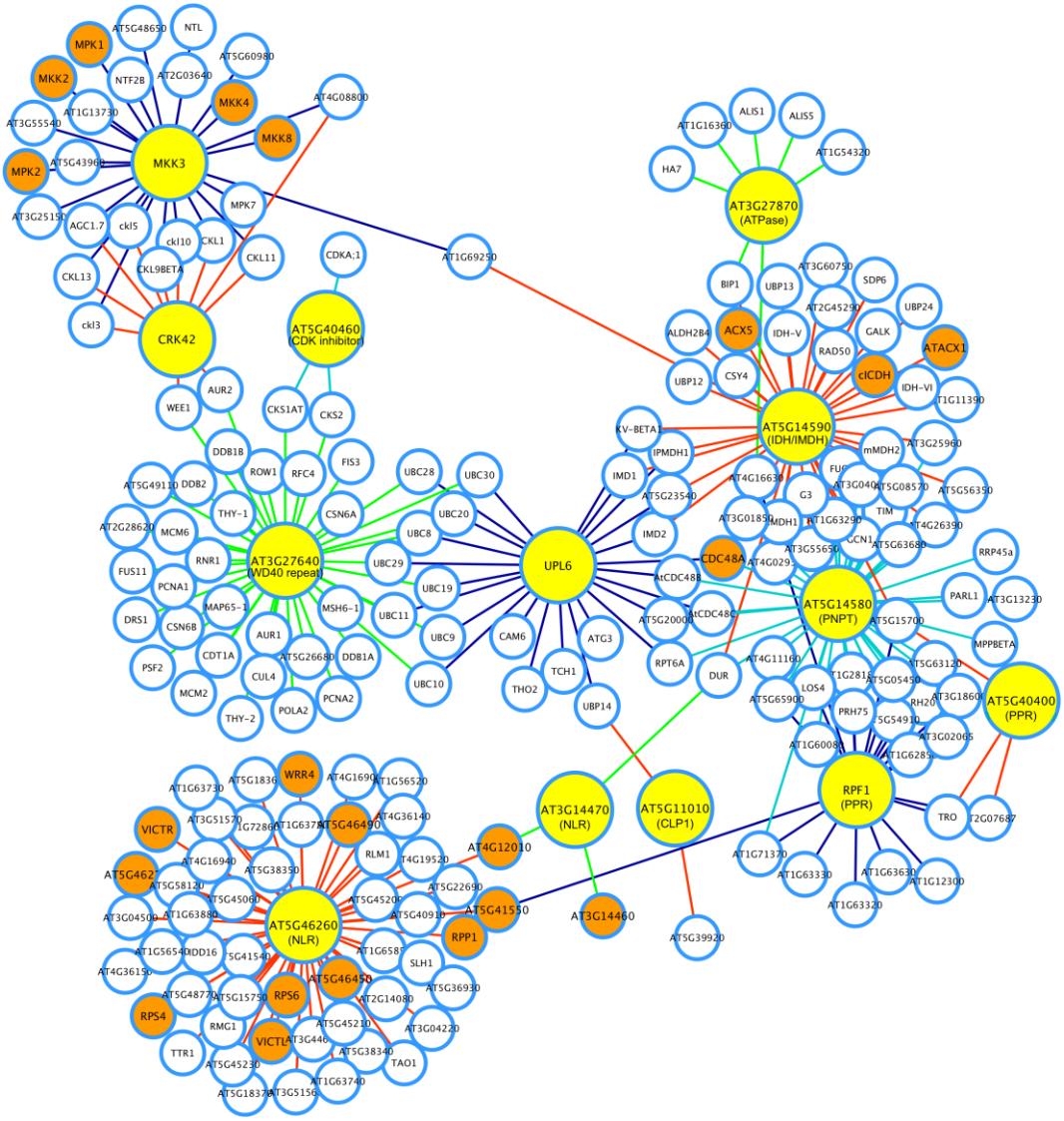


Figure 2.9. Bayesian probabilistic functional gene network showing predicted connections between 13 potential *Arabidopsis thaliana* orthologues (larger yellow nodes) of candidate genes located within the halo blight mapping interval on chromosome 4 of the *Phaseolus vulgaris* reference genome. Other functionally associated genes with defence-related functional annotations are depicted as orange nodes. Edges emanating from potential orthologues are coloured for visual distinction of connections. ATACX1/ACX5: ACYL-CoA OXIDASES 1 and 5; CDC: cell division cycle protein; CDK: cyclin-dependent protein kinase; CLP1: pre-mRNA CLEAVAGE AND POLYADENYLATION FACTOR 1; CRK: cysteine-rich receptor-like protein kinase; IDH/cICDH/IMDH: (cytosolic) isocitrate/isopropylmalate dehydrogenase; MKK: mitogen-activated protein kinase kinase; MPK: mitogen-activated protein kinase; NLR: nucleotide-binding site–leucine-rich repeat protein; PNPT: polyribonucleotide nucleotidyltransferase; PPR: pentatricopeptide repeat; RPF1: RNA PROCESSING FACTOR 1; RPP1: disease resistance protein with an N-terminal domain homologous to a Toll/interleukin-1 receptor (TIR–NLR class R protein) conferring resistance to *Peronospora parasitica* (causing downy mildew) in *Arabidopsis thaliana*; RPS4/RPS6: TIR–NLR class R proteins conferring resistance to *Pseudomonas syringae* pathovars in *Arabidopsis thaliana*; UPL6: E3 UBIQUITIN-PROTEIN LIGASE 6; VICTR/L: VARIATION IN COMPOUND TRIGGERED ROOT GROWTH RESPONSE(-LIKE) (TIR–NLR class R proteins); WRR4: TIR–NLR class R protein conferring broad-spectrum resistance to *Albugo candida* (causing white rust) in *Arabidopsis thaliana*.

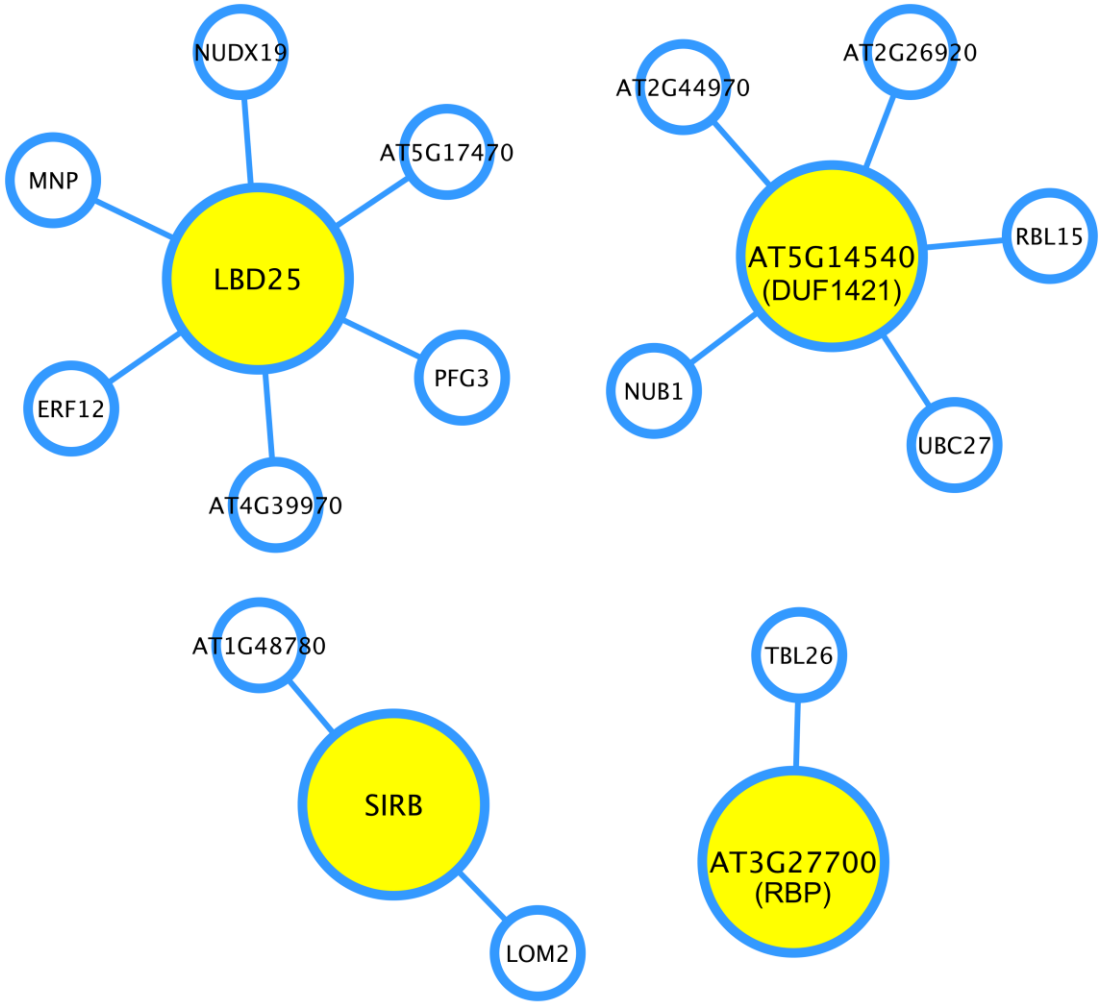


Figure 2.10. Bayesian probabilistic functional gene networks showing predicted connections between four potential *Arabidopsis thaliana* orthologues (yellow) of candidate genes located within the halo blight mapping interval on chromosome 4 of the *Phaseolus vulgaris* reference genome and other functionally associated genes. DUF: domain of unknown function; LBD: lateral organ boundaries domain-containing protein; RBP: RNA-binding protein; SIRB: SIROHYDROCHLORIN FERROCHELATASE B.

Map resolution

The physical and genetic lengths of each chromosome/linkage group, the recombination rate per chromosome, and the marker density for the common bean reference-genome genotype (G19833), and the Stampede × Red Hawk and SOA-BN × Edmund populations are summarised in Table 2.3.

Table 2.3. Summary of the physical (Mb) and genetic (cM) lengths of each chromosome/linkage group, the recombination rate (cM/Mb) per chromosome, and the marker density for the *Phaseolus vulgaris* reference-genome genotype (G19833), and the Stampede × Red Hawk (SR) and SOA-BN × Edmund (SE) populations

Chromosome/ linkage group	Size in reference genome (Mb)	Linkage group length (cM)		cM/Mb		Genetic markers (SE map)	Genetic markers/Mb
		SR map	SE map	SR map	SE map		
1	52.18	84.00	78.88	1.61	1.51	115	2.20
2	49.03	127.60	85.04	2.60	1.73	130	2.65
3	52.22	116.90	74.29	2.24	1.42	102	1.95
4	45.79	94.00	67.05	2.05	1.46	88	1.92
5	40.24	90.80	70.29	2.26	1.75	83	2.06
6	31.97	70.80	55.01	2.21	1.72	108	3.38
7	51.70	105.40	83.69	2.04	1.62	131	2.53
8	59.63	114.00	100.41	1.91	1.68	134	2.25
9	37.40	94.60	76.07	2.53	2.03	123	3.29
10	43.21	60.20	53.69	1.39	1.24	92	2.13
11	50.20	78.50	59.61	1.56	1.19	114	2.27
Unanchored[a]	N/A	N/A	189.80	N/A	N/A	28	N/A
Mean	46.69	94.25	73.09	2.04	1.58	110.91	2.42
Total/genome-wide	513.57	1036.80	993.83	2.02	1.94	1248	2.43

[a]Values for the linkage group containing unanchored markers were not included in the calculation of means.
SR: Stampede × Red Hawk F_2 and F_5-derived recombinant inbred populations genotyped for integration of the common bean reference assembly with a dense genetic map (Schmutz *et al*., 2014).

2.4. DISCUSSION

This research has demonstrated that the recent advances of genotyping-by-sequencing can provide an efficient genotyping platform to enable high-resolution linkage mapping of important traits in common bean. Specifically, a 500-kb mapping interval was defined for quantitative, potentially race-nonspecific, resistance to halo blight with a modest-sized mapping population.

The continuum of phenotypic variation observed in both the SE and JDT recombinant inbred populations following separate inoculations with *Psph* race 6 and race 1 suggests that minor-effect QTL may condition this quantitative resistance. Segregation for resistance derived from PI 150414 in the JDT population, which was selected for halo blight resistance at earlier generations of inbreeding, suggests that it is not behaving as a single recessive factor in this population. However, it is not possible to make a reliable inference of dominance or recessiveness using evidence collected from a sample of RILs, particularly where minor-effect genes may influence the perceived mode of inheritance of the major-effect resistance locus. Previous evidence from genetic experiments by Taylor *et al*. (1978) using F_1, F_2 and F_3 lines indicated simple inheritance of resistance in PI 150414. This resistance factor behaved as a single major recessive in separate crosses with the two susceptible cultivars Seafarer and, to a lesser extent, Cascade (Taylor *et al*., 1978; also see Patel and Walker, 1966; and Dickson and Natti, 1966, for further evidence of simply inherited recessive resistance in PI 150414). All F_1 plants of Cascade × PI 150414 were susceptible, which supported the hypothesis that this resistance is governed by a single recessive gene. Taylor *et al*. (1978) observed good to excellent fits to a 1:3 (resistant : susceptible) segregation ratio in F_2 single-plant progenies of Cascade × PI 150414 following challenge with different races (separately and in mixtures).[4] In the combined F_2 progenies of Cascade × PI 150414, however, there was a lower proportion of resistant individuals and therefore a significant deviation from 1:3, with a χ^2 value of 8.38 (P = 0.01–0.001; Taylor *et al*., 1978: 106–107). Nonetheless, in F_3 progenies of resistant F_2 plants, all 16 plants were classified as

[4] Race 1 isolates used by Taylor *et al*. (1978) were 31A, 98, 710A and 725A. Isolate 882 was the only race 2 isolate used. Race 5 isolate 99 and race 6 isolates 716B and 777A—each then designated race 2—were also used in their study. Race 2 has since been subdivided (Teverson, 1991; Taylor *et al*., 1996a).

resistant following inoculation with a mixture of races, indicating simple recessive inheritance.

The findings reported by Taylor *et al*. (1978) also indicate that the genetic background can influence the expression and perceived mode of inheritance of this resistance. They showed that the resistance from PI 150414 behaved as an incompletely dominant gene in crosses with another susceptible cultivar, Gallatin 50. In the F_1 progeny of Gallatin 50 × PI 150414, all plants exhibited symptoms 'intermediate' between those displayed by their resistant and susceptible parents, indicating allele dosage effects upon resistance; that is, a lower level of resistance was observed in the heterozygous condition. As Taylor *et al*. (1978: 107) described, "[p]lants scored as 'intermediate' showed fewer water-soaked lesions than susceptible plants and these lesions showed a blackening reaction." A blackening (necrotic) reaction is associated with a hypersensitive response that attempts to arrest pathogen infection, while a lower frequency of water-soaked lesions on 'intermediate' plants is indicative of localised infection. An excellent fit to a 1:2:1 (resistant : intermediate : susceptible) ratio was observed in the combined F_2 progenies of Gallatin 50 × PI 150414, which supported the hypothesis that this resistance is governed by a simply inherited, incompletely dominant gene in this cross. However, Taylor *et al*. (1978) reported considerable heterogeneity between single-plant progenies in the F_2 population, with significant deviations from 1:2:1 in two of seven F_2 progenies (χ^2 = 19.42, P = <0.001; and χ^2 = 9.9, P = 0.01–0.001). They speculated that this variation might have been attributable to the small size of single-plant progenies. They also suggested that the subjective phenotyping method might not have ensured accurate and consistent prediction of the genotype.

In the present research, a digital photographic phenotyping approach, using image analysis software such as ImageJ (Schneider *et al*., 2012), was considered as a potential means of increasing the objectivity with which plants are scored for infection. However, such an approach was discounted due to several confounding factors. Leaf puckering and curling following inoculation with *Psph* causes changes to the area and size of the leaf that would be under image analysis. This factor varied from plant to plant and did not appear to show distinct patterns that were consistent among resistant plants and among susceptible plants. Moreover, this factor would confound attempts to determine the ratio of leaf area with lesions (necrotic or water-soaked) to leaf area without lesions. Genotypic differences in unifoliolate leaf

morphology between the two parental lines Edmund (smaller) and SOA-BN (larger) would be another confounding factor; unifoliolate leaf expansion is inhibited in susceptible parent SOA-BN following inoculation with *Psph*. It would therefore be difficult to determine in the progeny whether unifoliolate leaves have reached their maximum potential size following inoculation. Consequently, the usefulness of the ratio of leaf area with lesions to leaf area without lesions determined for each RIL would be limited.

The five-point scale developed by Innes *et al*. (1984) was adapted in the current research to allow for decimal numbers to be assigned to plants exhibiting symptoms intermediate between two integer scores. This provides greater resolving capability than the more qualitative assessment method employed by Taylor *et al*. (1978). The five-point scale of Innes *et al*. (1984) has also been used successfully for genetic mapping of *Pse-1* and *Pse-2* on Pv10 (Miklas *et al*., 2009, 2011), *Pse-3* on Pv02 (Fourie *et al*., 2004; Miklas *et al*., 2011, 2014), and *Pse-6*, *Pse-(Race 1)* and *Pse-(Race 7)* on Pv04 (Miklas *et al*., 2014). Moreover, it is anticipated that resistance-linked markers identified through genetic mapping will enable effective use of these race-specific *R* genes in marker-assisted selection of halo blight-resistant cultivars.

In a study to map *Pse-2*, which confers resistance to all *Psph* races except races 1 and 6, Miklas *et al*. (2011) reported recessively inherited resistance to race 8, consistent with findings reported by Teverson (1991). However, this resistance cosegregated with *Pse-2* in a population of 79 RILs derived from the cross ZAA 12 (A43) × Canadian Wonder, with no recombinants detected. These data confirmed that resistance to race 8 is governed by *Pse-2*, but conflicts with the conclusion of Teverson (1991) that an independent recessive gene (*pse-5*) conferred resistance to race 8. Miklas *et al*. (2011) proposed an alternative explanation to account for these apparently contradictory findings: *Pse-2* confers resistance to race 8 but has epistatic interactions with other genes which partially suppress its expression, resulting in an apparent reversal of dominance in F_1 and segregating F_2 populations following challenge with race 8. In view of the previously documented effects that genetic background can have on the perceived mode of inheritance of halo blight resistance (Taylor *et al*., 1978), this model might provide a useful point of comparison for dissecting the genetic basis of resistance derived from PI 150414.

Pse-3 was previously reported to enhance the phenotypic expression of the quantitative resistance from PI 150414. Teverson (1991) and Taylor *et al*. (1996b) observed enhanced resistance in the presence of dominant *Pse-3* in Edmund and in accession Wis HBR 72 (also possessing resistance from PI 150414). This interaction reportedly induces a 'ghost effect' that enhances quantitative resistance to other *Psph* races besides races 3 and 4 (J. D. Taylor, personal communication). Teverson (1991: 124–131) observed a resistant reaction combined with a severe hypersensitive reaction in Edmund plants inoculated with race 3 and race 4 (mean disease score = 1.0; n = 60), and a resistant reaction without severe hypersensitivity in Edmund plants inoculated with race 6 (mean disease score = 1.5; n = 40). Taylor *et al*. (1996b) observed a corresponding pattern of severe hypersensitive resistance to races 3 and 4, and high-level resistance to races 1, 2 and 6 in Wis HBR 72 plants, which had a disease score of 1.0 following inoculation with each of these latter three races. They also reported that PI 150414, which lacks *Pse-3*, had a disease score of 1.0 in response to separate inoculations with race 1 and with race 2, a disease score of 2.0 in response to inoculation with race 4, and a disease score of 3.0 in response to separate inoculations with race 3 and with race 6.

Fourie *et al*. (2004) reported that a QTL conditioning resistance to a putative race 7 isolate and to an isolate of unknown race, first mapped by Ariyarathne *et al*. (1999), also corresponds to the location of the subsequently mapped gene *Pse-3* on linkage group 2 (chromosome Pv02). Their findings therefore indicate that *Pse-3* conditions resistance to other races besides races 3 and 4. For reactions to races 6 and 1 in the SE population used in the current study, however, the *Pse-3* locus on Pv02 was not found by two-dimensional genome scans to confer additive or epistatic effects with regard to the quantitative resistance from PI 150414. For reaction to race 6, an average disease severity score of 2.2 was calculated for the nine SE RILs predicted to possess resistance only from PI 150414, whereas an average score of 2.3 was calculated for the 29 RILs predicted to possess both resistance from PI 150414 and *Pse-3* (P = 0.25 in a two-sample, one-sided, equal-variance t-test). For reaction to race 1, however, the corresponding average SE RIL scores were 2.4 (PI 150414-derived resistance only; eight RILs) and 2.3 (PI 150414-derived resistance and homozygous or heterozygous at *Pse-3*; 29 RILs) (P = 0.056 in a two-sample, one-sided, equal-variance t-test). Any putative enhanced quantitative resistance

conditioned by *Pse-3* nonetheless requires confirmation using a larger subset of RILs predicted to possess resistance only from PI 150414.

On average, Teverson (1991) assigned a higher level of resistance to Edmund plants inoculated with race 6 (isolate 1299A collected from Tanzania by J. H. C. Davis, 1984) than was assigned in the current research. In the experiments of Teverson (1991: 124–131), 21 Edmund plants had a disease score of 1.0, and 19 had a disease score of 2.0 following inoculation with race 6 isolate 1299A (mean disease score: 1.48; SD = 0.51). In the present study, Edmund had a mean disease score of 2.1 in separate experiments with race 6 isolate 716B (SD = 0.12; n = 28) and race 1 isolate 725A (SD = 0.11; n = 31). This discrepancy may be due to the use of different isolates and/or the subjective method of assessment.

However, evidence to suggest that quantitative resistance is enhanced in the presence of *Pse-3* is more ambiguous in Edmund than in Wis HBR 72. According to Teverson (1991), in the F_2 progeny of the cross Edmund (quantitative resistance + *Pse-3*) × Canadian Wonder (susceptible), five plants (10%) were assigned a disease score of 1.0 (without severe hypersensitivity) following co-inoculation with both race 3 and race 6. The absence of the severe hypersensitive reaction following inoculation with race 3 (a symptom characteristic of plants possessing *Pse-3*) suggested that the disease score of 1.0 was due entirely to the presence in these F_2 plants of the quantitative resistance factor(s) from PI 150414. This is in contrast to the observation of Taylor *et al*. (1996b) that PI 150414 plants (lacking *Pse-3*) had a disease score of 3.0 following inoculation with the same race 6 isolate, 1299A. Taylor *et al*. (1978) and Innes *et al*. (1984) nonetheless reported higher levels of resistance in PI 150414 plants inoculated with isolates then designated race 2. Race 2 has since been subdivided into races 2, 6 and 8 based on interaction phenotypes exhibited by an extended series of Phaseolus host differential lines (Teverson, 1991; Taylor *et al*., 1996a). Subsequently re-classified 'race 2' isolates used in these two studies were 882 (race 2), 99 (race 5), 716B (race 6) and 777A (race 6). However, individual isolates used in each series of inoculations are not specified.

Fine-scale mapping

Repetitive and paralogous sequences in the Pv04 mapping interval present challenges to fine-scale mapping. To progress further in future research aimed at resolving this

locus, a set of 120-nucleotide RepeatMasked[5] probes has been designed at 2× and 3× tiling densities by MYcroarray, Ann Arbor, USA, for target enrichment sequencing of all genes (full-length, unspliced transcripts) within the interval. This exercise confirmed the relatively high proportion of non-unique gene sequences within the mapping interval compared with other regions of the common bean reference genome. Each probe candidate was subjected to BLASTN comparison with the reference genome, and a melting temperature (T_m) at which 50% of molecules are hybridised was estimated for each hit. When stringent BLAST filtering criteria[6] are applied, the average estimated target coverage obtained for the 38 predicted genes within the interval is 65%. The 13 predicted nucleotide-binding site–leucine-rich repeat (NLR)-encoding genes and pseudogenes are the most repetitive within the interval, with an average estimated target coverage of 19.51%. By contrast, an average estimated target coverage of 85.57% was obtained for a probe set targeting primarily the full complement of predicted NLR-encoding genes in the reference genome. Accordingly, variable-stringency BLAST filtering criteria and tiling density have been applied for probe selection to achieve optimal target specificity and breadth of coverage. A combination of long- and short-read target enrichment sequencing will enable elucidation of the haplotypes of resistant and susceptible genotypes at the problematic locus on Pv04, as well as improved subsequent marker design for fine mapping.

Candidate proteins

The physical region of the reference genome corresponding to the major-effect halo blight resistance locus on chromosome Pv04 contains a cluster of 13 genes predicted to encode nucleotide-binding site–leucine-rich repeat (NLR) proteins with putative coiled-coil (CC) N-terminal domains (Table 2.2; Schmutz *et al*., 2014). In the reference genome, four of these were classified as pseudogenes (Phvul.004G007900

[5] Based on the database of Fabaceae repeats in the Michigan State University Plant Repeat Databases, available at http://plantrepeats.plantbiology.msu.edu/index.html (accessed 1 Jul. 2016).

[6] A stringent filter ensures that only specific probes pass; qualifying probes meet one of the following criteria: (1) hits have $T_m \leq 60$ °C; (2) no more than two hits have T_m of 62.5–65 °C; (3) no more than ten hits have T_m of 62.5–65 °C and with at least one failing flanking probe; (4) no more than ten hits have T_m of 62.5–65 °C, no more than two hits have T_m of 65–67.5 °C, and with fewer than two passing flanking probes; or (5) no more than two hits have T_m of 62.5–65 °C, no more than one hit has T_m of 65–67.5 °C, no more than one hit has $T_m \geq 70$ °C, and with fewer than two passing flanking probes.

to Phvul.004G008200) due to the presence of premature stop codons and/or frameshift mutations (see Supplementary Figure 16 of Schmutz *et al*., 2014).

NLRs represent the most abundant class of plant R proteins and provide the second tier of defence against invading pathogens (Dangl and Jones, 2001; Dangl *et al*., 2013). However, the potential roles of most plant NLRs in disease resistance have yet to be experimentally validated and so other functions cannot be ruled out (McHale *et al*., 2006). NLRs confer effective, generally race-specific resistance almost exclusively against obligate biotrophic or hemibiotrophic pathogens (Glazebrook, 2005; Jones and Dangl, 2006; Mengiste, 2012).

It is conceivable that the predicted NLRs at the Pv04 locus confer different resistance specificities against different *Psph* races, with gene duplication, polymorphism and novel configurations potentially enabling recognition of novel pathogen variants (Richter *et al*., 1995; Parker *et al*., 1997; Noël *et al*., 1999; Wiesner-Hanks and Nelson, 2016). Miklas *et al*. (2014) proposed this potential explanation for *Pse-6*, which maps to the same region of Pv04 and confers dominant race-specific resistance to *Psph* races 1, 5, 7 and 9. They identified a putative recombinant in the *P. vulgaris* BelNeb-RR-1 × A55 RIL population with resistance to race 7 and susceptibility to races 1 and 9. Using different RIL populations, Miklas *et al*. (2014) also mapped separate race-specific resistances to race 1 and to race 7 to the same region of Pv04.

Similarly, novel gene configurations within the maize *Rp1 R* gene cluster resulting from unequal crossing-over gave rise to new rust resistance specificities absent from parental lines (Richter *et al*., 1995). Characterisation of the complex *RPP5* locus in *A. thaliana*, governing resistance to *Peronospora parasitica*, revealed that the haplotype of ecotype Landsberg *erecta* (L*er*; resistant) contains ten NLR-encoding homologues, while that of ecotype Columbia (Col-0; susceptible) contains eight (Noël *et al*., 1999). Noël *et al*. (1999) also observed pronounced structural and sequence polymorphisms at these loci and suggested that diversifying selection underlies hypervariability of the ligand-interacting leucine-rich repeat (LRR) residues. They proposed that selection for highly polymorphic *R* gene loci under pathogen pressure indicates that different *RPP5* homologues may have different recognition specificities with regard to pathogen avirulence effectors.

In the current research, however, symptoms on resistant bean plants in the SE and JDT populations following separate inoculations with *Psph* races 6, 1 and 3 (in

the absence of *Pse-3*) appeared to be distinct from those typically associated with race-specific interactions. Compared with the hypersensitive response observed on Red Mexican UI-3 plants (possessing *Pse-1*) following inoculation with race 1 (Figure 2.1A), for example, necrotic reactions at the sites of infiltrative inoculation were less pronounced and less localised, with traces of water-soaking in the areas immediately surrounding necrosis (Figure 2.1B). A broad-spectrum resistance mechanism might therefore underlie the lower-level, quantitative nature of this resistance phenotype. *RPM1* (*RESISTANCE TO P. SYRINGAE PV. MACULICOLA 1*; Grant *et al*., 1995) and *WRR4* (*WHITE RUST RESISTANCE 4*; Borhan *et al.*, 2008, 2010) provide examples of individual NLR-encoding genes that confer more broad-based resistance in Arabidopsis and brassicas against bacterial and oomycete pathogens expressing particular *avr* genes.

Shao *et al*. (2014) identified 337 genes containing nucleotide-binding site (NBS) domains in the common bean reference genome by hidden Markov model (HMM) and BLAST-based searches, although not all possess an LRR domain. As is common in other plant species (Choi *et al*., 2016), many of these genes are arranged in clusters throughout the common bean genome due to local tandem duplications, potentially produced by unequal crossover events (Shao *et al*., 2014; Schmutz *et al*., 2014). Clusters of NLR-encoding genes that are structurally polymorphic are known to exhibit suppressed meiotic recombination (Parker *et al*., 1997; Noël *et al*., 1999; Choi *et al*., 2016), which poses a challenge to resolving the differential recognition specificities of *R* genes within clusters.

The *I* locus governing BCMV resistance (tightly linked with *Pse-3*) on Pv02 provides an intriguing example of suppression of recombination within a cluster of NLR-encoding genes in common bean. Vallejos *et al*. (2006) did not identify any recombinants that would be informative for dissecting this cluster in their pooled-sample genotyping assay of 483 homozygous recessive (*ii*) individuals identified among 3,056 F_2 plants derived from a cross between the susceptible Andean line Calima (*ii*) and the resistant Mesoamerican line Jamapa (*II*). Variation in copy number of NLR-encoding genes with N-terminal domains homologous to Toll/interleukin-1 receptors (TIR) was observed between dominant resistant (~ 24 copies) and recessive susceptible (~ 12 copies) haplotypes at the *I* locus (Vallejos *et al*., 2006).

This two-fold copy-number variation may be indicative of hemizygosity, which inhibits recombination due to obfuscation of the process of homology search and strand invasion of a homologous non-sister chromatid (Vallejos *et al*., 2006; Ozias-Akins *et al*. 1998; Goel *et al*. 2003; Henderson, 2012). Such hemizygosity can result from unequal crossing-over, whereby meiotic misalignment of non-allelic sequences followed by intergenic or intragenic recombination causes unequal exchange of homologues between paired chromosomes (Nöel *et al.*, 1999). Nöel *et al.* (1999) proposed illegitimate recombination of this kind and/or gene conversion to be the most likely explanations for the complex structural polymorphisms observed between *Peronospora parasitica*-resistant and susceptible lines at the *RPP5* locus in *A. thaliana*. Inversions might provide an alternative explanation for inhibited recombination at the *I* locus, as observed at the structurally complex *Mla* locus governing powdery mildew resistance in barley (Wei *et al*., 2002, via Vallejos *et al*., 2006).

Suppressed recombination at the *I* locus may represent an adaptive strategy promoting balancing selection between alleles for BCMV resistance (where BCMV is present) and susceptibility (where other potyviruses and comoviruses are present and deleterious to *I*-bearing beans) (Vallejos *et al*., 2006). The reference-genome genotype (G19833) is susceptible to both BCMV and halo blight, and has the susceptible haplotype at the *I*/*Pse-3* locus (C. E. Vallejos, personal communication; Bello *et al*., 2014). This precedent for NLR copy-number variation in common bean therefore motivates future target enrichment sequencing experiments to determine and resolve the haplotypes of resistant and susceptible lines at the Pv04 locus governing quantitative resistance to halo blight.

Recessive resistance in plant–microbe interactions

In an important contribution to the re-discovery of Mendelian genetics and its application in plant breeding, Biffen (1907, 1912) explained resistance to wheat yellow rust (*Puccinia striiformis* f. sp. *tritici)* by the inheritance of a single recessive gene, and then deployed the resistance allele in new *Triticum* varieties. Most subsequent research has focused attention on dominant disease resistance, which has been shown in numerous plant pathosystems to involve host receptor-like proteins that enable detection of pathogen effector proteins and consequent elicitation of host defence (Kang *et al*., 2005). In contrast, recessive or incompletely dominant disease

resistance may indicate that a host protein, which is targeted by the pathogen in a compatible interaction, is either absent or has been mutated to a dysfunctional form (Fraser, 1986, 1990; Kang *et al*., 2005). This has been supported by increasing reports of mutations in non-receptor proteins conferring recessive resistance to viral, bacterial and fungal pathogens (Collmer *et al*., 2000; Kang *et al*., 2005; Iyer-Pascuzzi and McCouch, 2007; Orjuela *et al.*, 2013; Wang *et al.*, 2013). Combining alleles to confer receptor- and non-receptor-mediated resistance has been proposed as a strategy for providing durable disease control in crop plants (Leach *et al.*, 2001; Iyer-Pascuzzi and McCouch, 2007).

The CC–NLR class protein LOV1 (LOCUS ORCHESTRATING VICTORIN EFFECTS 1) in *A. thaliana* is an example of a functional protein encoded by a dominant gene that confers sensitivity to the toxin victorin and susceptibility to the necrotrophic fungus that produces it, *Cochliobolus victoria* (Sweat *et al*., 2008; Lorang *et al*., 2012; Gilbert and Wolpert, 2013). Sequence analysis of the *LOV1* gene from victorin-insensitive (*lov1*) *A. thaliana* mutants revealed alleles with mutations located in conserved NLR domains essential for R protein function (Sweat *et al.*, 2008). These mutations are comparable to those previously identified in *RPM1* as causing loss of R protein function against *P. syringae* strains possessing *avrRpm1* or *avrB* (Tornero *et al*., 2002, via Sweat *et al*., 2008), indicating that the toxin-mediated susceptibility mechanism governed by LOV1 may be analogous to ETI governed by R proteins (Sweat *et al*., 2008). In this proposed mechanism, effectors wielded by necrotrophs or hemibiotrophs induce necrosis by activating dominant NLR immune receptors, thereby promoting virulence and effector-triggered sensitivity/susceptibility (Toruño *et al*., 2016). NLR-encoding gene sequences identified as truncated or containing deletions in essential domains in halo blight-resistant *P. vulgaris* lines (obtained by target enrichment sequencing) may therefore be candidates that warrant further investigation.

However, Lorang *et al*. (2012) proposed that victorin functions as, or mimics the function of, a pathogen virulence effector, as it is essential for causing disease and triggering defence in *A. thaliana* plants possessing LOV1 (Lorang *et al*., 2012). Phaseolotoxin is a nonspecific phytotoxin produced by some strains of *Psph* and other *P. syringae* pathovars, and causes halo blight-infected plants to exhibit symptomatic chlorotic halos surrounding water-soaked lesions (Arnold *et al*., 2011). Rico *et al*. (2003) reported that non-toxigenic field strains of *Psph*, which lack the

ability to synthesize phaseolotoxin, were the predominant cause of halo blight of common beans grown in Spain. Accordingly, they concluded that phaseolotoxin is not required for pathogenicity or virulence in *Psph*. Its function therefore cannot be analogous to victorin in *C. victoria*. Moreover, ETI underlies disease resistance mechanisms that are effective against many hemibiotrophic pathogens (Glazebrook, 2005; Mengiste, 2012), including *Psph* and other *P. syringae* pathovars (Zhang *et al*., 2010; Baltrus *et al*., 2012; Lindeberg *et al*., 2012). A recessive resistance model premised on toxin-mediated, effector-triggered dominant susceptibility therefore may not be applicable to this interaction in the Phaseolus–Pseudomonas pathosystem. Thus, the quantitative, potentially recessive, halo blight resistance mechanism derived from PI 150414 may not be analogous to that possessed by *A. thaliana lov1* mutants.

Unlike in plant–virus pathosystems (Kang *et al*., 2005), there are few known examples of recessive genes governing resistance to bacterial plant pathogens (Iyer-Pascuzzi and McCouch, 2007). Consequently, the molecular basis of recessively inherited bacterial resistance mechanisms is poorly understood. However, at least nine recessive genes confer resistance to the rice bacterial blight pathogen *Xanthomonas oryzae* pv. *oryzae*, a Gram-negative bacterium that wields the type III secretion system (Iyer-Pascuzzi and McCouch, 2007). The proteins encoded by these recessively inherited genes do not belong to the five major classes of R proteins. For example, *xa5* and *xa13* each encode proteins with essential functions in developmental processes in both their recessive and dominant allelic forms (Iyer-Pascuzzi and McCouch, 2007, and references therein).

Drawing upon Fraser's (1986, 1990) hypothesis that recessive viral *R* genes encode mutated forms of host proteins required for viral replication, Iyer-Pascuzzi and McCouch (2007) proposed that *xa5* and *xa13* may represent mutations in dominant susceptibility alleles required for bacterial virulence. According to this model, some recessive resistance genes may govern a passive mechanism that occurs in the absence of a dominant gene targeted (directly or indirectly) by a corresponding bacterial virulence factor and required for a compatible plant–pathogen interaction. Resistant rice plants possessing either of *xa5* or *xa13* do not exhibit a typical hypersensitive response, suggesting that the resistance mechanisms governed by these recessive alleles induce defence responses that are distinct from those triggered by typical dominant *R* genes. Rice plants possessing *xa5*, for example, exhibit a

delayed and weak hypersensitive response following inoculation with *X. oryzae* pv. *oryzae* (Iyer-Pascuzzi and McCouch, 2007). These atypical reactions in rice may therefore provide useful parallels to those exhibited by bean plants possessing the potentially recessively inherited halo blight *R* gene from PI 150414 (i.e., compared with bean plants possessing dominant major *R* genes [Figure 2.1A], they exhibited less pronounced and less localised necrotic reactions at the sites of infiltrative inoculation, with traces of water-soaking in the areas immediately surrounding necrosis [Figure 2.1B]). Despite the atypical nature of these reactions, Yang *et al*. (2006) presented evidence for a possible gene-for-gene interaction between dominant susceptibility alleles such as *Xa13* and type III effectors. However, *xa5* and *xa13* confer race-specific resistance and therefore may govern defence response mechanisms that are different from those induced by the quantitative, potentially race-nonspecific, resistance from PI 150414 in common bean.

Recessively inherited *R* genes might provide more durable resistance. Under the above paradigm, dominant susceptibility alleles appear to enable virulence. Therefore, mutations in pathogen virulence factors in response to the absence of, or mutations in, corresponding host compatibility factors may cause the pathogen to incur a fitness penalty (Leach *et al.*, 2001; Iyer-Pascuzzi and McCouch, 2007). Enhanced and broader resistance to rice bacterial blight is achieved when *xa5* or *xa13* is deployed with other *R* genes (Iyer-Pascuzzi and McCouch, 2007, and references therein). However, the durability of the resistance provided by different *R* gene combinations involving these two recessive genes has not yet been reported.

Parental polymorphisms in an RNA-binding protein

Six quality-filtered SNVs with SnpEff-predicted (Cingolani *et al*., 2012) 'moderate' impact on protein function that are shared by Edmund and Capulet (and different from SOA-BN and the susceptible reference-genome genotype, G19833) were identified in a predicted RNA recognition motif (RRM)-containing protein (Phvul.004G007600) located within the Pv04 mapping interval in the reference genome (Table 2.2). Also known as RNA-binding proteins (RBPs), this functional class has been implicated in post-transcriptional regulation of plant immunity at various steps of RNA processing (i.e., pre-mRNA 3′ polyadenylation, splicing, editing, and 5′ capping, after which mature mRNAs can be exported to the cytoplasm as candidates for translation) (Qi *et al*., 2010; Woloshen *et al*., 2011; Staiger *et al*.,

2013). RBPs contain characteristic conserved motifs that are predicted to facilitate binding of RNA targets required for execution of RNA-processing functions (Woloshen *et al*., 2011). Post-transcriptional gene regulation, partly enabled by RBPs, can promote rapid responses to biotic and abiotic stimuli (Woloshen *et al*., 2011; Staiger *et al*., 2013).

In common bean cell cultures, for example, a transcript encoding the cell wall protein PvPRP1 (*Phaseolus vulgaris* PROLINE-RICH PROTEIN 1) was rapidly down-regulated following exposure to a fungal elicitor from *Colletotrichum lindemuthianum* (Zhang *et al*., 1993). Zhang and Mehdy (1994) subsequently identified a single 50-kDa protein (PvPRP1-BINDING PROTEIN, PRP-BP), which binds the 3′ untranslated region of the *PvPRP1* mRNA, as a candidate potentially causing destabilisation and degradation of this target mRNA. Reduced synthesis of PvPRP1 during the defence response is thought to be due to its lack of capacity for cell wall strengthening by means of isodityrosine cross-linking under stress conditions (Sheng *et al*., 1991; Zhang *et al*., 1993). Unfortunately, the sequence of PRP-BP has not been determined and so it was not possible to perform BLAST comparisons with common bean sequence databases. Based on TBLASTN comparison of the deduced amino acid sequence of *PvPRP1* (Sheng *et al*., 1991) with the common bean reference genome (Schmutz *et al*., 2014), the most likely physical position for this gene is on chromosome Pv02 at ~ 39.5 Mb (sharing 99% identity with gene Phvul.002G229200 across 44% of the length of the query sequence; predicted functional annotations: ARABINOGALACTAN PROTEIN 31 [PANTHER] containing an Ole e 1-like pollen protein domain [Pfam]).

In Arabidopsis, RBPs possessing RRMs (including glycine-rich GRP7) have been identified as targets for *P. syringae* pv. *tomato* type III-secreted effector HopU1 (Fu *et al*., 2007; Jeong *et al*., 2011). HopU1 is a mono-ADP-ribosyltransferase required by the pathogen for full virulence on *A. thaliana* (Nicaise *et al*., 2013). Fu *et al*. (2007) showed that HopU1 subverts host immunity by modifying RBPs and thereby interfering with RNA metabolism, suggesting that it may restrict the availability of mRNAs with immune response functions in the plant. *A. thaliana* knockout lines lacking a functional *Grp7* allele exhibited enhanced susceptibility to *P. syringae* (Fu *et al*., 2007). Nicaise *et al.* (2013) found that the RRM in GRP7 enables specific binding with *FLS2* and *EFR* transcripts. They revealed a novel bacterial virulence strategy by which HopU1 obstructs the interaction between GRP7

and these PRR-encoding transcripts *in vivo*. This correlated with a HopU1-dependent reduction in FLS2 protein accumulation following inoculation with *P. syringae* pv. *tomato*. GRP7 does not share notable amino acid identity with the RBP within the Pv04 mapping interval. However, this example provides an intriguing precedent for RBP-mediated plant susceptibility to *P. syringae* strains encoding a particular type III effector.

In the current research, all six SNVs within the candidate RBP were predicted to be missense (non-synonymous) mutations located within the coding sequences of exons, which could alter protein functioning or effectiveness. One of these six missense variants is located in exon 1 (encoding a predicted zinc finger domain), one is in exon 3, and four are in exon 4 (encoding a predicted RRM). A further three unfiltered missense variants (with alleles shared by Edmund and Capulet different from those shared by SOA-BN and G19833, but at coverage depths < 10) with predicted moderate impact were identified in exon 2, also predicted to encode an RRM. These missense mutations might affect the affinity with which the RBP binds defence-related transcripts, potentially impairing the accumulation of proteins involved in immune responses during infection with *Psph*.

A general limitation of this variant annotation approach is that it must be performed relative to the common bean reference genome, obtained from a halo blight-susceptible genotype. Classification of variant impact is therefore an estimation of its deleteriousness or disruptiveness to a predicted functional gene in a reference genome. This could nonetheless be a useful approach for localising recessively inherited *R* genes, which may represent mutations in dominant susceptibility alleles. However, the failure to identify SNVs shared by Edmund and Capulet located within the candidate NLR-encoding genes may be due to a programmatic inability to predict gain-of-function mutations relative to predicted truncated pseudogenes in the reference genome, or to copy-number variation between resistant and susceptible genotypes.

Further research is required to determine whether the RBP and/or other candidate proteins play a role in conferring halo blight resistance. However, polymorphisms in the predicted RBP-encoding gene will be useful as cosegregating markers for this resistance.

Candidates within a predicted interactome of defence-related genes

Other candidates within the Pv04 interval with predicted defence-related functional annotations include a cysteine-rich receptor-like protein kinase (a possible defensin; Marshall *et al*., 2011; Phvul.004G011000), a mitogen-activated protein kinase kinase (MAPKK; Phvul.004G010400; Dóczi *et al*., 2007; Berr *et al*., 2010), a WD40 repeat-containing protein (Phvul.004G010500; Miller *et al*., 2016), an E3 ubiquitin protein ligase (UPL6; Phvul.004G009400; Duplan and Rivas, 2014), and a protein with homology to an *A. thaliana* Armadillo/β-catenin-like repeat protein (Phvul.004G009700, containing one missense variant; Sharma *et al*., 2014). Possible orthologues of all except Phvul.004G009700 are connected to the potentially orthologous NLR in *A. thaliana* within the wider interactome predicted using AraNet version 2 (Figure 2.9). This may indicate their involvement in a common defence signalling pathway, and in regulation of immune responses during and/or downstream of ETI-mediated pathogen recognition.

Proposed further experiments

Minor-effect additive QTL were detected on chromosomes Pv03 and Pv08 for reaction to *Psph* race 1. Further experiments to investigate the reproducibility of race 1 reaction phenotypes exhibited by selected RILs possessing particular combinations of resistance and susceptibility haplotypes at loci on Pv04, Pv03 and Pv08 would help to validate these minor-effect QTL. Additionally, seed increase of the complete SE RIL population will permit greater replication in repeated QTL mapping experiments. This will also permit characterisation of RIL phenotypes following inoculation with additional *Psph* races and different race 6 isolates, thereby enabling the specificity of the major-effect QTL on Pv04 to be determined.

Chapter 3

Genome-wide association genetics to identify conserved virulence effectors and targets for host resistance in a globally important bacterial pathogen of common bean

3.1. INTRODUCTION

Rapid advances in genome sequencing and computational biology provide new opportunities to mine genomes for pathogenicity- and immunity-related loci, with an aim for ultimately improving strategies to achieve durable disease resistance in crops. The efficient application of these new tools will be important for securing the advances in crop improvement needed to feed a growing world population.

Genome-wide association (GWA) genetics emerged from the Human Genome project in mid-2000. It has become an increasingly important approach for facilitating the identification of hundreds of genetic polymorphisms that are responsible for inherited diseases and other phenotypic variation in humans (Read and Massey, 2014). Parallel advances have emerged in plant biology, through seminal GWA research in *Arabidopsis thaliana* and maize.

Although bacteria have much simpler genomes than humans and plants, the application of association genetics in prokaryotic organisms has only recently been demonstrated. However, bacterial GWA research is now intensifying thanks to the reduced costs of genome sequencing and public availability of massive microbial reference datasets (Read and Massey, 2014). For example, in research supporting future sustainable food supplies, high-throughput genomic sequencing of diversity in a bacterial pathogen of cassava (*Xanthomonas axonopodis* pv. *manihotis*, *Xam*) has identified highly conserved virulence effectors (Bart *et al*., 2012). These provide primary targets for the development of durable resistance to combat a devastating disease of a major staple food in tropical countries. Similarly, the aim of the current research is to apply GWA genetics to identify candidate genes that encode molecular determinants in the broadly pathogenic race 6 of *Pseudomonas syringae* pathovar *phaseolicola* (*Psph*), which correspond to novel halo blight resistance in common bean (*Phaseolus vulgaris* L.).

Psph uses the type III secretion system (T3SS), a specialised protein structure, to inject molecular determinants of virulence and pathogenicity into the cytoplasm of a host plant cell. These pathogen proteins are known as type III effectors (T3Es) and are essential for causing disease in a compatible host. Some effectors can be detected within the cytoplasm by a matching host receptor (R) protein, which can then elicit host defences and, often, rapid cell death in a resistant

host genotype. These effectors are often expendable (present, absent, or exhibiting allelic variation) as the pathogen evolves to evade detection by the host.

Using the Phaseolus–Pseudomonas pathosystem, Taylor *et al*. (1996a,b) identified five major-effect *R* genes in the bean host (four of which have been subsequently mapped; Miklas *et al*., 2009, 2011, 2014), which were used to group strains of *Psph* bacterium into nine races. As a conclusion, they proposed a summary model of gene-for-gene relationships (Flor, 1942, 1971) based on interaction phenotypes observed on eight bean lines possessing these *R* genes (either singly or in combinations) following inoculation with 175 *Psph* isolates selected from an extensive collection. The *Psph* collection was assembled by Taylor and colleagues over two decades from five continents (Europe, North America, South America, Africa and Australasia).

Other effectors are determinants of pathogenicity in a host species (pathovar determinants), and are most likely highly conserved within a pathovar. *Psph* race 6 defines a broadly virulent group of strains that is undetected by any of the known race-specific major *R* genes (Taylor *et al*., 1996a,b; Miklas *et al*., 2009, 2011, 2014). Thus, based on the current paradigm, the hypothesis is that highly conserved genes that encode T3Es in *Psph* race 6 are candidates for pathovar-level determinants. These candidates would provide excellent biomarkers for identifying plant genes that determine race-nonspecific (pathovar-level) resistance to *Psph*.

Whole-genome sequencing has greatly advanced our understanding of pathogenicity effectors in *Psph* and provides an excellent genomics platform for this research. One complete reference genome has been assembled for *Psph* race 6 isolate 1448A (Joardar *et al*., 2005) and a draft genome has been assembled for race 1 isolate 1644R (Baltrus *et al*., 2012). Twenty-seven T3E-encoding genes have been previously identified in the genome of *Psph* race 6 (isolate 1448A), with validation by mutational (loss-of-function) and transgenic (gain-of-function) experiments (Chang *et al*., 2005; Vencato *et al*. 2006; Zumaquero *et al*. 2010). Effector proteins that elicit host defences following recognition by halo blight resistance proteins Pse-1 (HopF1, formerly AvrPphF), Pse-2 (HopX1, formerly AvrPphE), and Pse-3 (HopAR1, formerly AvrPphB) in Phaseolus have been molecularly characterised (see Tsiamis *et al*., 2000; Stevens *et al*., 1998; and Jenner *et al*., 1991, respectively).

This study exploits whole-genome, next-generation sequencing of 32 isolates of *Psph*, selected from a pathogenically and geographically diverse collection, to

enable identification of candidate effector genes, with a specific aim of identifying candidate determinants of a pathovar or pathotype level of pathogenic capability. Eleven of the newly sequenced isolates are from *Psph* race 6 and of wide geographic origin, and the remainder is a representative sample of isolates from the eight other described *Psph* races. This research reports sequence polymorphisms (allelic sequence and presence/absence variation) occurring across *Psph* races in Pseudomonas T3SS proteins. This provides the basis for the identification of candidate effectors, including (**1**) those conserved amongst all *Psph* races (candidate targets for race-nonspecific resistance), and (**2**) those conserved amongst a diverse collection of *Psph* race 6 isolates but polymorphic in other races (candidate targets for specific resistance to race 6).

A complementary objective is to apply full-genome SNP-based phylogenetic analysis to investigate the geographic history of *Psph* race 6 relative to other *Psph* races and public reference genomes of other *P. syringae* pathovars. This analysis will reveal the extent of historical global movement within the pathovar, the results of which will have implications for the development of future strategies for achieving durable, globally effective halo blight resistance.

3.2. MATERIALS AND METHODS

Bacterial isolates

Thirty-two isolates of *Psph* were selected for whole-genome sequencing (Tables 3.1 and A1.1, APPENDIX 1). The isolates were selected from a pathogenically and geographically diverse collection, assembled by Dr John Taylor and colleagues over two decades of collecting from 17 countries across four continents. They include isolates from each of the nine described pathotypic races (Taylor *et al*., 1996a), with 11 of the selected isolates from race 6. Their research concluded with a summary of race designations, which have been used in the current research. In view of the predicted absence of HopF1 from *Psph* isolate 1308 (Table 3.2), originally designated as race 1, lines of the halo blight differential series (except for *P. acutifolius* accession 1072) were assessed for their reaction to this isolate as described in Chapter 2 (Taylor *et al*., 1996a). Isolate 1308 was virulent on each differential bean line tested and, accordingly, was re-classified as race 6.

PCR amplification, Sanger sequencing and BLASTN analysis of a partial, 572-bp sequence of the constitutive sigma factor gene *rpoD* were performed to confirm that each bacterial isolate was of *Psph*. The PCR reaction mix was composed of 10 mM of Tris-HCl (pH 8.3), 50 mM of KCl, 1.5 mM of $MgCl_2$, 0.001% gelatin, 0.2 mM of each dNTP, 1 µM of each primer, 20–50 ng of genomic DNA, and 0.75 U of Taq DNA polymerase (1X REDTaq ReadyMix; Sigma Aldrich, Gillingham, UK), in a final volume of 25 µl. Amplifications were performed on a GeneAmp PCR System 9700 thermocycler (Applied Biosystems, Foster City, USA) using primers described by Rico *et al*. (2003), with an initial denaturation step at 94 °C for 10 min; followed by 30 cycles of 94 °C for 1 min, 50 °C for 1 min, and 72 °C for 1 min; and with a final step at 72 °C for 10 min. Amplification of the *rpoD* sequence in some isolates required optimisation of PCR conditions. For these isolates, the PCR reaction mix contained 0.2 mM of each dNTP, 1 µM of each primer, 20–50 ng of genomic DNA, 1X Fermentas Long PCR Buffer with 1.5 mM of $MgCl_2$, and 1.25 U of Fermentas Long PCR Enzyme Mix (Fisher Scientific, Loughborough, UK), in a final volume of 25 µl. Cycling conditions consisted of an initial step at 94 °C for 2 min; followed by 30 cycles of 94 °C for 1 min, 55 °C for 1 min, and 68 °C for 1 min; and a final step at 68 °C for 10 min.

Genome sequencing

DNA for the preparation of sequencing libraries was isolated from bacterial cultures grown overnight on King's Medium B (King *et al.*, 1954) at 25 °C. Specifically, bacteria were suspended in sterile high-purity water to give turbid suspensions between 2 and 4 on the McFarland scale (McFarland, 1907). Total DNA of each bacterial isolate was extracted using the DNeasy Blood & Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's protocol for Gram-negative bacteria. Double-stranded DNA was quantified using the Qubit dsDNA BR Assay Kit with a Qubit 2.0 Fluorometer (Invitrogen, Waltham, USA). DNA quality was determined using a NanoDrop Spectrophotometer (NanoDrop, Wilmington, USA) and gel electrophoresis. Libraries were constructed using the Illumina Nextera XT DNA Library Preparation Kit (Illumina, San Diego, USA) following the manufacturer's instructions and validated on an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA). The average DNA fragment length was between 600 bp and 700 bp, with fragments ranging from 200 bp to > 1,000 bp. Following PCR amplification, the libraries were purified using AMPure XP beads (Beckman Coulter, High Wycombe, UK). Libraries were pooled and sequenced by paired-end sequencing on two lanes (one pool per lane) of an Illumina HiSeq 2500 instrument (at Illumina Cambridge Ltd., Saffron Walden, UK) using the 500-cycle HiSeq Rapid SBS Kit v2 (supplemented to run 600 cycles), generating 300-bp paired-end reads.

Quality processing of sequencing reads

The quality of raw and cleaned sequencing reads was assessed using FastQC version 0.11.2 (Andrews, 2010). Illumina adapters, PCR primers and low-quality bases were removed using Trimmomatic version 0.32 (Bolger *et al*., 2014) in paired-end and palindrome modes. Read-cleaning settings informed by FastQC review were applied as follows: a maximum read–adapter sequence alignment mismatch count of 2 bases; a palindrome trimming accuracy score threshold of ≥ 30; a simple trimming accuracy score threshold of ≥ 10; a minimum length of adapter fragments of 1 base; removal of the first 20 bases from the beginning of reads; cropping of reads to 210 bases; a sliding window size of 4 bases with an average Phred-scaled sequence quality threshold of ≥ 15; removal of low-quality bases from the beginning and end of reads,

with a Phred-scaled quality threshold of ≥ 3; and a read length threshold of ≥ 50 bases.

De novo genome assembly

Thirty-two high-quality draft *Psph* genomes were assembled *de novo* from cleaned reads using SPAdes version 3.6.2 (Bankevich *et al*., 2012) with read error correction and the '--careful' option applied. Multiple k-mer lengths (sequences of 21, 33, 55, 77, 99 and 127 bases) were specified for optimal assembly. The multi-kmer de Bruijn graph-based genome assembler SPAdes was chosen in view of the advantages conferred by its strategy to optimise assembly at each genomic region, and its favourable performance in evaluations of different *de novo* assemblers (Magoc *et al*., 2013; Algorithmic Biology Lab, undated). By providing for the use of different values of 'k' at different genomic regions, this strategy seeks to reduce both the inclusion of false overlaps (with larger values of k) and the omission of true overlaps (with smaller values of k) between sequences (Bankevich *et al*., 2012; Magoc *et al*., 2013). Statistics summarising the quality of each assembly were generated using QUAST version 3.2 (Gurevich *et al*., 2013) (Table 3.1). Average depth of coverage was estimated for each assembly by dividing the total number of contributing bases by the assembly size (~ 6 Mb). *De novo*, rather than reference-based (*Psph* race 6), assemblies were performed to enable identification of effectors and other T3SS components that are absent from race 6 isolates and present in isolates from other races. These proteins may be determinants of avirulence in other races.

T3SS protein prediction

A comprehensive database of 19,024 protein sequences from plant and animal pathogens was compiled using keyword searches (APPENDIX 6) in the Reference Sequence (RefSeq) non-redundant protein database of the National Center for Biotechnology Information (NCBI; Tatusova *et al*., 2014), as well as from databases containing previously characterised *P. syringae* T3SS proteins. This compilation includes known and putative T3Es and other components of the T3SS secretory apparatus in the reference assembly for *Psph* race 6 isolate 1448A (Joardar *et al*.,

2005) [7] downloaded from the Pseudomonas Genome Database (http://www.pseudomonas.com; Winsor *et al*., 2016), and in the hypersensitive response and pathogenicity (Hrp) and Hrp outer protein (Hop) databases downloaded from the *Pseudomonas syringae* Genome Resources web site (http://www.pseudomonas-syringae.org/pst_func_gen2.htm; Lindeberg *et al*., 2005). To reduce redundancy within the database, amino acid sequences were clustered using the UCLUST algorithm (Edgar, 2010) with an identity threshold of 90% for a final database of 5,869 representative query protein sequences.

This approach took advantage of the well-curated and annotated T3SS protein databases that have been developed for Pseudomonas. The compilation of a comprehensive database of proteins with experimentally validated or computationally predicted T3SS functions was also favoured over *de novo* prediction because a universally applicable type III secretion signal pattern has not yet been defined (Arnold *et al.*, 2009; Bart *et al*., 2012).

Contiguous sequences (contigs) were scanned for the presence of predicted T3SS protein-encoding sequences by TBLASTN comparison with the query protein database using the Biotools version 1.2.12 suite of Python scripts as described in Bart *et al*. (2012). To account for polymorphisms in functionally homologous protein domains between different pathogens (Zhao *et al*., 2011, via Bart *et al*., 2012), a protein hit was defined according to low-stringency criteria, based on a sequence having ≥ 45% amino acid identity over ≥ 80% of the length of a query protein within the T3SS database (Bart *et al*., 2012). Subsequent gene prediction permitted the use of alternate start codons that are different from that represented in the putatively homologous query protein. The minimum sequence length permitted for the prediction of open reading frames was 300 bases. The corresponding nucleotide sequence for each TBLASTN-identified protein hit within a contig was obtained and its coding sequence translated using Biotools version 1.2.12.

A database was compiled to summarise the presence and absence of all T3SS protein hits in each newly assembled *Psph* draft genome, in two previously assembled *Psph* genomes including a completed reference genome for race 6 isolate 1448A (NCBI BioSample accession: SAMN02603162; Joardar *et al*., 2005) and a

[7] See rows 213–285 of Table S3 in Joardar *et al*. (2005), and rows 269–381 of the table "Comparative pathogenomics of Pseudomonas" available at http://www.mgc.ac.cn/VFs/download.htm (Chen *et al*., 2016).

draft genome for race 1 isolate 1644R (SAMN02471563; Baltrus *et al*., 2012), as well as in previously assembled genomes for *P. syringae* pathovars *glycinea* (strain race 4; SAMN02471825), *maculicola* (strain CFBP 1657; SAMN03328955), *pisi* (strain PP1; SAMN02471574), *syringae* (strain B728A; SAMN02604347) and *tomato* (strain DC3000; SAMN02604017) downloaded from the Pseudomonas Genome Database (http://www.pseudomonas.com; Winsor *et al*., 2016) (Table A7.1, APPENDIX 7).

Translated amino acid sequences that represent hits corresponding to previously identified *P. syringae* avirulence (Avr), Hop and Hrp proteins (named according to the unified nomenclature proposed by Lindeberg *et al*., 2005) were aligned using T-Coffee version 11.00.8cbe486 (Notredame *et al*., 2000) to enable assessment of allelic variation across *P. syringae* strains. Multiple sequence alignments were manually inspected for all types of mutation, including premature stop codons or frameshift mutations (such as those resulting in the use of alternate start codons to produce truncated protein hits), insertions and deletions not causing reading frame shifts, and missense mutations. Resulting annotations denoting the allelic sequence and presence/absence variation of these virulence and pathogenicity proteins across *P. syringae* strains are summarised in Table 3.2.

Assemblies were also compared and visualised by aligning *de novo* contigs to the chromosome and plasmids of *Psph* race 6 isolate 1448A by BLASTN comparison with an E-value threshold of 1×10^{-6} using BLAST Ring Image Generator (BRIG; Alikhan *et al*., 2011). Labels for effector-encoding genes were retrieved from GenBank annotation files for *Psph* 1448A for inclusion in the output images.

SNP-based phylogenomic and association analysis

A matrix of concatenated pan-genome SNPs identified amongst the 34 newly and previously assembled *Psph* genomes, the genomes of the five other *P. syringae* pathovars listed above, and the reference genome for *Pseudomonas fluorescens* strain SBW25 (NCBI BioSample accession: SAMEA2272316) was generated using kSNP version 3.02 (Gardner *et al*., 2015). The choice of *P. fluorescens* as an outgroup for phylogeny inference was informed by its close phylogenetic relationship with *P. syringae* and its non-pathogenic status (Stavrinides, 2009). The kSNP3 program 'Kchooser' was used to determine an optimum k-mer size of 21 to

apply to SNP matrix construction for these input assemblies. This defined the length of the nucleotide sequences found in all input assemblies. The identification of putative SNPs within k-mers of a specified size ensures that variant calling is not affected by any potential assembly errors (Timme *et al*., 2013). The fraction of 21-mers found in all 40 genome assemblies was 0.453, indicating that SNP detection efficiency would be high (> 90%) and that subsequent estimation of a phylogenetic tree would be accurate (> 97%) (Gardner and Hall, 2016). SNP positions and corresponding gene annotations were determined relative to the completed reference assembly for *Psph* race 6 isolate 1448A in the first instance, or the draft assembly for *Psph* race 1 isolate 1644R where annotations were absent from *Psph* 1448A. All identified core and accessory SNPs were included to create a pan-genome SNP matrix.

A full-genome maximum likelihood phylogenetic tree was inferred from the SNP matrix using RAxML-HPC version 8.0.26 (Stamatakis, 2014) under the general time reversible–gamma ('GTRGAMMA') model of nucleotide substitution rate heterogeneity. Estimated median substitution rates were used to represent each rate class in determining the discrete gamma distribution. A rapid bootstrap analysis was performed under the extended majority-rule consensus tree criterion ('autoMRE'; terminating after an automatically determined number of 650 bootstrap replicates) followed by a maximum likelihood analysis of the original alignment of concatenated SNPs to search for the best-scoring tree. The resulting maximum likelihood tree was visualised and formatted in FigTree version 1.4.2 (Rambaut, 2014). A RAxML-inferred phylogeny was favoured over the maximum likelihood phylogeny generated by kSNP3 because the former calculates bootstrap percentages rather than Bayesian posterior probabilities. Bootstrap values are thought to provide more conservative and accurate estimates of branch support (i.e., tree reliability) than do Bayesian posterior probabilities, which often are excessively high (Cummings *et al*., 2003; Simmons *et al*., 2004).

Two additional pan-genome SNP matrices were generated using kSNP3 for downstream analysis with the Predict Phenotypes From SNPs version 2 package (PPFS2; Hall, 2014a,b) for statistical association analyses of microbial genomes. This enabled identification of SNP alleles that are predictive of, and potentially causally related to, the *Psph* race 6 virulence phenotype. PPFS2 identifies SNP alleles with the lowest χ^2 probabilities of being randomly distributed with regard to a

phenotype of interest. Such diagnostic SNPs have the greatest resolving power to predict the phenotypes of strains represented in the kSNP3-generated matrix.

One SNP matrix was constructed for the 34 *Psph* assemblies only, and another was created for the 34 *Psph* assemblies and the assembly for *P. fluorescens* strain SBW25, each with a Kchooser-estimated optimum k-mer size of 19. *P. fluorescens* was included in the latter matrix as an outgroup genome. The fractions of 19-mers found in all 34 and 35 genome assemblies were 0.823 and 0.497, respectively. For each SNP matrix, a consensus set of the smallest number of SNPs that is collectively predictive of the *Psph* race 6 phenotype was identified using the bootstrapping approach implemented in PPFS2. PPFS2 determines the accuracy of SNP-based predictions by comparing the predicted phenotypes of a subset of strains whose phenotypes have been randomly designated 'unknown' with their known actual phenotypes. The allelic information for the subset of strains whose phenotypes are designated 'known' (unchanged by PPFS2) provides the basis for these predictions (Hall, 2014b).

An initial single-replicate PPFS2 analysis was performed to determine optimal settings for identifying diagnostic SNPs. Based on the results of this preliminary analysis, the following settings were applied to PPFS2 analyses of each SNP matrix: 50 bootstrap replicates to identify a consensus set of predictive SNPs; a maximum of 1,000 predictive SNPs may be included in the consensus set (default); a maximum positive predictive value (PPV) of 98% (default), at which addition of SNPs is terminated; PPV may decrease to ≥ 90% of the highest recorded PPV, after which addition of SNPs is terminated; inclusion of SNP positions uniquely present or absent in phenotype-positive or phenotype-negative strains; a SNP must be present in ≥ 50% of bootstrap replicates for inclusion in the final consensus set of diagnostic SNPs (default). As defined by Hall (2014b: 9), PPV refers to "the probability that a strain that is predicted to be positive is actually positive" with regard to the phenotype of interest, and is given by the number of true positives (TP)/(TP + the number of false positives, FP).

The 'CausalSNPs' module of PPFS2 determines which of those diagnostic SNPs have the highest probability of being causally related to the phenotype. SNPs with diagnostic capability may be indicative only of evolutionary history (Hall, 2014a,b). Diagnostic SNPs identified as potentially causal are assumed to be "more likely than the average diagnostic SNP to change allele state along the same branches

along which the phenotype changes" (Hall, 2014a: 2). For each kSNP3-generated SNP matrix, a reduced set of SNPs was imported into MEGA version 7 (Kumar *et al*., 2016) for estimation of the ancestral state of each SNP at each internal node of a maximum likelihood tree. This PPFS2-generated subset consisted of the diagnostic SNPs and a random sample of 975 SNPs (the number of diagnostic SNPs multiplied by 25) from the kSNP3-generated matrix. This step enabled identification of the change in the allelic state of each SNP along a branch from an ancestral node to a descendant node of a manually rooted tree (Hall, 2014a). The maximum likelihood tree inferred from the subset of the 34-genome (*Psph* only) SNP matrix was rooted along the branch from *Psph* 1644R to the other *Psph* isolates. *Psph* 1644R was estimated to be the most closely related to the *P. fluorescens* outgroup genome by analysing an equivalent subset of the 35-genome SNP matrix in MEGA7. This approach improves the predictive power of PPFS2 by excluding SNPs that would otherwise be contributed by the outgroup genome (Hall, 2014a,b).

Estimation of maximum likelihood trees and ancestral states was performed in MEGA7 essentially as described by Hall (2014b). However, in separate analyses, each of the maximum likelihood and parsimony trees generated by kSNP3 for each SNP matrix was used as the 'Initial Tree' for maximum likelihood estimation. SNP-based maximum likelihood trees are favoured over parsimony trees for bacterial genomes, which contain thousands of genes. This is because rapidly evolving loci constitute a relatively small proportion of bacterial genomes compared with viral genomes. Consequently, bacterial genomes generally are not characterised by high levels of heterotachy (i.e., pronounced changes in the rates at which site-specific nucleotide substitutions evolve over time; Gardner and Hall, 2013). However, according to apparently contradictory recommendations from Hall (2014b: 14), the parsimony trees inferred by kSNP3 are more accurate than the equivalent maximum likelihood trees. By using a parsimony tree to initialise the tree search, MEGA will reportedly generate a maximum likelihood tree that is approximately "97–98% congruent with the [parsimony] tree estimated by kSNP" (Hall, 2014b: 14).

For each analysis, the 'Most Probable Sequences' output file generated by MEGA7 was processed with the PPFS2 program 'CausalSNPs'. This enabled identification of diagnostic SNPs with the highest likelihood of being causal, based on low χ^2 probabilities that the change in their allelic state occurred "randomly across the same branches across which the phenotype changed" (Hall, 2014a: 3).

Gene annotations derived from GenBank files for the genomes of *Psph* 1448A and 1644R were retrieved for corresponding non-synonymous SNPs identified as potentially causal of the *Psph* race 6 virulence phenotype.

A maximum likelihood phylogeny was inferred using RAxML-HPC version 8.0.26 (Stamatakis, 2014) as described above (but with the rapid bootstrap analysis terminating automatically after 550 replicates) from the SNP matrix generated by kSNP3 for the 34 *Psph* assemblies. The tree was visualised and formatted in FigTree version 1.4.2 (Rambaut, 2014).

All memory-intensive analyses not requiring a graphical user interface were performed using the command-line interface of the Bio-Linux 8 operating system on Ubuntu Server LTS 14.04 (64-bit, 384 GB of RAM, and two 16-core processors).

3.3. RESULTS

Summary of assembly quality

This research has exploited Illumina next-generation sequencing and multi-kmer de Bruijn graph-based *de novo* assembly to generate high-quality draft genomes for 32 pathogenically diverse isolates of *Psph*. Assembly quality metrics are summarised in Table 3.1. An average estimated depth of coverage across all assemblies of 585 × was obtained (range = 82–1,652 ×; SD = 472 ×). Advances in next-generation sequencing exemplified by the Illumina HiSeq 2500 instrument and in *de novo* assembly algorithms exemplified by SPAdes (Bankevich *et al*., 2012) have enabled considerable improvements in genome assembly quality within the past five years. For example, the contiguities of the assemblies produced in the current research represent improvements on that of the *de novo* genome assembly for *Psph* race 1 isolate 1644R reported by Baltrus *et al*. (2012). The average N50 contig length for the 32 new *Psph* assemblies is 85.6 kb (range = 79.3–98.9 kb; SD = 4.0 kb), while the N50 value for the 1644R assembly is 13.1 kb. The N50 statistic calculated by QUAST version 3.2 (Gurevich *et al*., 2013) defines the length of the smallest contig in a collection of contigs that covers at least half of an assembly. An average number of contigs of 275 (range = 195–733; SD = 124) for the new *Psph* assemblies also represents a marked improvement on the 1,169 contigs in the 1644R assembly. On average, 5,036 (97.4%) of the 5,172 predicted full-length genes in the reference genome for *Psph* race 6 isolate 1448A were present in the 32 genomes assembled in this study.

Table 3.1. Summary of genome assemblies for 34 isolates of *Pseudomonas syringae* pv. *phaseolicola*

Isolate	Race	Origin (date of collection)	Host	Depth of coverage (×)	Contigs	N50 length (bp)	Total length (bp)	Genome fraction (%)	Predicted genes	Partial genes
1644R	1	USA (1971)	*Vigna radiata*	70	1,169	13,068	5,930,644	N/A	6,429	N/A
1281A	1	UK (1984)	*Phaseolus coccineus*	1,181	216	85,782	6,578,666	96.5	5,002	152
1816A	1	Madagascar (1987)	*Phaseolus lunatus*	118	179	98,175	6,554,289	96.3	5,008	138
725A	1	UK (1973)	*Phaseolus vulgaris*	1,641	194	85,782	6,139,329	96.4	5,004	145
Average				**980**	**196**	**89,913**	**6,424,095**	**96.4**	**5,005**	**145**
1502A	2	Mauritius (1986)	*Phaseolus vulgaris*	101	164	85,948	5,993,933	97.2	5,068	128
1678	2	Sweden (1956)	*Phaseolus vulgaris*	298	186	82,094	5,982,745	97.5	5,068	123
2698A	2	Zimbabwe (1990)	*Phaseolus vulgaris*	510	170	84,153	5,961,177	97.2	5,069	127
882	2	USA (1975)	*Phaseolus vulgaris*	123	166	87,138	5,989,312	97.1	5,066	131
Average				**258**	**172**	**84,833**	**5,981,792**	**97.3**	**5,068**	**127**
1301A	3	Tanzania (1984)	*Phaseolus vulgaris*	82	186	79,377	6,071,513	97.8	5,064	133
1567A	3	Burundi (1986)	*Phaseolus vulgaris*	474	192	85,939	6,142,881	97.1	5,061	129
2475A	3	Colombia (1989)	*Phaseolus vulgaris*	1,403	195	86,218	6,189,128	97.2	5,068	123
Average				**653**	**191**	**83,845**	**6,134,507**	**97.4**	**5,064**	**128**
1302A	4	Rwanda (1984)	*Phaseolus vulgaris*	1,150	292	82,889	6,271,043	97.1	5,055	128
1334A	4	Uganda (1984)	*Phaseolus vulgaris*	986	267	82,037	6,612,400	97.2	5,068	122
2240A	4	Burundi (1988)	*Phaseolus vulgaris*	743	212	89,907	6,130,764	97.1	5,066	127
Average				**960**	**257**	**84,944**	**6,338,069**	**97.1**	**5,063**	**126**
11	5	Canada (1941)	*Phaseolus vulgaris*	809	192	85,945	6,113,243	96.8	5,031	138
1516A	5	Tanzania (1986)	*Neonotonia wightii*	438	180	84,187	5,995,698	91.3	4,736	188
Average				**623**	**186**	**85,066**	**6,054,471**	**94.1**	**4,884**	**163**

Table 3.1. Continued

Isolate	Race	Origin (date of collection)	Host	Depth of coverage (×)	Contigs	N50 length (bp)	Total length (bp)	Genome fraction (%)	Predicted genes	Partial genes
1448A	6	Ethiopia (1985)	*Phaseolus vulgaris*	N/A	N/A	N/A	6,112,448	100.0	5,172	N/A
1010	6	USA (1979)	*Phaseolus vulgaris*	246	188	83,963	6,196,706	96.9	5,060	123
1294	6	Colombia (1984)	*Phaseolus vulgaris*	114	176	86,164	6,002,101	97.3	5,070	127
1299A	6	Tanzania (1984)	*Phaseolus vulgaris*	197	189	83,969	6,112,148	97.0	5,057	128
1308	6	Colombia (1984)	*Phaseolus vulgaris*	1,652	196	84,026	6,515,328	97.3	5,068	124
1599A	6	UK (1986)	*Phaseolus coccineus*	263	170	87,180	6,121,559	97.3	5,071	122
1715A	6	Ethiopia (1986)	*Phaseolus vulgaris*	1,526	176	81,983	6,033,241	96.9	5,049	128
1769A	6	Bulgaria (1987)	*Phaseolus vulgaris*	474	178	83,934	5,953,463	97.2	5,067	128
2109	6	Spain (1988)	*Phaseolus vulgaris*	124	188	87,180	6,167,797	96.9	5,046	126
2654A	6	Lesotho (1990)	*Phaseolus vulgaris*	755	177	84,118	5,943,475	97.1	5,060	130
2693A	6	Zimbabwe (1990)	*Phaseolus vulgaris*	99	180	86,218	5,923,768	95.9	4,980	126
716B	6	UK (1973)	*Phaseolus vulgaris*	535	200	82,889	6,099,238	97.5	5,062	128
Average				**544**	**183**	**84,693**	**6,097,166**	**97.0**	**5,054**	**126**
1314	7	Colombia (1984)	*Phaseolus vulgaris*	206	184	87,288	6,057,432	96.3	4,997	145
1354A	7	Kenya (1985)	*Phaseolus vulgaris*	333	183	83,815	6,017,838	96.4	5,002	151
1549A	7	UK (1986)	*Phaseolus vulgaris*	430	177	87,138	6,023,625	96.4	5,000	147
Average				**323**	**181**	**86,080**	**6,032,965**	**96.4**	**5,000**	**148**
1645	8	Tanzania (1964)	*Dolichos* sp.	586	178	82,889	6,025,632	97.2	5,065	123
2654C	8	Lesotho (1990)	*Phaseolus vulgaris*	733	186	81,983	6,211,774	97.5	5,062	135
Average				**659**	**182**	**82,436**	**6,118,703**	**97.4**	**5,064**	**129**
2732E	9	Colombia (1990)	*Phaseolus vulgaris*	389	185	98,934	6,044,295	96.3	5,003	143

Depth of coverage represents the estimated average across an assembly. Genome fraction is the proportion of the reference genome (1448A; Joardar *et al*., 2005) covered by assembled contigs. Red text denotes isolates with previously assembled genomes (Joardar *et al*., 2005; Baltrus *et al*., 2012). Values for isolates 1644R (NCBI BioSample accession: SAMN02471563) and 1448A (SAMN02603162) were excluded from the calculation of averages for race 1 and race 6 isolates, respectively.

Pseudomonas T3SS protein database

Compilation of a summary database of sequence polymorphisms in Pseudomonas T3SS proteins has enabled rapid identification of core and accessory proteins implicated in pathogenicity and virulence in *Psph*. Table 3.2 summarises the computationally predicted allelic sequence and presence/absence variation of named Avr, Hop and Hrp proteins in each of the 32 draft *Psph* genomes, in two previously assembled *Psph* genomes from race 6 and race 1 (Joardar *et al*., 2005; Baltrus *et al*., 2012), and in five other *P. syringae* pathovars pathogenic on leguminous, cruciferous and solanaceous hosts.

A primary objective of this research was to identify T3SS proteins that are highly conserved amongst all races of the pathovar, as candidate targets for race-nonspecific resistance to halo blight in common bean. With the exception of the previously characterised avirulence elicitors HopAR1 and HopF1, only proteins that were found in all races (and their paralogues) are shown in Table 3.2. The predicted full T3SS protein inventories of these genomes are summarised in Table A7.1 (APPENDIX 7). Inspection of multiple sequence alignments of individual Avr, Hop and Hrp proteins has revealed those which are highly conserved within the pathovar (denoted by rows consisting of cells marked primarily with 'C' in Table 3.2). T3SS proteins that are conserved amongst all *Psph* races represent possible targets for hypothesised race-nonspecific resistance to halo blight in common bean, such as that derived from PI 150414.

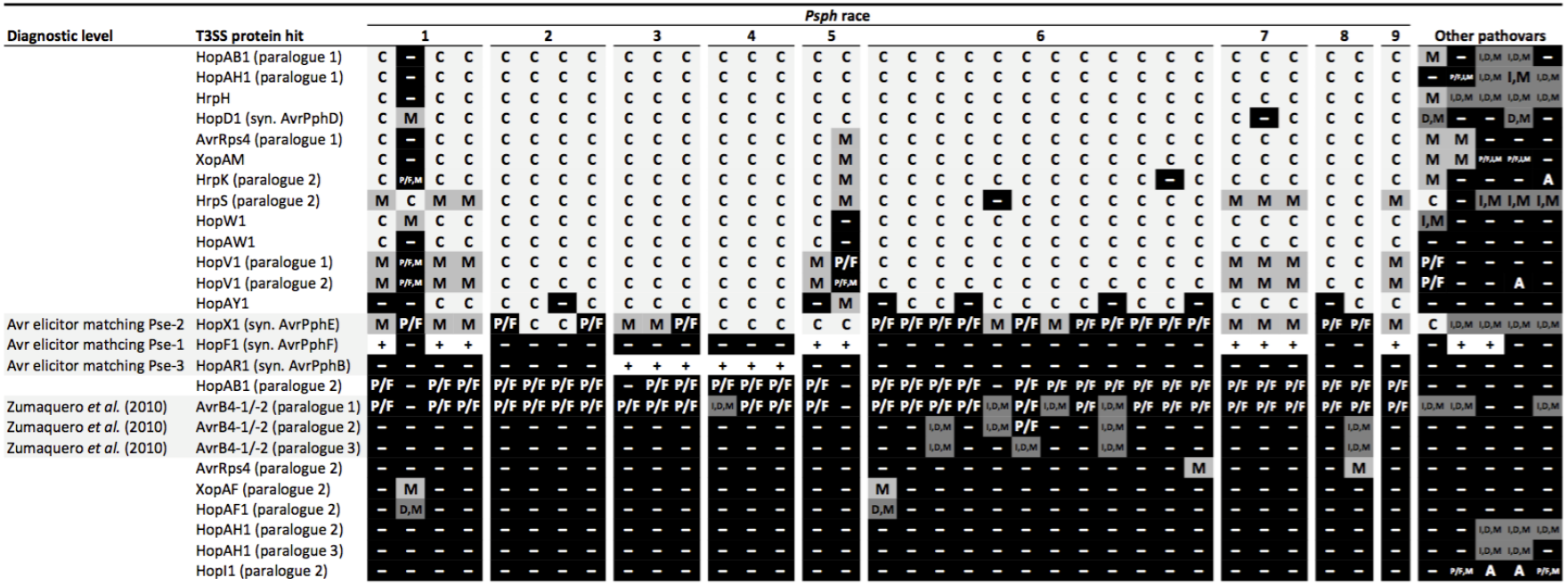
Psph race 6 defines the most broadly virulent pathotype, and has been useful for identifying sources of race-nonspecific resistance to halo blight. Thus, a complementary objective was to determine whether T3SS proteins can be identified that have unique alleles in, or are uniquely absent from, race 6 isolates, as candidate determinants of the broad virulence of race 6. None of the T3SS proteins in the database was found to be uniquely present in all race 6 isolates. However, T3E HopAR1 is present only in isolates from races 3 and 4 (Table 3.2). HopAR1 confers an avirulence phenotype by inducing a severe hypersensitive resistance response in plants possessing the cognate R protein Pse-3. Similarly, T3E HopF1 is present only in isolates from races 1, 5, 7 and 9 (Table 3.2), and triggers the hypersensitive response in plants carrying Pse-1. These findings are consistent with those presented by Jenner *et al*. (1991) on HopAR1 (formerly AvrPphB), and by Tsiamis *et al*. (2000) on HopF1 (formerly AvrPphF). Additionally, Stevens *et al.* (1998) reported

that while homologues of HopX1 (formerly AvrPphE) are present in all *Psph* races, non-synonymous single-nucleotide substitutions in their encoding genes confer virulence towards beans possessing Pse-2 in races 1, 3, 6 and 9. These three avirulence proteins are thus either absent from, or putatively mutated to virulence effectors in, race 6 isolates. Alleles of *hopX1* found in race 6 isolates may therefore provide useful biomarkers for identifying novel sources of race-specific and race-nonspecific halo blight resistance.

Table 3.2. Summary of sequence comparisons of type III secretion system (T3SS) protein hits in 34 pathogenically and geographically diverse isolates of *Pseudomonas syringae* pv. *phaseolicola* (*Psph*) and control isolates of five pathovars of *P. syringae*

		Psph race																																Other pathovars				
		1				2				3			4			5		6													7			8		9					
Diagnostic level	**T3SS protein hit**	1281A	1644R	1816A	725A	1502A	1678	2698A	882	1301A	1567A	2475A	1302A	1334A	2240A	11	1516A	1010	1294	1299A	1308	1448A	1599A	1715A	1769A	2109	2654A	2693A	716B	1314	1354A	1549A	1645	2654C	2732E	*glycinea*	*pisi*	*maculicola*	*tomato*	*syringae*
1. *P. syringae*	HopAJ2 (paralogue 1)	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	M	M
1. *P. syringae*	HopAN1	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	M	M	M
2. *Psph, Psg, Pspi*	HrpA2	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	–	–	–
2. *Psph, Psg, Pspi*	HrpG	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	P/F,I,M	–	–
2. *Psph, Psg, Pspi*	HopI1 (paralogue 1)	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	A	A	M
2. *Psph, Psg, Pspi*	HrpP	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	D,M	–	–
2. *Psph, Psg, Pspi*	HrpL	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	–	–
2. *Psph, Psg, Pspi*	HrpV	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	I,M	–	–
2. *Psph, Psg, Pspi*	HrpD	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	I,D,M	–	–
2. *Psph, Psg, Pspi*	HopAJ1	C	C	C	C	C	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	–	M	–
2. *Psph, Psg, Pspi*	HopAJ2 (paralogue 2)	C	C	C	C	C	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	–	M	–
3. *Psph, Psg*	HrpS (paralogue 1)	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	M	M	M
3. *Psph, Psg*	HrpR (paralogue 2)	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	–	–	M
3. *Psph, Psg*	HrpK (paralogue 1)	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	D,M	I,D,M	I,D,M	D,M
3. *Psph, Psg*	HrcQa	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	I,D,M	I,D,M	–	–
3. *Psph, Psg*	HrpK/HrpW1	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	–	–	–	–
3. *Psph, Psg*	HopAU1	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	–	–	–
3. *Psph, Psg*	HopG1 (paralogue 1)	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	–	–	–	–
3. *Psph, Psg*	HopG1 (paralogue 2)	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	–	–	A	–
3. *Psph, Psg*	HopQ1-2	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	–	–	–	–
3. *Psph, Psg*	HrpZ1	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	D,M	I,D,M	–	D,M
3. *Psph, Psg*	HopAS1	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	–	P/F,I,M	P/F,I,M	–
3. *Psph, Psg*	HrpR (paralogue 1)	M	C	M	M	C	C	C	C	C	C	C	C	C	C	C	M	C	C	C	C	C	C	C	C	C	C	C	C	M	M	M	C	C	M	C	–	–	–	I,M
4. *Psph*	XopAF (paralogue 1)	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	M	M	–	–
4. *Psph*	HopAV1	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	–	–	–	–
4. *Psph*	HopF3	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	–	–	–	–
5. *Psph* races 1, 7, 9	HopQ1	M	M	M	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	M	C	C	M	–	–	–	M	–
6. *Psph* races 1, 5, 7, 9	HopAE1 (paralogue 1)	M	M	M	M	C	C	C	C	C	C	C	C	C	C	M	M	C	C	C	C	C	C	C	C	C	C	C	C	M	M	M	C	C	M	I,M	M	–	–	I,M
6. *Psph* races 1, 5, 7, 9	HopAE1 (paralogue 2)	M	M	M	M	C	C	C	C	C	C	C	C	C	C	M	M	C	C	C	C	C	C	C	C	C	C	C	C	M	M	M	C	C	M	I,M	M	–	–	I,M
7. *Psph* race 6	AvrD1	C	M	C	C	C	C	C	C	C	C	M	M	C	C	C	M	M	M	M	M	C	M	C	M	C	C	M	M	C	C	C	C	C	C	M	–	M	–	–
	HopJ	C	I	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	M	–	–
	HopAF1 (paralogue 1)	C	I	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	M	M	M	M
	HopAK1	C	P/F	C	C	C	C	C	C	C	C	C	C	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	I,D,M	I,D,M	I,D,M	I,D,M
	HopR1	C	P/F,M	C	C	C	C	C	C	C	C	C	C	C	C	C	M	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	M	M	P/F,I,M	P/F,I,M	–
Baltrus *et al.* (2012)	HopM1	C	P/F,D	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	F,I	A	A	A	A
Baltrus *et al.* (2012)	AvrB2 (syn. AvrPphC)	C	–	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	–	–	–	–
	AvrB1	C	–	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	C	–	–	–	–	–

Table 3.2. Continued



+ = homologue present.
C = homologue is fully conserved relative to the alignment consensus sequence.
M = missense mutation(s).
I = non-frameshift insertion(s).
D = non-frameshift deletion(s).
P/F = premature stop codon(s) or frameshift mutation(s).
A = all listed mutation types.
– = homologue not found.

Psph: *Pseudomonas syringae* pv. *phaseolicola*; *Psg*: *Pseudomonas syringae* pv. *glycinea*; *Pspi*: *Pseudomonas syringae* pv. *pisi*. Red text denotes *Psph* isolates with previously assembled genomes (Joardar *et al.*, 2005; Baltrus *et al.*, 2012). Previously assembled genomes for *P. syringae* pathovars *glycinea* strain race 4 (NCBI BioSample accession: SAMN02471825), *maculicola* strain CFBP 1657 (SAMN03328955), *pisi* strain PP1 (SAMN02471574), *syringae* strain B728A (SAMN02604347) and *tomato* strain DC3000 (SAMN02604017) were included for identification and comparison of sequences with homology to T3SS proteins.

Candidate virulence effectors in *Psph* race 6

A set of 39 SNPs was identified as diagnostic of *Psph* race 6 using PPFS2, based on the kSNP3-generated genome-wide SNP matrix for the 34 *Psph* assemblies. Among these, five non-synonymous substitutions were identified as having high probabilities (> 0.9999999) of being casually related to the race 6 phenotype using the PPFS2 program 'CausalSNPs'. The potentially causal role of these variants was determined based on low χ^2 probabilities that their allelic state changed randomly along the same branches of a maximum likelihood tree along which the phenotype changed (Table 3.3; Figure 3.1). Each of these five non-synonymous SNPs was identified as potentially causal using both the parsimony and maximum likelihood phylogenies inferred by kSNP3 to initialise the tree search in MEGA7.

None of the *Psph* isolates belonging to races other than race 6 was incorrectly predicted by PPFS2 to be from race 6. Four isolates previously designated as 'race 6' were predicted not to be from race 6 (Figure 3.1; isolates 1448A, 1715A, 2109 and 2654A). The reliability with which PPFS2 identifies diagnostic and potentially causal SNPs depends on the correct designation of phenotype-positive and phenotype-negative strains. The developer of PPFS2 advises that the accuracy with which a set of diagnostic SNPs predicts the phenotypes of strains of unknown phenotype should be ≥ 90% in order for the potentially causal SNPs to be considered valid (Hall, 2014a,b). The 39 diagnostic SNPs predicted the phenotypes of all of the 34 *Psph* isolates (12 designated as race 6 and 22 designated as not from race 6) with 88.2% accuracy. Accuracy increased to 96.7% when the analysis was repeated with the same settings, but with isolates 1448A, 1715A, 2109 and 2654A designated as being of unknown phenotype. Race 3 isolate 2475A (carrying HopAR1) was incorrectly predicted to be from race 6. However, this prediction is consistent with the close phylogenetic relationship between isolate 2475A and some race 6 isolates, as depicted in Figures 3.2 and 3.3. This analysis identified 22 diagnostic SNPs. Among these, the five non-synonymous substitutions that were identified as potentially causal by the original analysis (Table 3.3) were identified as having high probabilities (> 0.9999999999999) of being causally related to the race 6 phenotype. This analysis did not identify any novel potentially causal SNPs not identified by the original analysis.

Table 3.3. Summary of non-synonymous substitutions identified by PPFS2 analyses with the lowest χ^2 probabilities that their allelic state changed randomly along the same branches of a maximum likelihood tree along which the phenotypes of *Pseudomonas syringae* pv. *phaseolicola* strains changed to race 6

SNP position	Location	Locus tag	Predicted functional annotation	Base change	Amino acid change	Parsimony			Maximum likelihood		
						χ^2	d.f.	*P*	χ^2	d.f.	*P*
97,404	Plasmid A	PSPPH_RS26985	AvrD1 (syringolide synthesis)	A383G	Q128R	40.67	1	1.81E−10	36.27	2	1.33E−08
1,508,902	Chromosome	PSPPH_1296	Type III effector HopX1	C208T	Q70X	40.67	1	1.81E−10	36.27	2	1.33E−08
1,744,233	Chromosome	PSPPH_RS07590	Short-chain dehydrogenase	G589T	A197S	40.67	1	1.81E−10	36.27	2	1.33E−08
4,473,029	Chromosome	PSPPH_RS19790	D-alanine:D-alanine ligase A	G463T	A155S	40.67	1	1.81E−10	36.27	2	1.33E−08
116,931	Plasmid A	PSPPH_RS27065	Hypothetical protein	A3550G	G1184R	40.67	1	1.81E−10	36.27	2	1.33E−08
1,777,757	Chromosome	PSPPH_RS27065	Hypothetical protein	A3550G	G1184R	40.67	1	1.81E−10	36.27	2	1.33E−08

Variant sites and annotations were deduced from the reference genome and corresponding GenBank record for *Pseudomonas syringae* pv. *phaseolicola* isolate 1448A. Protein annotations were predicted by automated computational analysis based on protein homology by Joardar *et al*. (2005). Each listed substitution is predicted to be located within a protein-coding sequence. χ^2 statistics correspond to separate analyses performed using each of the parsimony and maximum likelihood trees estimated by kSNP3 to initialise the maximum likelihood tree search in MEGA7. d.f.: degrees of freedom.

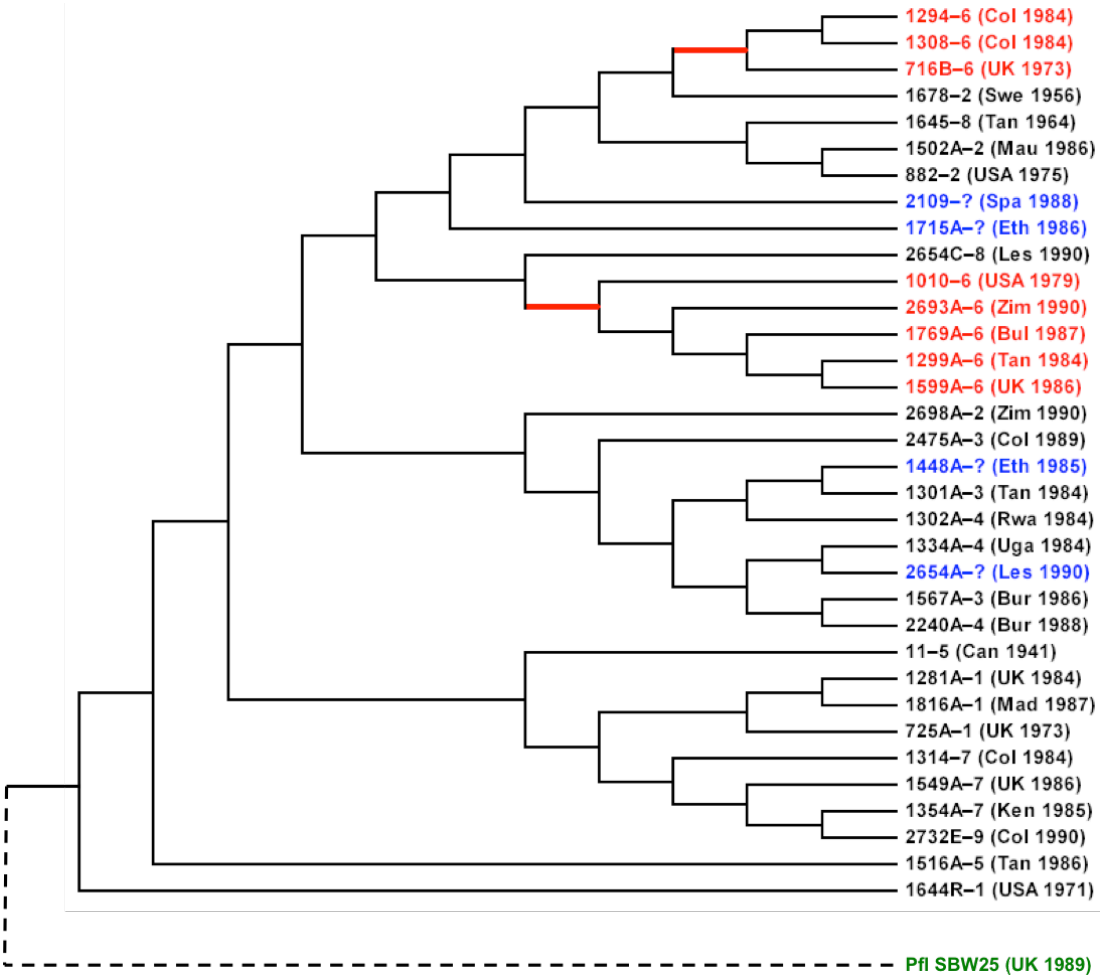


Figure 3.1. Maximum likelihood tree inferred from a PPFS2-generated matrix of 39 SNPs diagnostic of race 6 and a random sample of 975 SNPs from the kSNP3-generated full-genome SNP matrix for 34 isolates of *Pseudomonas syringae* pv. *phaseolicola* using MEGA7. The parsimony tree estimated by kSNP3 was used to initialise the tree search. Red text denotes isolates predicted by PPFS2 to be from race 6. Blue text denotes isolates previously designated as 'race 6' but which were predicted by PPFS2 not to be from race 6. Branches along which the phenotype changed to race 6 are indicated in red. The tree was virtually rooted using the outgroup *P. fluorescens* strain (*Pfl* SBW25; dashed line, green text). Where known, race designation and country and year of collection for each isolate are indicated (Bul: Bulgaria; Bur: Burundi; Can: Canada; Col: Colombia; Eth: Ethiopia; Ken: Kenya; Les: Lesotho; Mad: Madagascar; Mau: Mauritius; Rwa: Rwanda; Spa: Spain; Swe: Sweden; Tan: Tanzania; Uga: Uganda; Zim: Zimbabwe).

Analyses performed using both the statistical GWA tool PPFS2 and multiple-sequence alignments of predicted T3E proteins present in all *Psph* races revealed potentially causal non-synonymous substitutions within two candidate determinants (*avrD* and *hopX1*) of the race 6 virulence phenotype (Tables 3.3 to 3.5).

avrD (syn. *avrD1*) is a unique avirulence gene that encodes an enzyme which directs the synthesis of bacterial metabolites known as syringolides, which are low-molecular-weight *C*-glycosyl lipids that elicit a hypersensitive response from

soybean cultivars carrying *R* gene *Rpg4* (Keen *et al*., 1990; Smith *et al*., 1993; Midland *et al*., 1993, 1995; Yucel *et al*., 1994a,b). Non-functional *avrD* alleles occur in seven races of *P. syringae* pv. *glycinea* (Keith *et al*., 1997). Point mutations relative to the functional avirulence allele in some *P. syringae* pv. *tomato* strains are summarised in Table 3.4. Keith *et al.* (1997) proposed that these missense and nonsense mutations represent different routes to virulence taken by *P. syringae* pv. *glycinea* to evade detection by host soybean plants possessing *Rpg4*. Yucel and Keen (1994) showed by site-directed mutagenesis that three amino acid changes (cysteine 19 to arginine, alanine 280 to valine, and leucine 304 to serine) are critical for restored avirulence function in *P. syringae* pv. *glycinea*. Additionally, a mutation from leucine 301 to phenylalanine is required for enhanced avirulence activity.

Importantly, the 36 *Psph* isolates examined here possess all four of the residues required for avirulence activity on *Rpg4*-bearing soybeans. In *Psph*, the diagnostic missense mutation (g.383A→G transition) causes a glutamine 128 to arginine change in eight of the 12 isolates designated as race 6, as well as in race 3 isolate 2475A (Table 3.4). The four exceptional isolates originally designated as 'race 6' are those that were predicted by PPFS2 not to be from race 6, and have the glutamine residue found in other races.

AvrD was predicted to be encoded in the genomes of six out of nine *P. syringae* pv. *tomato* strains available from the Pseudomonas Genome Database (http://www.pseudomonas.com; Winsor *et al*., 2016). The predicted protein is identical in all six strains. However, AvrD was not found in *P. syringae* pv. *tomato* strains DC3000 (SAMN02604017), ICMP2844 (SAMN03976246) or NCPPB 1108 (SAMN02472091). Where present, the protein (311 amino acids) is more conserved within this pathovar than in *Psph* (reported to be active in syringolide elicitor production in isolates 3121 and G50; Keen *et al*., 1990; Yucel *et al*., 1994a) and in *P. syringae* pv. *glycinea* (reported to be non-functional; Yucel *et al*., 1994a; Keith *et al*., 1997). AvrD in *P. syringae* pv. *tomato* shares 95.2% amino acid identity with its homologue in *P. syringae* pv. *maculicola* strain CFBP 1657 (SAMN03328955). Collectively, these seven isolates (AvrD homology class I, as designated by Yucel *et al*., 1994a) differ from isolates of pathovars *phaseolicola* and *glycinea* (homology class II) at 60 amino acid positions of AvrD. Only those positions at which variation was detected amongst *Psph* (potentially harbouring functional $avrD^{p}$ allele[s]) and *P. syringae* pv. *glycinea* (harbouring non-functional $avrD^{g}$ alleles) and at which critical

residues are located are shown in Table 3.4 in view of their potential impact on protein function.

None of the deduced amino acid sequences of AvrD variants represented in Table 3.4 was predicted to possess an N-terminal type III secretion signal peptide characteristic of a T3E with translocation potential, as determined by the EffectiveT3 Naïve Bayes model 2.0.1 (Eichinger *et al*., 2016). This is consistent with the reported absence of a signal peptide sequence from AvrD in *P. syringae* pv. *tomato* (Keen *et al*., 1990) and is unsurprising in view of the atypical role of this protein in conferring avirulence.

Table 3.4. Amino acid sequence variations encoded by *avrD* alleles present in 36 isolates of *Pseudomonas syringae* pv. *phaseolicola*, isolates representing seven races of *P. syringae* pv. *glycinea*, and six isolates of *P. syringae* pv. *tomato*

		Position																
Allele[a]	**Isolate–race**	**19**	**41**	**44**	**66**	**128**	**161**	**166**	**168**	**226**	**243**	**245**	**250**	**280**	**283**	**294**	**301**	**304**
avrD$^{p1–2}$	11–5, 725A–1, 882–2, 1281A–1, 1301A–3, 1314–7, 1334A–4, 1354A–7, 1448A–6, 1502A–2, 1549A–7, 1567A–3, 1645–8, 1678–2, 1715A–6, 1816A–1, 2109–6, 2240A–4, 2654A–6, 2654C–8, 2698A–2, 2732E–9	R	E	A	H	Q	D	N	Y	H	G	S	M	V	K	G	F	S
avrD^{p6}	716B–6, 1010–6, 1294–6, 1299A–6, 1308–6, 1599A–6, 1769A–6, 2475A–3, 2693A–6	R	E	A	H	R	D	N	Y	H	G	S	M	V	K	G	F	S
avrD^{p1}	3121–1	R	E	R	H	K	D	S	Y	H	D	S	M	V	K	G	F	S
avrD^{p2}	G50–2	R	E	R	H	Q	D	N	Y	H	G	S	V	V	K	G	F	S
avrD^{p4}	1302A–4	R	E	A	H	Q	D	N	C	H	G	S	M	V	K	G	F	S
avrD^{p5}	1516A–5	R	A	A	H	Q	D	N	Y	H	G	S	M	V	K	G	F	S
avrD$^{p1–3}$	1644R–1	R	A	A	H	Q	G	N	Y	H	G	S	M	V	K	G	F	S
avrD^{g0}	Pathovar *glycinea* race 0	R	A	A	R	Q	D	N	Y	H	G	S	M	V	N	*	–	–
avrD^{g1}	Pathovar *glycinea* race 1	R	A	A	H	Q	D	N	Y	Y	G	S	M	V	N	G	F	S
avrD$^{g2/4/5}$	Pathovar *glycinea* races 2, 4 and 5	C	A	A	H	Q	D	N	Y	H	G	L	M	A	K	G	L	L
avrD$^{g3/6}$	Pathovar *glycinea* race 3 and 6	H	A	A	H	Q	D	N	Y	H	G	S	M	A	K	G	L	S
avrD^{t}	Pathovar *tomato*: 6 isolates	R	A	A	H	Q	D	S	Y	H	D	S	M	V	K	G	L	S

[a]An alphanumeric suffix in superscript is given for each allele name, denoting the pathovar and race in which the allele was identified, as proposed by Keith *et al*. (1997). *avrD*^{p2} in isolate G50 and all *avrD*$^{g\#}$ (pathovar *glycinea*) alleles were reported in Keith *et al*. (1997). *avrD*^{p1} in isolate 3121 was reported in Yucel *et al*. (1994a). *avrD*^{t} (encoding protein with RefSeq accession WP_007247583) was found in *P. syringae* pv. *tomato* strains A9 (SAMN04256406), 407 (SAMN04256407), K40 (SAMN02472092), Max13 (SAMN02472093), NYS-T1 (SAMN03093524) and T1 (SAMN02472090).

T3E HopX1 (or its absence) is another candidate determinant of the broad virulence of *Psph* race 6. A non-synonymous g.208C→T substitution results in a premature translation termination site at amino acid 70 in eight of the 12 isolates originally designated as race 6 and in race 3 isolate 2475A (*hopX1*^{p6}; Table 3.5). Each of these nine isolates also possesses allele *avrD*^{p6}. Avirulence effector HopX1 is matched by cognate bean R protein Pse-2 (Mansfield *et al*., 1994), which confers resistance to races 2, 3, 4, 5, 7, 8 and 9 (Miklas *et al*., 2011). Other single-nucleotide substitutions identified within the coding sequence of *hopX1* in different *Psph* isolates and races have been reported to render these alternate alleles determinants of virulence (Stevens *et al*., 1998). Some of these alleles were shown to encode functional effectors, with quantitative variation in plant reactions to expression of different alleles attributed to variable receptor–effector binding affinity.

The truncated allele *hopX1*^{p6} is unlikely to be functional in encoding a virulence effector that traverses the T3SS, despite containing an alternate start codon at amino acid 155. From this alternate start codon to the translation termination site at amino acid 381, *hopX1*^{p6} is identical to the functional avirulence allele *hopX1*^{p4}. This suggests that the predicted full-length protein is required for avirulence function. Moreover, the amino acid translations of *hopX1*^{p6}, *hopX1*$^{p6–3}$ (isolate 2109) and *hopX1*$^{p1–2}$ (isolate 1644R) were not predicted to possess an N-terminal type III secretion signal peptide (Table 3.5), as determined by the EffectiveT3 Naïve Bayes model 2.0.1 (Eichinger *et al*., 2016). The amino acid composition within the first 154 residues of other HopX1 variants therefore appears to be critical in conferring T3E secretion potential. As summarised by Arnold *et al*. (2009: 2011), to have translocation potential, "the first 50 amino acids of the effector N-terminal region should contain > 10% serines, an aliphatic aa (isoleucine, leucine, valine, alanine or proline) in the 3rd or 4th position and no acidic aa (aspartic acid or glutamic acid) in the first 12 aa."

Table 3.5. Amino acid sequence variations encoded by *hopX1* alleles present in 34 isolates of *Pseudomonas syringae* pv. *phaseolicola*

Allele[a]	Isolate–race	Avr function[b]	T3E[c]	Position													
				38	69	70	155	170	176	191	235	259	305	310	343	374	381
hopX1^{p4}	11–5, 1302A–4,1334A–4, 1678–2, 2240A–4, 2698A–2	Yes	Yes	Q	G	Q	M	S	A	G	W	Y	G	E	G	K	*
hopX1^{p1}	725A–1, 1281A–1, 1816A–1	No	Yes	Q	G	Q	M	S	A	R	W	Y	G	E	G	K	*
hopX1^{p3}	1301A–3, 1448A–6	No	Yes	Q	G	Q	M	S	A	G	W	Y	G	K	G	K	*
hopX1$^{p3–2}$	1567A–3	Unknown	Yes	Q	G	Q	M	L	A	G	W	Y	G	E	G	K	*
hopX1^{p5}	1516A–5	Yes	Yes	Q	G	Q	M	S	A	G	W	Y	G‡	E	G	K	*
hopX1$^{p6–2}$	1715A–6	Unknown	Yes	Q	G	Q	M	S	V	G	W	Y	G	E	G	K	*
hopX1$^{p7–2}$	1354A–7	Unknown	Yes	Q	G	Q	M	S	A	G	W	Y	G	E	D	K	*
hopX1^{p9}	1314–7, 1549A–7, 2732E–9	No	Yes	Q	G	Q	M	S	A	G	L	Y	G	E	G	K	*
hopX1^{p8}	882–2, 1502A–2, 1645–8, 2654C–8	No	Yes	Q	G	Q	M	S	A	G	W	Y	G	E	G	Insertion	
hopX1$^{p1–2}$	1644R–1	Unknown	No	Q	Frameshift		M	S	A	G	W	Y	G	E	G	K	*
hopX1$^{p6–3}$	2109–6	Unknown	No	Frameshift			M	S	A	G	W	Y	G	E	G	K	*
hopX1$^{p6–4}$	2654A–6	Unknown	Yes	Q	G	Q	M	S	A	G	W	*	–	–	–	–	–
hopX1^{p6}	716B–6, 1010–6, 1294–6, 1299A–6, 1308–6, 1599A–6, 1769A–6, 2475A–3, 2693A–6	Unknown	No	Q	G	*	M	S	A	G	W	Y	G	E	G	K	*

[a]An alphanumeric suffix in superscript is given for each allele name, denoting the pathovar and race in which the allele was identified. *hopX1*^{p1} = *avrPphE1*; *hopX1*^{p3} = *avrPphE6*; *hopX1*^{p4} = *avrPphE4* (functional avirulence gene); *hopX1*^{p5} = *avrPphE5* (functional avirulence gene); *hopX1*^{p8} = *avrPphE8*; *hopX1*^{p9} = *avrPphE9* (as inferred from Stevens *et al*., 1998).

[b]Avirulence (Avr) function of each allele is listed according to experimental evidence reported by Stevens *et al*. (1998).

[c]T3E refers to the predicted presence (with high confidence) or absence of an N-terminal type III secretion signal peptide in each HopX1 variant, as determined by the EffectiveT3 Naïve Bayes model 2.0.1 (Eichinger *et al*., 2016; http://www.effectivedb.org/method/effectivet3).

‡*hopX1*^{p5} (formerly *avrPphE5*) contains a synonymous g.915C→T substitution, which does not abolish avirulence on beans carrying *Pse-2* (Stevens *et al*., 1998).

Other candidate determinants of *Psph* race 6 virulence

A possible role for the short-chain dehydrogenase (locus tag PSPPH_RS07590; Table 3.3) variant present in race 6 might be in microbial quorum sensing, or in preventing quorum quenching, which control expression of virulence factors and biofilm formation in pseudomonads. For example, Bijtenhoorn *et al.* (2011) reported that *Pseudomonas aeruginosa* expressing *bpiB09* (encoding an NADP-dependent short-chain dehydrogenase) exhibited impaired motility, biofilm formation and quorum sensing-dependent pyocyanin toxin production, and caused reduced paralysis of the nematode *Caenorhabditis elegans*.

Another candidate determinant of race 6 virulence is a variant of D-alanine:D-alanine ligase A (PSPPH_RS19790). Antibiotics target enzymes within this functional class, which have roles in cell wall biosynthesis. Modified ligases producing D-alanine:D-lactate in enterococcal animal pathogens confer resistance to vancoymcin through reduced binding affinity (Lessard and Walsh, 1999; Kuzin *et al*., 2000). It is therefore conceivable that the mutated variant in *Psph* race 6 might play a role in conferring resistance to plant antimicrobial compounds.

Development of diagnostic markers

Two cleaved amplified polymorphic sequence (CAPS) markers targeting the substitutions within *avrD*^{p6} and *hopX1*^{p6} potentially causally related to the *Psph* race 6 phenotype have been designed (Table 3.6). Each SNP introduces either a six-base or five-base recognition site for a restriction enzyme that is expected to cleave PCR amplicons at the variant position. However, the restriction enzyme targeting the *hopX1*^{p6} SNP allele (Hpy99XIII) is not widely available, and so Sanger sequencing of amplicons may be necessary. Although these markers have yet to be tested, it is anticipated that they will provide tools with which to further test the validity of the finding that these alleles are diagnostic of *Psph* race 6. If validated, these markers could serve as useful molecular diagnostic tools for detection of a broadly virulent pathogen, and for identifying novel sources of race-specific and race-nonspecific halo blight resistance.

Table 3.6. Markers for genotyping of alleles diagnostic of *Pseudomonas syringae* pv. *phaseolicola* race 6

Marker name	Primers (5′–3′)[a]	Physical position[b]	Location	Annealing temp. (°C)[c]	Amplicon size (bp)	SNP position[d]	Restriction enzyme[e]	Race 6 expected fragments (bp)
avrD^{p6}	TTGGGACCCGCTAAAGATCG	97,750	Large plasmid (NC_007274)	55	424	347 (allele G)	AvaI / BsoBI	344 & 80
	GAGCGGCGGGAATGATAAAC	97,327						
hopX1^{p6}	ATCATCATCGGCCCAGAACC	1,508,769	Chromosome (NC_005773)	55	372	134 (allele T)	Hpy99XIII[f]	136 & 236
	CGCTCGGCATCTTTTCTCAAG	1,509,140						

[a]Primers were designed using Geneious version 9.1.4 (Kearse *et al*., 2012), specifying optima of 20 bases in length, 60 °C melting temperature and 50% GC content, as well as a terminal 3′ G or C base. Primer specificity to the intended template was confirmed by comparing primer pairs with the reference genome for *Pseudomonas syringae* pv. *phaseolicola* isolate 1448A (RefSeq accessions NC_005773, NC_007274 and NC_007275) using the National Center for Biotechnology Information Primer-BLAST tool (https://www.ncbi.nlm.nih.gov/tools/primer-blast/).

[b]The physical position of each primer refers to the genomic coordinate of the 5′-most base.

[c]Optimum annealing temperature was calculated using the New England BioLabs Tm Calculator version 1.9.4 (http://tmcalculator.neb.com/) and assumes use of standard Taq polymerase and buffer, and 0.4 µM of each primer.

[d]SNP position refers to the location of the missense substitution within the expected PCR amplicon.

[e]Restriction enzymes for use in cleaved amplified polymorphic sequence (CAPS) marker assays were identified using the command-line version of SNP2CAPS (Thiel *et al*., 2004), as modified at http://radish.plantbiology.msu.edu/uploaded/0/05/SNP2CAPS_mod.txt .

[f]Hpy99XIII is not widely available.

Phylogenomic analysis reveals racial clades within *Psph*

Figures 3.2 and 3.3 depict the phylogenetic distribution of *Psph* isolates, which cluster primarily by race and, to a lesser extent, by region of collection. Omission of the outgroup *P. fluorescens* strain SBW25 and other *P. syringae* pathovars resulted in an altered tree topology with regard to the distribution of clades (Figure 3.3). However, clustering of *Psph* isolates remained comparable.

The wide geographic distribution of race 6 isolates, which lack a functionally characterised avirulence gene, is reflected in their broad phylogenetic distribution. Collected in Colombia in 1989, isolate 2475A is genetically more similar to race 6 than to its race 3 counterparts, including at *avrD* and *hopX1* (Tables 3.4 and 3.5). Its phylogenetic proximity to race 6 isolates also collected in Colombia suggests that the presence of *hopAR1* (Table 3.2) may be the result of horizontal gene transfer from a strain of East or Central African origin, possibly after dispersal on internationally traded seed. *Psph* races 1, 5, 7 and 9 may be the most proximal to the ancestral phenotype of the pathovar. The genetic similarities between these four races are also illustrated by the shared absence of sequences present in races 2, 3, 4, 6 and 8 (Figures 3.4 to 3.6). The level of sequence conservation across the pathovar is considerably lower with regard to the small plasmid of *Psph* 1448A compared with the chromosome and large plasmid.

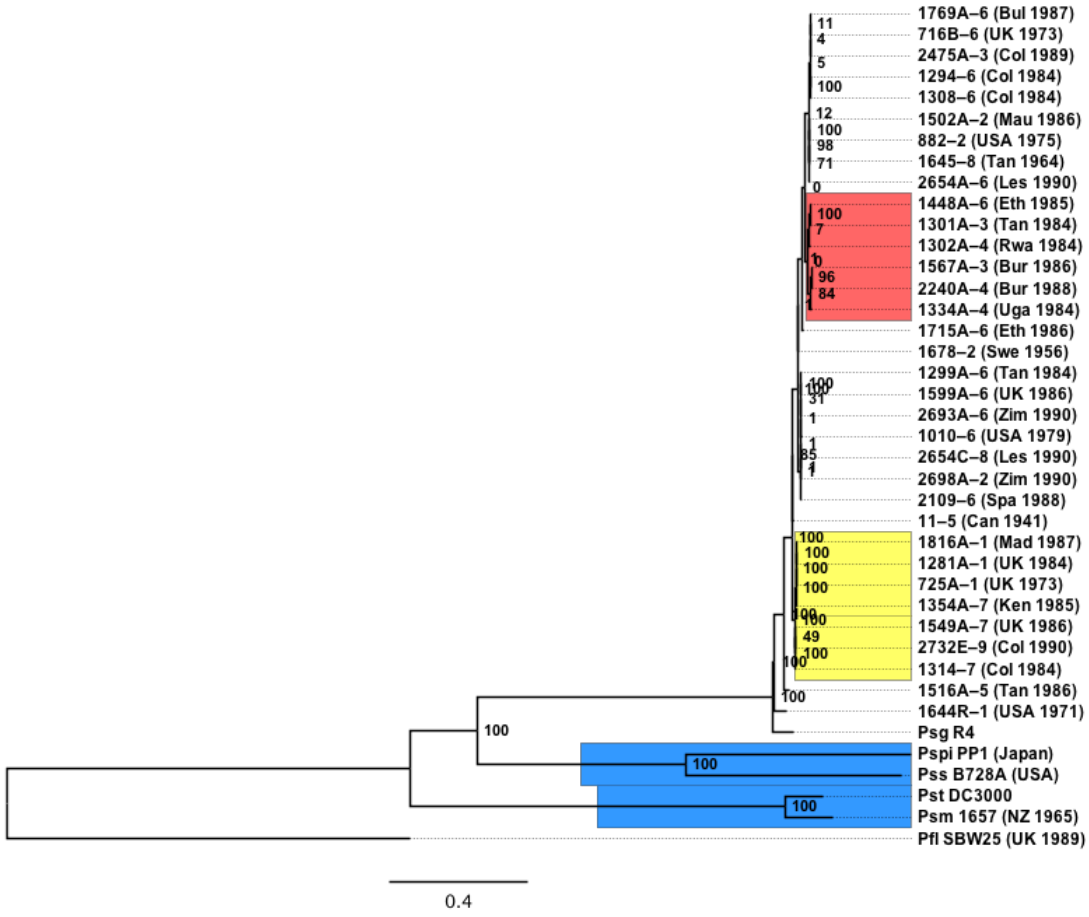


Figure 3.2. Full-genome SNP-based maximum likelihood phylogeny of 34 isolates of *Pseudomonas syringae* pathovar *phaseolicola* (*Psph*), *P. syringae* pathovars *glycinea* strain race 4 (*Psg* R4), *maculicola* strain CFBP 1657 (*Psm* 1657), *pisi* strain PP1 (*Pspi* PP1), *syringae* strain B728A (*Pss* B728A) and *tomato* strain DC3000 (*Pst* DC3000), and the outgroup *P. fluorescens* strain SBW25 (*Pfl* SBW25). Where known, race designation and country and year of collection for each isolate are indicated (Bul: Bulgaria; Bur: Burundi; Can: Canada; Col: Colombia; Eth: Ethiopia; Ken: Kenya; Les: Lesotho; Mad: Madagascar; Mau: Mauritius; NZ: New Zealand; Rwa: Rwanda; Spa: Spain; Swe: Sweden; Tan: Tanzania; Uga: Uganda; Zim: Zimbabwe). SPAdes was used for *de novo* assemblies of the 32 newly sequenced *Psph* genomes. RAxML was used to infer the maximum likelihood tree under the general time reversible–gamma model from a kSNP3-generated pan-genome SNP matrix. Branch support values are bootstrap percentages calculated by RAxML (650 replicates) and are shown as node labels for legibility. The tree was rooted with the outgroup *P. fluorescens* strain SBW25. *Psph* isolates from races 1, 7 and 9 are highlighted in yellow. *Psph* isolates from races 3 and 4 collected in East Africa are highlighted in red. With the exception of *P. syringae* pv. *glycinea* strain race 4, other *P. syringae* pathovars are highlighted in blue.

Figure 3.3. Full-genome SNP-based maximum likelihood phylogeny of 34 isolates of *Pseudomonas syringae* pathovar *phaseolicola* (*Psph*). Where known, race designation and country and year of collection for each isolate are indicated (Bul: Bulgaria; Bur: Burundi; Can: Canada; Col: Colombia; Eth: Ethiopia; Ken: Kenya; Les: Lesotho; Mad: Madagascar; Mau: Mauritius; Rwa: Rwanda; Spa: Spain; Swe: Sweden; Tan: Tanzania; Uga: Uganda; Zim: Zimbabwe). SPAdes was used for *de novo* assemblies of the 32 newly sequenced *Psph* genomes. RAxML was used to infer the maximum likelihood tree under the general time reversible–gamma model from a kSNP3-generated pan-genome SNP matrix. Branch support values are bootstrap percentages calculated by RAxML (550 replicates) and are shown as branch labels. The tree was rooted with *Psph* race 1 isolate 1644R (Baltrus *et al.*, 2012) based on its phylogenetic proximity to the outgroup *P. fluorescens* strain SBW25, as shown in Figure 3.2. *Psph* isolates from races 1, 7 and 9 are highlighted in yellow. *Psph* isolates from races 3 and 4 collected in East Africa are highlighted in red.

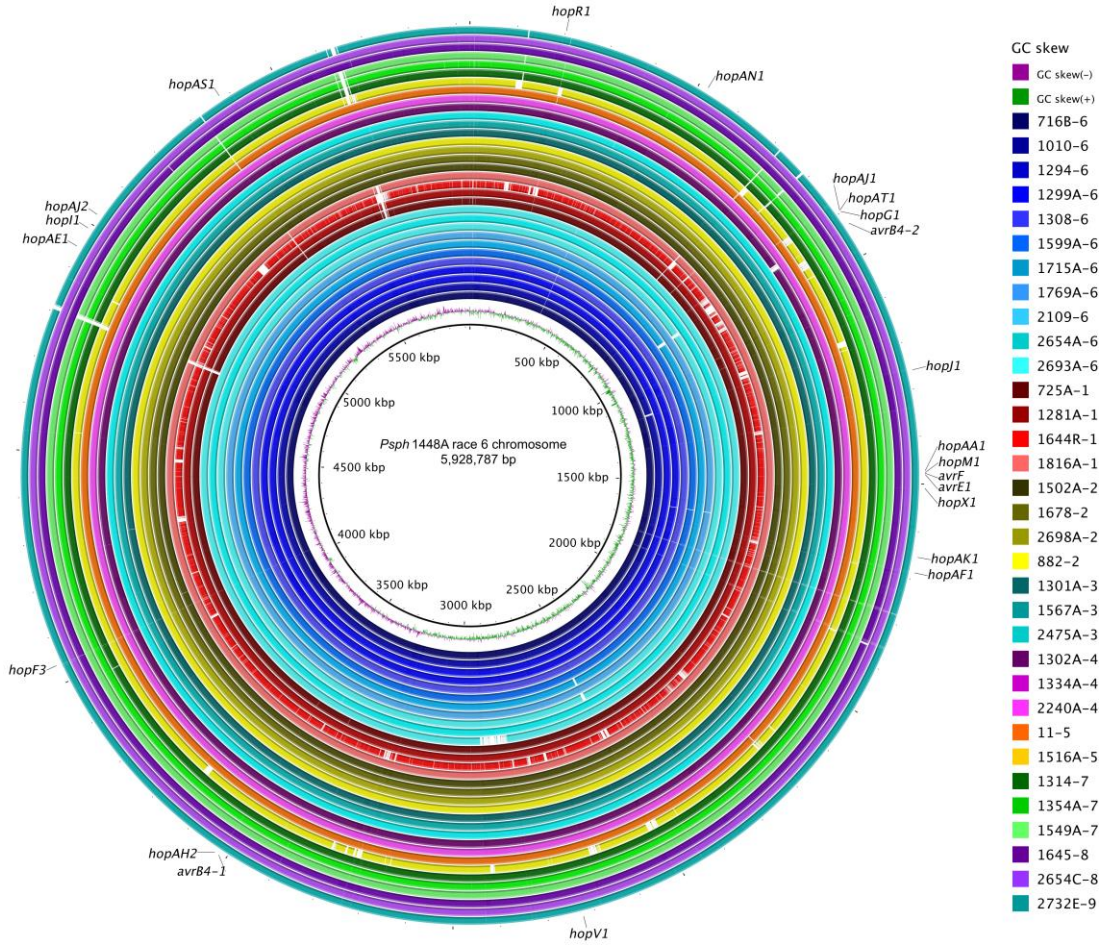


Figure 3.4. Summary of chromosomal DNA sequence conservation amongst *Pseudomonas syringae* pv. *phaseolicola* isolates and races. *De novo*-assembled contigs were aligned to the chromosome of reference isolate 1448A by BLASTN comparison with an E-value threshold of 1×10^{-6} using BLAST Ring Image Generator (BRIG; Alikhan *et al*., 2011). Labels for effector-encoding genes were retrieved from GenBank annotation files for isolate 1448A.

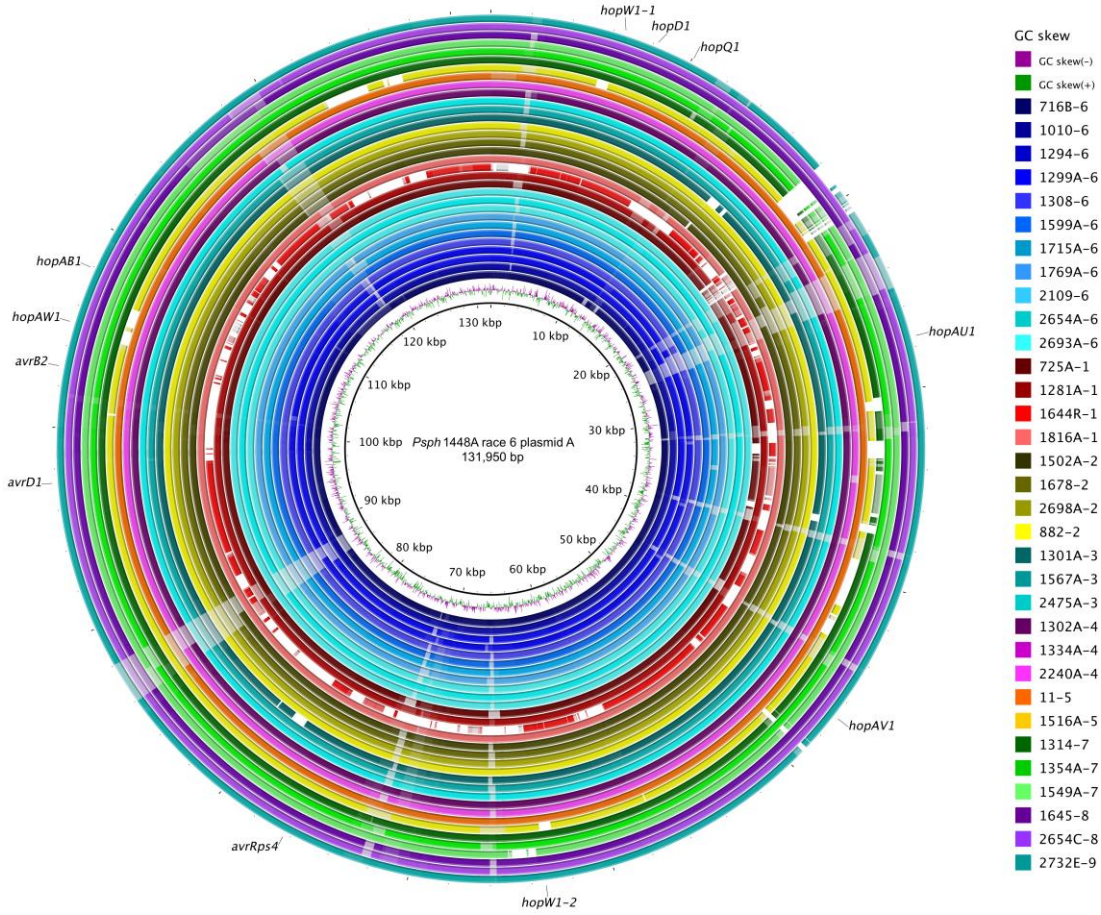


Figure 3.5. Summary of plasmid DNA sequence conservation amongst *Pseudomonas syringae* pv. *phaseolicola* isolates and races. *De novo*-assembled contigs were aligned to the large plasmid of reference isolate 1448A.

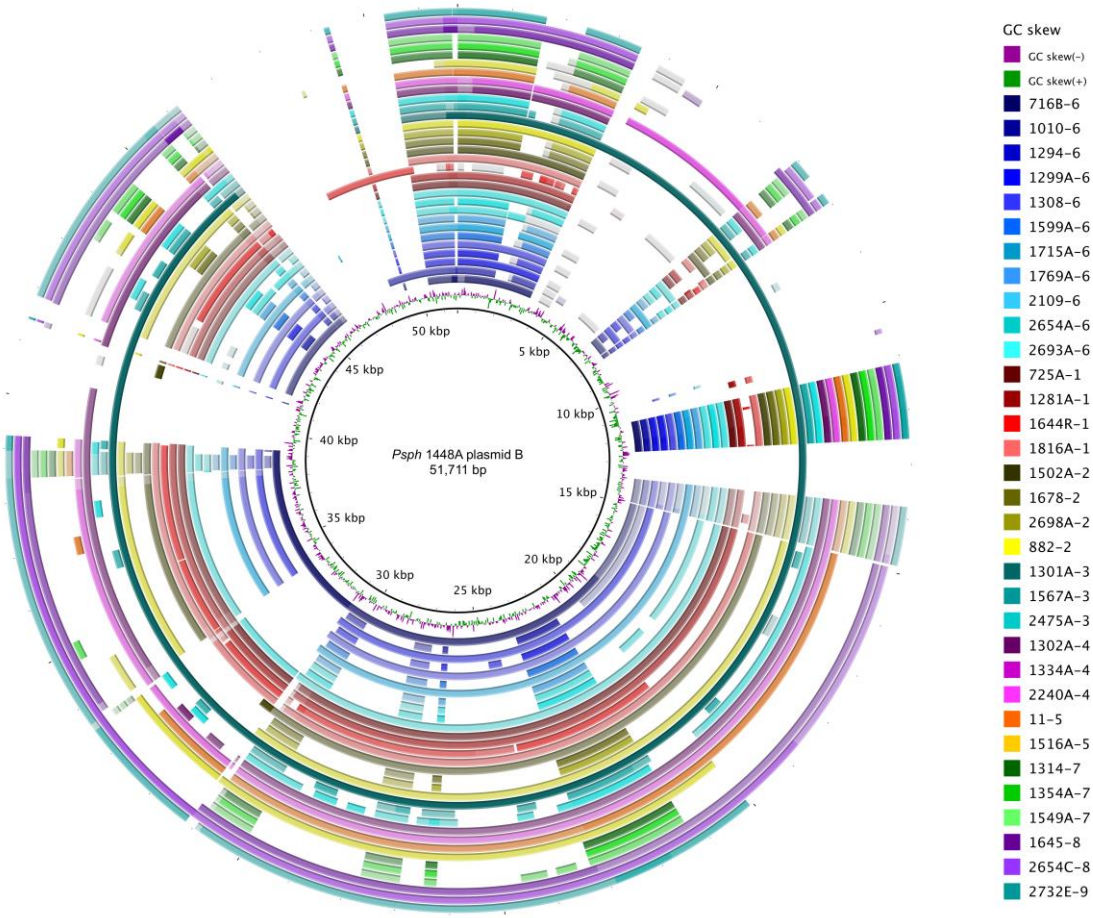


Figure 3.6. Summary of plasmid DNA sequence conservation amongst *Pseudomonas syringae* pv. *phaseolicola* isolates and races. *De novo*-assembled contigs were aligned to the small plasmid of reference isolate 1448A.

3.4. DISCUSSION

Candidate targets for race-nonspecific resistance

This research has applied whole-genome, next-generation sequencing to identify candidate determinants of avirulence that are ubiquitous and highly conserved within *Pseudomonas syringae* pv. *phaseolicola* (Tables 3.2 and A7.1, APPENDIX 7). These T3SS proteins represent candidate targets for quantitative, potentially race-nonspecific, resistance in common bean. Consequently, they may help to elucidate the molecular basis of plant–microbe interactions that govern potentially durable, pathovar-level resistance in this pathosystem.

It is tempting to speculate that reported differences in the level of resistance observed in PI 150414 plants following inoculation with different *Psph* isolates and races might have a basis in allelic variation at an effector present in all races. Taylor *et al*. (1996b) reported disease severity scores of 1.0 following inoculation with race 1 isolate 1281A (*avrD*$^{p1–2}$, *hopX1*^{p1}, *hopF1*) or race 2 isolate 882 (*avrD*$^{p1–2}$, *hopX1*^{p8}), 2.0 with race 4 isolate 1302A (*avrD*^{p4}, *hopX1*^{p4}, *hopAR1*), and 3.0 with race 3 isolate 1301A (*avrD*$^{p1–2}$, *hopX1*^{p3}, *hopAR1*) or race 6 isolate 1299A (*avrD*^{p6}, *hopX1*^{p6}). For example, variable affinity of binding between variants of effector HopX1 and its receptor has been proposed to underlie variation in defence responses elicited from bean plants (Stevens *et al*., 1998). Alternatively, phenotypic variation exhibited by PI 150414 plants in response to different *Psph* isolates may be attributable to different interactions between variable repertoires of effector-encoding alleles. For example, the absence of functional avirulence effectors from race 6 isolates (e.g., HopF1, HopAR1, HopX1) or race 3 isolates (HopF1, HopX1) might result in enhanced hypersensitive response suppression activity of functional effectors. By the same token, the presence of functional avirulence effectors in race 1 isolates (HopF1), race 2 isolates (HopX1) and race 4 isolates (HopX1, HopAR1) might have the inverse effect.

A prominent example of *Psph* effector interdependencies that have quantitative effects on interaction phenotypes exhibited by bean plants is the *hopF1–avrB2–hopAB1* pathogenicity island. Jackson *et al.* (1999) reported that deletion of an indigenous 157-kb plasmid harbouring this pathogenicity island from race 7 strain 1449B conferred avirulence on formerly susceptible bean cultivar Canadian Wonder,

which is not predicted to possess any of the characterised halo blight *R* genes (Taylor *et al.*, 1996a,b; Miklas *et al.*, 2009). Re-introduction of *hopAB1* (syn. *virPphA*) resulted in only partially restored virulence, eliciting a delayed hypersensitive response after the appearance of water-soaking on pods of Canadian Wonder. By contrast, a transconjugant of the plasmid-cured strain expressing *hopF1* (syn. *avrPphF*) elicited a rapid hypersensitive response from leaves and pods of Canadian Wonder (Tsiamis *et al*., 2000). Virulence on Canadian Wonder was fully restored in transconjugants of the plasmid-cured strain harbouring a genomic clone comprising *avrD*, *hopF1*, *avrB2* (syn. *avrPphC*) and *hopAB1*. *avrB2* was found to mask the avirulence activity of *hopF1* on Canadian Wonder, suppressing a hypersensitive response (Tsiamis *et al*., 2000).

Gene-for-gene models based on qualitative interaction phenotypes have provided the basis upon which to develop our understanding of molecular plant–microbe interactions (Flor, 1942, 1971). However, effector interactions and the partial restoration of virulence suggest that different effector targets in the plant may contribute quantitatively to the resistance response (Jackson *et al*., 1999). Thus, some of the approaches used in the above example involving trans-complementation of plasmid-cured strains may provide useful means of dissecting the molecular basis of quantitative phenotypic variation exhibited by PI 150414 plants in response to inoculation with different *Psph* isolates.

In view of the potentially broad-spectrum nature of the resistance derived from PI 150414, characterisation of the phenotypes of resistant bean plants following inoculation with other *P. syringae* pathovars could shed more light on the nature of this resistance. Phenotypes comparable to those exhibited by resistant plants following inoculation with *Psph* might indicate that a species-level resistance mechanism effective against all *P. syringae* pathovars could underlie these symptoms. This could provide the basis for refining the list of candidate avirulence targets for this resistance to include only those identified as conserved amongst different *P. syringae* pathovars. However, if qualitatively distinct phenotypes were observed following challenge with other *P. syringae* pathovars, it would not be possible to exclude pathogen proteins that are conserved across *Psph* and other pathovars from the list of candidate targets for race-nonspecific resistance. The presence of additional effectors in other pathovars, and their absence from *Psph*, could underlie distinct interaction phenotypes.

Variable conservation of T3SS proteins within and amongst pathovars provides the basis for the provisional classification of protein groups that serve at different diagnostic levels with regard to host range differences. Expanding on work by Baltrus *et al.* (2012), which compared the effector inventories of *Psph* 1644R (isolated from *Vigna radiata*), *Psph* 1448A (*Phaseolus vulgaris*) and *P. syringae* pv. *glycinea* strain race 4 (R4; *Glycine max*), the diagnostic levels indicated in Table 3.2 highlight potential molecular determinants of host specialisation. Baltrus *et al.* (2012) found five T3Es (AvrB2, HopAW1, HopV1, AvrB4-2 and HopW1-2) to be present in *Psph* 1448A and absent from the two other isolates. Expression of only *avrB2* from *Psph* 1448A in a mutant of *P. syringae* pv. *glycinea* R4 (lacking avirulence genes *hopH1*/*hopC1* and *hopM1*) increased growth of this avirulent strain on common bean cultivar Canadian Wonder (Baltrus *et al*., 2012). This suggests a possible role for AvrB2 as a candidate determinant of host specificity in *Psph* 1448A. Complete conservation of AvrB2 amongst *Psph* isolates collected from *Phaseolus* spp. supports this hypothesis (Table 3.2). HopAW1, HopV1, AvrB4 and HopW1-2 were not detected, or contain premature translation termination sites or frameshift mutations, in *Psph* race 5 isolate 1516A, collected from *Neonotonia wightii*.

Curiously, AvrB4-1 was reported by Baltrus *et al*. (2012) to be present in *Psph* 1644R and 1448A, but absent from *P. syringae* pv. *glycinea* R4. In the current study, AvrB4-1 was found in *Psph* 1448A and *P. syringae* pv. *glycinea* R4, but was not detected in *Psph* 1644R by either automated (Biotools version 1.2.12) or manual TBLASTN comparison due to the absence of a conserved N-terminal domain containing a translation initiation site. Most *Psph* isolates examined here share a conserved N-terminal truncation and an alternate translation start site within AvrB4 relative to the variants present in *Psph* 1448A (AvrB4-1 and AvrB4-2) and *P. syringae* pv. *glycinea* R4 (AvrB4-1). Zumaquero *et al*. (2010) showed that AvrB4-1 and AvrB4-2, which differ by three amino acids, are functionally redundant and contribute quantitatively to virulence in *Psph* 1448A. AvrB4 variants with alternate translation initiation sites were not detected in either *Psph* 1644R or 1516A.

Most T3Es identified in *Psph* 1644R, *Psph* 1448A and/or *P. syringae* pv. *glycinea* R4 by Baltrus *et al.* (2012) were predicted to be similarly present in or absent from these strains in the current research. However, several other discrepancies between the TBLASTN results obtained in this study and those

reported by Baltrus *et al*. (2012) were found with regard to the presence/absence of T3Es in *Psph* 1644R and/or *P. syringae* pv. *glycinea* R4. Manual TBLASTN searches detected only truncated segments (lacking translation initiation sites) and/or rearranged segments of proteins (HopAV1, HopF3 and HopAH1) reported by Baltrus *et al.* (2012) to be conserved amongst all three isolates examined. Similarly, HopAY1 was reported to be conserved in *Psph* 1644R and 1448A, but only truncated and rearranged segments of this protein were detected in 1644R by manual TBLASTN analysis.

The current research extends the list of T3SS proteins conserved amongst *Psph* isolates collected from *Phaseolus* spp. and not found or mutated in *P. syringae* isolates collected from other hosts. It has also enabled identification of additional T3SS proteins that are conserved amongst different *P. syringae* pathovars, which might represent targets for a potential species-level or otherwise broad-spectrum resistance mechanism in common bean. This study demonstrates the importance of exploiting increasingly affordable whole-genome sequencing of multiple bacterial isolates within pathovars to enable GWA genetics. Selection of closely related strains of wide geographic origin and pathogenic capability provides increased power for the identification of pathovar-level determinants of pathogenicity.

AvrD

Two computational approaches to GWA genetics (based on T3SS protein homology searches and SNP-based statistical phylogenomics) were used to identify candidate pathogenicity determinants that are uniquely conserved within the most broadly virulent race of the pathovar, *Psph* race 6 (Tables 3.2 to 3.5). These candidates may provide excellent biomarkers for identifying novel resistance sources effective against *Psph* race 6, a devastating pathogen that threatens common bean production worldwide.

For example, AvrD is exceptional in its avirulence activity, which unlike T3Es, is not determined by direct interaction with a target plant receptor following translocation into the host plant cell. Instead, AvrD catalyses the synthesis of extracellular syringolide elicitors that induce a hypersensitive cell death response in soybean plants carrying *R* gene *Rpg4* (Keen *et al*., 1990; Midland *et al*., 1993, 1995). Vencato *et al*. (2006) nonetheless demonstrated with translocation reporters that AvrD from *Psph* 1448A can be delivered into plant cells via the T3SS. They

suggested that, in T3SS-competent *P. syringae*, syringolide production might occur following translocation of AvrD into the plant cell (Vencato *et al*., 2006). They also found *avrD* to be strongly induced by the HrpL alternative sigma factor, which activates expression of genes essential for pathogenicity in *P. syringae*.

Despite the apparent absence of a T3SS-targeting signal peptide from all examined variants of AvrD, Keen *et al*. (1990) found that *P. syringae* pv. *glycinea* race 0 possessing a functional *avrD* transgene depends on *hrp* genes (critical for assembly of the T3SS apparatus) for the induction of *avrD*-dependent hypersensitive necrosis in soybeans expressing *Rpg4*. *hrp* mutant strains expressing *avrD* were nonetheless able to produce the syringolide elicitor in culture fluids, indicating that elicitor synthesis is not *hrp*-dependent. Conversely, *E. coli* cell cultures expressing *avrD* and lacking the T3SS *hrp* gene cluster elicited a hypersensitive response from resistant soybean plants. Infiltration of leaves of resistant plants with partially purified AvrD failed to elicit a hypersensitive response, highlighting the importance of elicitor production for induction of hypersensitivity (Keen *et al*., 1990).

The existence of isolates originally defined as race 6 that lack *avrD*^{p6} (instead possessing *avrD*$^{p1–2}$) suggests that *Psph* 'race 6' might be subdivided if a source of race-specific resistance were to be identified and used as a novel host differential. Under the hypothesis that, within race 6, *avrD*^{p6} confers virulence towards bean plants lacking quantitative resistance, the broad virulence of the four exceptional isolates originally designated as race 6 might be attributable to another virulence factor. Isolates 1448A, 1715A, 2109 and 2654A might therefore be re-classified with a novel race designation based on interaction phenotypes with a new host differential. As Taylor *et al*. (1996a) note in their seminal *Psph* race differentiation study, "the ability to differentiate [races] is entirely dependent on the availability of an appropriate cultivar" possessing specific resistance. Findings from several subsequent studies surveying *Psph* populations in different regions indicate that an expanded bean differential set is required to enable further race differentiation and accurate race designations (González *et al*., 2000; Lamppa *et al*., 2002; Rico *et al*., 2003; Félix-Gastélum *et al*., 2016). However, in view of the otherwise high level of conservation of potentially functional alleles encoding AvrD amongst all *Psph* races (Table 3.4), this protein cannot be ruled out as a candidate target for a hypothesised race-nonspecific R protein in common bean.

HopX1

The truncated allele *hopX1*^{p6}, which is not predicted to encode an N-terminal T3SS signal peptide, was also identified as potentially causally related to the *Psph* race 6 virulence phenotype in the current research (Table 3.5). *hopX1* is located near the chromosome-borne cluster of *hrp* genes required for the production of the T3SS apparatus (Mansfield *et al*., 1994). In *Pseudomonas syringae* pv. *tabaci* strain 11528, *hopX1* encodes a cysteine protease that promotes degradation of JASMONATE ZIM DOMAIN (JAZ) proteins, key repressors of the jasmonic acid (JA) plant defence signalling pathway (Gimenez-Ibanez *et al.*, 2014). In the absence of a cognate R protein in Arabidopsis, T3E protein HopX1 induces JA-dependent genes and represses the generally antagonistic salicylic acid (SA) defence pathway, thereby promoting susceptibility to biotrophic and hemi-biotrophic plant pathogens such as *P. syringae*.

Common bean *R* gene *Pse-2* encodes the receptor protein matching HopX1, encoded by functional avirulence alleles in *Psph* (Mansfield *et al*., 1994; Stevens *et al*., 1998). *Pse-2* confers resistance to *Psph* races 2, 3, 4, 5, 7, 8 and 9 (Miklas *et al*., 2011). Stevens *et al*. (1998) reported that individual missense mutations convert alternate alleles of *hopX1* in *Psph* races 1, 3, 6 and 9 into virulence determinants. They inoculated halo blight differential line A43 (*Pse-2*, *Pse-3*, *Pse-4*) with transconjugants of broadly virulent isolate 1448A harbouring plasmids expressing different *hopX1* alleles. Loss of avirulence function of given *hopX1* alleles was attributed to a p.G191R mutation in race 1 isolate 1281A (*hopX1*^{p1}), p.E310K in each of race 3 isolate 1301A and race 6 isolate 1448A (*hopX1*^{p3}), and p.W235L in race 9 isolate 2709A (*hopX1*^{p9}). Transconjugant 1448A expressing *hopX1*^{p5} (then designated *avrPphE5*) or *hopX1*^{p4} (*avrPphE7*) elicited a hypersensitive cell death response on pods of differential line A43, showing that these alleles encode determinants of avirulence. *hopX1*^{p9} (*avrPphE9*) was proposed to be a determinant of reduced virulence compared with *hopX1*^{p1} and *hopX1*^{p3}, which elicit a fully compatible (susceptible–virulent) interaction. Stevens *et al*. (1998: 168) observed "a slight reduction in the spread of expansion of water-soaked lesions caused by the transconjugant with the *avrPphE9* allele", compared with a fully compatible interaction. This is inconsistent with the presence of *hopX1*^{p9} in race 7 isolates 1314 and 1549A (Table 3.5), which are, by definition, purportedly avirulent on bean lines possessing *Pse-2*. These contradictory findings suggest that race typing of isolates

1314 and 1549A should be repeated and their *hopX1* alleles confirmed by Sanger sequencing. However, inferences made at the level of pathovar race based on evidence from interaction phenotypes elicited by individual isolates may not be valid.

Stevens *et al*. (1998) reported another route to virulence: a 104-bp insertion at the C-terminus of *hopX1*^{p8} (*avrPphE8*), which causes a reading frame shift and loss of a translation termination site (also see Inoue and Takikawa, 1999). They observed that A43 pods did not exhibit a fully compatible interaction following inoculation with transconjugant 1448A expressing *hopX1*^{p8}, with water-soaked lesions remaining more localised than those caused by wild-type 1448A. In the current research, this insertion was predicted to abolish the T4SS translocation potential of HopX1 in race 2 isolates 882 and 1502A, and in race 8 isolates 2654C and 1645, based on the C-terminal amino acid composition analysed using the T4SEpre model (Wang *et al*., 2014; Eichinger *et al*., 2016). The role of the T4SS in plant pathogenicity has been demonstrated in only a few systems, such as in the tumour-inducing capacity of *Agrobacterium tumefaciens* (Christie, 1997) and in the delivery of a heat-stable protein from *Burkholderia cenocepacia* into onion tissues that causes water-soaked lesions (Engledow *et al.*, 2004). However, genomic analyses have identified T4SS components in numerous phytopathogens, suggesting potentially important and evolutionarily conserved roles in pathogenicity (Stavrinides, 2009, and references therein).

The presence of *hopX1*^{p8} in *Psph* isolates designated as race 2, which are purportedly avirulent on plants carrying *Pse-2*, conflicts with the reported role of this allele in conferring reduced virulence. Alternatively, these findings may reflect shortcomings in the subjective method of halo blight disease scoring. This method has enabled investigation of proposed Phaseolus–Pseudomonas gene-for-gene complementarity relationships, which are based on qualitative interaction phenotypes. However, Miklas *et al*. (2011) found *Pse-2* to have broader effect than was previously reported (Teverson, 1991; Taylor *et al*., 1996a), with resistance to races 3, 8 and 9 (isolates not specified) cosegregating with the previously described resistance to races 2, 4, 5 and 7. This broader 'resistance' might correspond to the reduced virulence elicited from A43 pods following inoculation with transconjugants expressing *hopX1*^{p8} or *hopX1*^{p9}. Race 3 isolates possess variable *hopX1* alleles (Table 3.5) and so the broadening of the resistance conferred by *Pse-2* to encompass

race 3 might not be valid. For example, it is unlikely that transconjugants expressing the allele diagnostic of race 6 and present in race 3 isolate 2475A (*hopX1*^{p6}) would elicit a hypersensitive response from plants possessing *Pse-2* only; the *hopX1*^{p6}-encoded truncation of 154 amino acids from the N-terminal region of HopX1 is predicted to abolish its T3SS translocation potential (Table 3.5). Alternatively, it is possible that *Pse-2* encodes the receptor target for a different avirulence effector in isolate 2475A and others.

Phylogenomics

The phylogenies inferred from pan-genome SNPs provide a high level of resolution for characterising relatedness, which is particularly informative when analysing closely related strains (Bart *et al*., 2012). Figures 3.1 to 3.3 indicate that *Psph* race 1, 5, 7 or 9 may be the most closely related to the ancestral phenotype of the pathovar. A larger and random sample of sequenced *Psph* isolates, unbiased towards the characterisation of a particular race, would be required to substantiate this hypothesis. Nonetheless, this theory is consistent with reports that *Psph* is able to lose avirulence genes (by deletion or other mutations to virulence) in order to overcome host defence responses (Teverson, 1991; Taylor *et al*., 1996a,b; Jackson *et al*., 2000; Pitman *et al*., 2005; Rivas *et al*., 2005). The global ubiquity of broadly virulent *Psph* race 6 is mirrored by its wide phylogenetic distribution. The race 6 phenotype exemplifies an adaptive strategy by which the loss of functional avirulence genes permits the pathogen to expand its range of virulence. This strategy allows the pathogen to disseminate in the absence of a widely deployed *R* gene with specificity against this race.

The close phylogenetic relationships amongst races 1, 5, 7 and 9 indicate that the similarities between these races extend beyond the presence of functional avirulence alleles encoding HopF1. Their wide geographical distribution might reflect global movement with intensifying agricultural trade flows. They may also have been allowed to proliferate in countries where *R* gene *Pse-1* has not historically been widely deployed in common bean cultivars.

Rivas *et al*. (2005) presented evidence to suggest that the absence of *hopF1* from races 2, 3, 4, 6 and 8 represents a single deletion event in the plasmid-borne pathogenicity island, indicating that isolates from these races derive from a common ancestor. They reported that while the pathogenicity island is otherwise well

conserved amongst *Psph* isolates and races, deletion of a continuous 9.5-kb fragment located between *avrD* and *avrB2* in 1448A resulted in the loss of *hopF1* (see Figure 1 of Rivas *et al*., 2005). Rivas *et al*. (2005) proposed that the global distribution of races lacking *hopF1* implies that its deletion aided their ability to spread, following probable initial dispersal on infected seeds through international trade. A novel *Psph* pathotype, 'race 2', was described in 1964 (Walker and Patel, 1964), which represented the break down of resistance conferred by *Pse-1*. This was precipitated by the wide deployment of *Pse-1* in the United States since the 1940s (Rivas *et al*., 2005). Race 2 was subsequently subdivided into races 2, 6 and 8 (Teverson, 1991; Taylor *et al*., 1996a). Thus, a proposed common ancestor of races lacking *hopF1* suggests that mutations that provided the basis for the more recent subdivision of *Psph* occurred after the *hopF1* deletion event (Rivas *et al*., 2005). These genetic changes may include all or some of those at *hopX1* (which is purportedly mutated to virulence in races 1, 3, 6, 8 and 9; Stevens *et al*., 1998) and the unique presence of avirulence gene *hopAR1* in races 3 and 4. The latter is thought to be a recent acquisition, occurring almost exclusively in isolates collected in East and Central Africa (Taylor *et al*., 1996a,b; Jackson *et al*., 2000; Rivas *et al*., 2005; also see Table 3.2).

Curiously, *hopF1* was not detected in the draft assembly for 'race 1' isolate 1644R (scaffolds with RefSeq accessions NZ_JH137509 to NZ_AGAS01001154; Baltrus *et al*., 2012) using automated computational analysis based on sequence homology in the current research (Table 3.2). Its absence from the corresponding GenBank annotation file (Baltrus *et al*., 2012) suggests that race typing of this isolate should be repeated and that PCR amplification of *hopF1* should be attempted. However, the comparatively close phylogenetic relationship of isolate 1644R with isolates from races 1, 5, 7 and 9, as well as with other *P. syringae* pathovars and non-pathogenic *P. fluorescens*, might suggest that it represents the ancestral *Psph* phenotype (Figures 3.1 to 3.3). It is possible that 1644R harbours another avirulence effector matched by Pse-1. In Arabidopsis, for example, R protein RPM1 guards RIN4, the virulence target of *P. syringae* type III effectors AvrRpm1 and AvrB (Mackey *et al*., 2002). Among the 34 *Psph* genomes investigated, isolate 1644R is the most distantly related to others within the pathovar. This is also evidenced by the apparent absence of large sections of sequence from the chromosome and plasmids relative to other isolates (Figures 3.4 to 3.6). However, this might be attributable to

improvements in next-generation sequencing and *de novo* assembly algorithms that have emerged within the past five years.

The phylogenetic clustering of isolates from races 3 and 4 (in which avirulence effector HopAR1 is uniquely present; Table 3.2) collected in East Africa is consistent with previous reports. Taylor *et al*. (1996b) suggested that the prevalence of isolates from races 3 and 4 was attributable to the historical absence of corresponding *R* gene *Pse-3* in beans grown in this region, noting its complete absence from Rwandan farmers' mixtures. In parts of Africa where *Bean common mosaic necrosis virus* (BCMNV) occurs, the deployment of the *I* gene (tightly linked with *Pse-3*) can result in a severe, systemic hypersensitive response that is detrimental to the plant (Vallejos *et al*., 2006; Bello *et al*., 2014). Deployment in combination with the *bc* recessive *R* genes effective against the virus is necessary to attenuate this harmful response (Bello *et al*., 2014). Race typing of isolates of *Psph* present in this region currently and in the future would provide an intriguing insight into the pathogen population following wider deployment of *Pse-3* (and, by linkage, *I*). It is possible that some *Psph* strains may have lost *hopAR1* to evade detection by *Pse-3*-bearing beans. Pitman *et al*. (2005) found that serial passaging of *Psph* race 4 isolate 1302A through leaves of bean cultivar Tendergreen (*Pse-3*) caused the loss of a 106-kb genomic island (PPHG-1) harbouring *hopAR1*. Loss of PPHG-1 conferred the bacteria with virulence towards Tendergreen plants. The antimicrobial conditions engendered by the plant hypersensitive cell death response exerted strong selective pressure on bacteria to lose *hopAR1*. Consequently, virulent variants of the strain were favoured and became predominant within leaf tissue (Pitman *et al*., 2005).

Hypothesis testing

Genetic manipulations are required to test hypotheses formulated as a result of bioinformatics-enabled GWA analyses. Mutational (loss-of-function) experiments would enable investigation of the potential contributions of candidate effectors to the virulence phenotype of *Psph* race 6 on common bean plants lacking resistance. Zumaquero *et al*. (2010) generated single- and double-knockout mutants of *Psph* 1448A lacking functional effector-encoding genes, including *avrD*. To investigate the potential role of AvrD as a determinant of race-nonspecific avirulence, a useful future experiment would be to examine the effects on the virulence and pathogenicity

of mutant isolates lacking AvrD on common bean plants possessing the resistance factor(s) derived from PI 150414.

Additionally, gain-of-function experiments involving non-race 6 transconjugants expressing candidate virulence effectors from race 6 could be performed to investigate their contributions to virulence on bean plants possessing R proteins corresponding to native avirulence proteins. Such experiments might help to determine whether exogenous candidate virulence effectors from race 6 (e.g., *avrD*^{p6}) suppress recognition of indigenous avirulence effectors, such as those encoded by particular alleles of *hopX1*. Alternatively, transgenic experiments involving other *P. syringae* pathovars that lack a candidate effector could provide insights into its contribution to virulence on a non-host common bean plant. Stevens *et al*. (1998) found that *in planta* expression of *hopX1*^{p4} (*avrPphE4*) or *hopAR1* (*avrPphB*) triggered a hypersensitive response in beans carrying *Pse-2* or *Pse-3*, respectively. Thus, *Agrobacterium tumefaciens*-mediated delivery and expression of candidate determinants of race 6 or race-nonspecific avirulence into plants possessing resistance only from PI 150414 would provide a useful test of their ability to cause a gene-specific hypersensitive response. Experiments in which soybeans carrying *Rpg4* are inoculated with *P. syringae* pv. *glycinea* carrying different alleles of the transgene *avrD*^{p} from *Psph* would reveal whether these alleles encode proteins required for soybean hypersensitive response induction, as is the case for *avrD*^{p1}, *avrD*^{p2} and *avrD*^{t} (Yucel and Keen, 1994; Yucel *et al*., 1994a).

It would also be informative to inoculate common bean plants possessing resistance from PI 150414 with isolates 1448A, 1715A, 2109 and 2654A, which lack *avrD*^{p6} and all functionally characterised halo blight avirulence alleles. This might provide an indication of the specificity of the resistance. As discussed in Chapter 2, *R* genes clustered within the mapping interval on common bean chromosome Pv04 might each confer different resistance specificities against different *Psph* races, including race 6. These four isolates might not possess avirulence genes corresponding to an *R* gene with specificity against race 6 and, consequently, may be virulent on PI 150414.

Molecular markers targeting alleles predictive of *Psph* race 6 virulence will enable rapid detection of this globally distributed pathogen in farmers' fields. These diagnostic tools will therefore provide opportunities to identify and deploy novel and previously identified sources of halo blight resistance effective against race 6.

Chapter 4

Developing molecular breeding capability to enable selection of physiological and morphological characteristics in common bean for low-input production in the UK

4.1. INTRODUCTION

Common bean is a potentially viable legume break crop for farmers in the United Kingdom. A successfully adapted British cultivar would be characterised by several physiological resilience and morphological traits. For example, it would be tolerant to low temperatures (< 15 °C) for seedling establishment in late May to early June, and would acquire and use water and nutrients efficiently to minimise the requirement for external inputs.

Low temperature is a major limiting factor on common bean production throughout northern Europe (Austin and MacLean, 1972a,b; Drijfhout *et al*., 1991; Rodiño *et al.*, 2007). Originally domesticated in Latin America and Mesoamerica, common bean cultivars are generally poorly adapted to low temperatures (up to 15 °C day/8 °C night) and crops are prone to failure during cool summers (Austin and MacLean, 1972a,b; Harwick *et al*., 1978; Hardwick and Andrews, 1981; Rodiño *et al.*, 2007). Runner bean (*Phaseolus coccineus* L.) cultivars are generally better adapted to these conditions, and are grown more frequently in the UK (commercially and in private gardens) for their reliable crop of green beans. As is the case for common bean, however, sub-optimum temperatures delay or prevent germination and emergence, which can increase production costs and losses by extending the crop cycle, potentially beyond the growing period in which the crop is viable (Rodiño *et al.*, 2007).

Ideally, a UK-adapted cultivar would reach harvest maturity within 100 days of planting to avoid spoilage of seed caused by fungal infection (mainly *Botrytis* spp.) and discolouration that accompanies autumnal wet weather (Hardwick, 1988; Dodd and Taylor, 1991, unpublished data). Studies of early maturity of beans have focused on selecting for rapid development during the reproductive growth stages (e.g., early flowering and early pod maturation; Rodiño *et al.*, 2007). The potential importance of seedling vigour during the early vegetative growth stages in determining time to harvest maturity has received less attention. In view of the importance of seedling resilience in cooler climates, phenotypic and genetic characterisation of physiological and morphological traits during both the vegetative and reproductive growth stages may enable marker-assisted breeding of common bean cultivars adapted to UK growing conditions.

A determinate, bush-type growth habit and plant architecture amenable to easy mechanical harvest are also critical agronomic traits for ensuring the viability of this novel rotational crop for UK farming. Selection of plants with an upright rather than prostrate habit, and which set pods at a sufficient height from the ground will help to minimise losses during harvest.

A parental line, SOA-BN, combines many of these desirable physiological and growth habit characteristics. It has a determinate, upright bush habit (type I) with pod set occurring in the upper and middle parts of the plant, whereas a registered haricot variety from the NVRS named Edmund has an indeterminate climbing habit (type IV). SOA-BN was previously identified as a superior source of cold tolerance in controlled-environment experiments (R. C. Hardwick, personal communication), and of earliness of maturity and potential drought tolerance based on findings from three seasons of field evaluations from 1988 to 1990 (Dodd and Taylor, 1991, unpublished data). Under the drought conditions of the 1990 field trial (with rainfall during the main growing period totalling only 30% of average), SOA-BN produced a seed yield of 143 g m^{-2}, whereas yields of all other lines were ≤ 80 g m^{-2}. SOA-BN also produced a significantly higher yield than all lines tested in 1988, the coolest of the three seasons studied, with 317 g m^{-2} ($P = 0.05$). SOA-BN is harvestable within 100 days and was earlier maturing than Edmund, all but two of the 75 experimental inbred lines tested, and the North American haricot cultivar Seafarer in 1988 and 1989.

In the 1988 and 1989 trials, which served as realistic tests of cold tolerance and growth under high temperatures, respectively, inbred line 15 (hereafter referred to as Capulet) was identified as one of the most promising lines, with more consistent yields under different conditions (263 g m^{-2} in 1988 and 271 g m^{-2} in 1989). Capulet also reached harvest maturity earlier than cultivar Seafarer, one of five white-seeded haricot cultivars used as controls (Dodd and Taylor, 1991, unpublished data).

Numerous genes involved in plant responses to low temperatures are also implicated in the regulation of water use (Seki *et al*., 2001, 2002, 2003; Guy, 2003; Narusaka *et al*., 2004; Shinozaki and Yamaguchi-Shinozaki, 2006). It is therefore plausible that the distinct physiological characteristics exhibited by SOA-BN and Capulet might have a basis in the efficient use of water. Water-use efficiency (WUE) is the efficiency with which a plant can assimilate CO_2 into biomass per unit

of water transpired (Passioura, 2006). Along with root development, stomatal control of transpiration is one of the two primary mechanisms by which C_3 plants like beans adapt to water stress (Kuruvadi and Aguilera, 1990; Laffray and Louguet, 1990). Stomatal conductance is a physiological trait that describes the rate of passage of water vapour exiting, or CO_2 entering, the plant through the leaf stomata. It can therefore be used as an indicator of WUE where biomass or yield data for a given quantity of water transpired are lacking. Rapid stomatal closure in response to water stress in a photosynthetically efficient genotype is a desirable drought avoidance mechanism (Pimentel *et al*., 1999).

Evaluation of phenotypic variation amongst SOA-BN, Capulet and Edmund with regard to WUE was therefore **an initial aim** of this research chapter. Quantitative measurements of phenotypic variation in WUE were taken for the parental lines SOA-BN and Edmund. WUE phenotypic data were also collected for Capulet, an improved haricot-type line derived from this parentage that combines multiple disease resistance (including to halo blight) with potential cold tolerance and earliness of maturity. Detection of significant differences in WUE between parent lines would provide the basis for characterising phenotypic variation within recombinant inbred populations to enable QTL mapping and, ultimately, marker-assisted breeding for enhanced WUE.

Evaluation of root architecture may also help to elucidate the physiological basis of resilience traits exhibited by SOA-BN. Drought tolerance is a complex phenotype governed by multiple genes involved in numerous constitutive traits (Beebe *et al*., 2013). However, drought-adapted genotypes that increase root growth under water stress could maintain high stomatal conductance, serving to decrease their intrinsic WUE (White *et al*., 1990). Although the WUE experiment was carried out under non-stress conditions, a more extensive root system in SOA-BN might underlie a lower mean intrinsic WUE (and higher mean stomatal conductance) compared with Capulet and Edmund.

Root characteristics are important morphological and physiological determinants of the ability of plants to acquire water and nutrients. Under intermittent drought, bean plants with extensive tap, basal, lateral and adventitious roots have an enhanced capacity for soil exploration and to uptake available water during infrequent periods of rain (Beebe *et al.*, 2013). Furthermore, numerous drought-adapted common bean genotypes exhibit tolerance to low soil phosphorus

(Beebe *et al.*, 2008). Improved phosphorus acquisition is correlated with several root phenotypes in common bean, including shallow basal root growth angles (Liao *et al*., 2004) and more basal root whorls (Miguel *et al*., 2013, 2015) and adventitious roots (Miller *et al.*, 2003) for enhanced topsoil foraging. Whorls are "distinct tiers of basal roots that emerge in a tetrarch fashion along the base of the hypocotyl" (Miguel *et al*., 2013: 973) and are depicted in Figure 4.1. However, while a more extensive root architecture enables uptake of more moisture and nutrients, efficient photosynthate partitioning to the developing grain is important for providing yield benefits under abiotic stress conditions (Beebe *et al*., 2009). Drought avoidance strategies combining increased rooting depth and water uptake with reduced stomatal conductance, evaporation surface area and radiation absorption (e.g., by leaf rolling) are also desirable under water stress (Beebe *et al.*, 2013). Thus, **a second aim** of this research chapter was to evaluate root architecture characters amongst the parents and recombinant inbred lines (RILs) of the SOA-BN × Edmund population to enable mapping of putative QTL.

Finally, although seed quantities of the RILs were limited, **a third aim** was to conduct a pilot experiment in the final year (2015) to evaluate developmental, morphological and reproductive characteristics amongst the inbreds under polytunnel and field growing conditions.

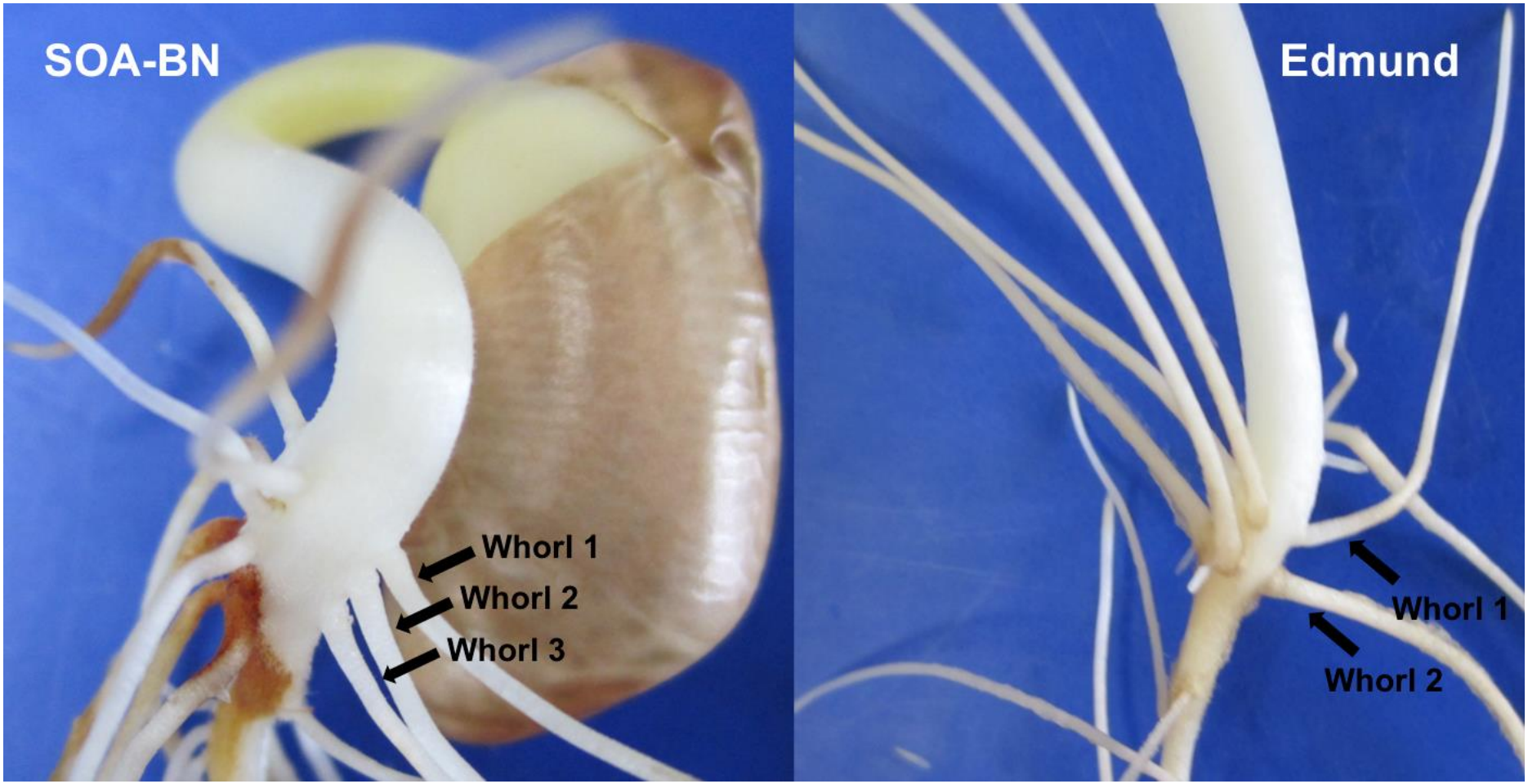


Figure 4.1. *Phaseolus vulgaris* parental genotypes SOA-BN and Edmund with contrasting basal root whorl phenotypes, with basal roots emerging from discrete whorls.

4.2. MATERIALS AND METHODS

4.2.1. Water-use efficiency evaluation

Plant material and growing conditions

Sixty-eight seeds of each of the genotypes Edmund, SOA-BN and Capulet were sown in 15-cm-diameter plastic pots (four seeds per pot, with 17 biological replicates of each genotype) containing John Innes No. 2 loam-based compost. A 3-mm soil sieve was used to homogenise the growing medium. Pots were filled to 1 kg (including a 73-g pot and drip dish) and watered to field capacity (FC; soil water content after complete gravitational drainage of excess water). Seeds were placed on top of moist soil, ensuring good soil contact, with an evenly distributed 5-g vermiculite covering. Following drainage of excess water, each pot and drip dish was weighed to determine FC weight. After emergence, pots were thinned to one plant per pot.

Plants were grown under glasshouse conditions with a heating temperature of 18 °C and a venting temperature of 20 °C from 18 August to 3 October 2014 at Warwick Crop Centre, Wellesbourne, UK (52°12.54′N, 1°36.23′W). Pots were watered every two days for a period of 30 days from sowing. Seven replicate plants of each genotype were harvested at the end of the 30-day period to determine the average 'starting aboveground biomass' and 'starting leaf area' of each genotype. At this point, each pot of the remaining ten replicates was covered with cling film (with a small opening at the plant stem and bamboo-cane guide) in order that only water lost trough leaf transpiration (and not through evaporation from the soil surface) affected pot weight measurements. One 'blank' control pot containing only soil, water and a bamboo cane was included to ensure that its weight remained constant throughout the experiment, without additional watering, due to the prevention of water loss by its cling-film covering.

The weight of each of the remaining ten replicate pots for each genotype was then recorded and 400 g of water added. Each pot was re-weighed following complete gravitational drainage of excess water to determine its 'target' FC weight for subsequent watering. For the remainder of the experiment, each pot was weighed and watered every two days (including once a week with a 1:200 dilution of a

nutrient solution containing 15 g L^{-1} of N, 7 g L^{-1} of P, and 30 g L^{-1} of K) to the target FC weight + 100 g and re-weighed after complete gravitational drainage of excess water. This enabled calculation of the total water transpired by each plant up to harvest.

Experimental design and layout

Pots were arranged in two randomised complete block designs (RCBDs). The first RCBD consisted of seven biological replicates of each genotype, which were harvested 30 DAS (days after sowing). The first harvest was done immediately prior to the period during which defined quantities of water were applied to remaining plants in order to determine the average starting aboveground biomass and average starting leaf area of each genotype. The second RCBD consisted of ten biological replicates of each genotype, which were harvested 44 DAS (two replicates), 45 DAS (four replicates) and 46 DAS (four replicates), at the flowering stage, to determine the 'final aboveground biomass' and 'final leaf area' of each plant.

WUE measurements

Gas exchange measurements were taken from the middle leaflet of the youngest fully expanded trifoliolate leaf. Pimentel *et al.* (1999) found that the youngest fully expanded leaf across four common bean lines had the maximum photoassimilate export. In the present study, it was not possible to consistently determine leaf number across genotypes due to their contrasting growth habits. Stomatal conductance (g_s) and net photosynthetic rate (A) measurements were taken at the pollination stage (35 DAS) using a CIRAS-2 infrared gas analyser fitted with a PLC5-B broad-leaf cuvette (PP Systems, Amesbury, USA). Avoiding the midrib, the leaflet attached to the plant was placed in the temperature-controlled cuvette chamber under forced ventilation to obtain a high boundary layer conductance. The mean ambient air temperature in the chamber was 29.3 °C (SD = 1.05 °C). The chamber, fitted with an LED unit for automatic control of light intensity, was illuminated with a photon flux density of 1,000 µmol m^{-2} s^{-1}.

Intrinsic water-use efficiency (IWUE; i.e., the rate of photosynthesis obtained for a given stomatal conductance) at the pollination stage (35 DAS) was calculated by dividing A by g_s (Osmond *et al.*, 1980; Pimentel *et al.*, 1999). The pollination

stage was chosen as the first time point for gas exchange measurements in view of evidence presented by Pimentel *et al*. (1999) indicating that this developmental stage is optimal for discriminating between efficient and inefficient common bean genotypes with regard to IWUE, under both water-stress and non-stress conditions. Pimentel *et al.* (1999) found that net photosynthetic rate reached its maximum level at this stage across four common bean genotypes, indicating that carbohydrate accumulation during pollination will be critical to subsequent pod filling. Seed production would be severely reduced in the event of a climatic constraint at this stage, as a reduction in photosynthesis would cause the abscission of early flowers, from which the majority (90%) of mature pods would otherwise have developed (Norman *et al*., 1995).

Stomatal conductance was also measured at the flowering stage (44 DAS) using a steady-state SC-1 Leaf Porometer (Decagon Devices, Inc., Pullman, USA). For measurement, the sensor head of the Porometer is clamped onto the leaflet (avoiding the midrib) and the instrument compares humidity measurements between the leaf and two known conductance elements to determine the actual water vapour flux from the leaf to the environment.

Data for other WUE metrics (including fresh and dry weight of aboveground biomass, leaf area, and total water transpired from 31 to 46 DAS) were collected for each plant after harvest. Aboveground biomass was dried at 80 °C for 72 h. Average starting biomass and starting leaf area for each genotype were deducted from individual final biomass and final leaf area measurements to calculate the 'effective' biomass and leaf area of each of the 30 plants that were harvested at the end of the experiment. The developmental stage of each plant was also noted at 44 DAS using the common bean growth stage guide of Schwartz *et al*. (2004).

The extent to which genotypes differed with regard to each WUE metric was determined by one-way analysis of variance (ANOVA), assumptions tested (homogeneity of variance, approximately normal distribution and independence of residuals), and box-and-whisker plots drawn using R statistical computing software version 3.1.1 (R Core Team, 2016). Significantly different means were segregated using the Tukey's honest significant difference (HSD) multiple comparison test.

4.2.2. Root architecture evaluation

Plant material and growing conditions

Twenty-eight seeds of each of the parental genotypes SOA-BN and Edmund were sown in 15-cm-diameter plastic pots (four seeds per pot, with seven biological replicates of each genotype) containing Sinclair vermiculite with medium-sized granules (2.0–5.0 mm). Vermiculite was used as the growing medium to enable easy root rinsing and assessment. Following an initial top-watering drench at sowing, pots were bottom-watered every two days with a 1:200 dilution of a nutrient solution containing 15 g L^{-1} of N, 7 g L^{-1} of P, and 30 g L^{-1} of K (Vitafeed Standard, Vitax, Leicester, UK). Seeds were sown into moist vermiculite approximately 1.5 cm deep. Pots were arranged in a RCBD. After emergence, pots were thinned to one plant per pot. Plants were grown under glasshouse conditions at Warwick Crop Centre, Wellesbourne, UK (52°12.54′N, 1°36.23′W), with a heating temperature of 18 °C and a venting temperature of 20 °C, and with supplementary lighting to provide a 16-h photoperiod.

Phenotypic measurements

Roots and shoots were harvested 21 DAS. Roots were rinsed with water to remove vermiculite, and were evaluated immediately after rinsing. Traits were assessed using methods adapted from the common bean "shovelomics" protocols developed by the laboratory of Jonathan Lynch and Kathleen Brown (Lynch and Brown, c2016; Burridge, 2012; Burridge *et al*., 2016). Whereas these protocols involve excavating field-grown bean plants from the soil for evaluation, the current research investigates seedling root and shoot traits in plants grown under glasshouse conditions in view of their potential role in conditioning resilience characteristics during early vegetative growth.

Traits for evaluation were adventitious (hypocotyl) root number, length (maximum, approximate median, and minimum), and diameter at 1 cm from the junction with the hypocotyl (maximum, approximate median, and minimum); basal root number, length (maximum, approximate median, and minimum), diameter at 1 cm below the base of the hypocotyl, growth angle (maximum and minimum), and branching density at a representative 2-cm section; mean basal root whorl number

per file; taproot length, diameter at 1 cm below the hypocotyl base, and branching density at a representative 2-cm section; and root and shoot dry weights.

A standard tape measure (1-mm resolution) and a phenotyping board featuring a protractor and 2-cm length scale were used for quantitative measurements of root length, growth angle and branching density (Burridge *et al*., 2016). A digital caliper (0.01-mm resolution) was used to measure root diameter. Root biomass and shoot biomass (separated at the base of the hypocotyl) were dried at 80 °C for 72 h after assessment of root morphological traits.

All statistical analyses were performed using R version 3 (R Core Team, 2016). Differences between parental genotypes were assessed by one-way ANOVA, and means, 95% confidence intervals and least significant differences (LSD) were plotted. The assumption of homogeneity of variances was tested by *F*-tests.

QTL mapping of root traits

Further experiments were performed to characterise phenotypic variation within the SOA-BN × Edmund $F_{6:7}$ RIL population with regard to root and shoot traits in which parental lines were found to differ significantly. Basal root number, basal root whorl number per file, maximum and minimum basal root growth angles, taproot diameter at a depth of 1 cm below the hypocotyl base, shoot diameter at 1 cm above the hypocotyl base, and root and shoot dry weights were evaluated in these experiments to enable genetic mapping. Seed coat colour was evaluated on a 0–10 (light to dark) scale, where a score of 0 corresponds to a white seed coat (absence of pigmentation) and a score of 10 corresponds to a black seed coat. Lines that were still segregating for seed coat colour were noted. Hypocotyl colour was assessed on a 0–5 scale, where a score of 0 corresponds to a green hypocotyl and a score of 5 corresponds to a purple hypocotyl, with intermediate scores assigned to RILs intermediate between these two extremes.

RIL and parental seeds were sown in 4-L plastic pots (one seed per pot, with three biological replicates per line) containing Sinclair vermiculite with medium-sized granules (2.0–5.0 mm) and plants were grown under glasshouse conditions as described for the parental assay. Variable germination rates resulted in fewer than three replicates for some RILs. Pots were arranged in three RCBDs along one bench of a glasshouse compartment; due to space limitations, it was necessary to divide the

lines between three consecutive experiments, with genotype replication occurring within each experiment. Plants were harvested 20–22 DAS.

The mean phenotype of each RIL for each trait was calculated. Fixed- and random-effects variance components were estimated and tested for each trait by restricted maximum likelihood (REML) analysis as described in Chapter 2. Least-squares means were generated using lmerTest version 2.0-32 where random-effects variance components were estimated to be significant within mixed models (Table A8.1, APPENDIX 8). Pearson product-moment correlation coefficients were calculated for each pairwise trait combination.

QTL analyses using the SOA-BN × Edmund RIL phenotypic data and genotyping-by-sequencing-derived linkage map were performed essentially as described in Chapter 2. Briefly, this involved both one- and two-dimensional genome scans for individual QTL and pairs of QTL. Loci with significant or higher-than-average LOD scores were noted to enable further evaluation of their significance in the context of multi-QTL models fitted and refined using the multiple imputation method. All reported statistics for each term within a QTL model (derived by ANOVA) account for the estimated additive and cumulative pairwise interactive effects of the locus, which may be individually insignificant but collectively significant (Broman and Sen, 2009).

Seed coat colour was mapped by two separate analyses: (1) one- and two-dimensional genome scans identified significant loci ($P < 0.05$) based on quantitative phenotypic data collected for all RILs and a multi-QTL model was fitted and refined; and (2) a one-dimensional genome scan identified significant loci ($P < 0.05$) based on quantitative phenotypic data collected for RILs with non-white seed coats and a multi-QTL model was fitted and refined. Forty white-seeded lines predicted to lack the '*P*' allele that potentiates seed coat pigmentation were excluded from this analysis to enable mapping of other factors governing pigmentation in the presence of dominant *P*.

4.2.3. Polytunnel and field evaluation

Polytunnel

In a Gallas Ley polytunnel (Warwick Crop Centre, Wellesbourne, UK; 52°12.51′N, 1°36.07′W), seeds of each SOA-BN × Edmund $F_{6:7}$ RIL were sown in 4-L plastic pots (one seed per pot) containing Levington M2 compost with a covering of vermiculite on 5 June and 6 June 2015. Plants were arranged sequentially according to line number. Two blocks were arranged in this way along one bench of the polytunnel to give two pseudo-replicates per RIL. Pots were top-watered at sowing and bottom-watered thereafter, including once a week with a 1:200 dilution of a nutrient solution containing 15 g L^{-1} of N, 7 g L^{-1} of P, and 30 g L^{-1} of K (Vitafeed Standard, Vitax, Leicester, UK) until early pod fill. Plants were treated for aphids with 10 ml L^{-1} of potassium salts of fatty acids (Savona, Koppert B.V., Berkel en Rodenrijs, Netherlands) on 15 July, 22 July, 29 July and 5 August 2015.

Growth habit was assessed 75 DAS according to the four-point evaluation scale described by Van Schoonhoven and Pastor-Corrales (1987), where a score of 1 corresponds to a determinate bush habit and a score of 4 corresponds to an indeterminate climbing habit. Flower colour was evaluated 75 DAS on a 0–3 scale, where a score of 0 corresponds to white, a score of 1 corresponds to white–pink, a score of 2 corresponds to pink, and a score of 3 corresponds to purple. Pod speckling was scored qualitatively (present or absent) 100 DAS. Reproductive developmental stage was also assessed 100 DAS according to the guide of Schwartz *et al*. (2004). Days to harvest maturity and yield (total seed weight per harvested plant) were also recorded.

Field

In the field (Warwick Crop Centre, Wellesbourne, UK; 52°12.30′N, 1°36.55′W), seeds of each SOA-BN × Edmund $F_{6:7}$ RIL were hand-sown into moist soil in a 100-m^2 area on 4 June 2015. Five seeds of each RIL were sown ~ 2–3 cm deep in individual rows of 1.5 m in length (one seed per ~ 37 cm), with a width of 0.5 m between each row, and with 1-m borders on each side of the 100-m^2 plot. Rows were ordered sequentially by RIL number.

A pre-emergence herbicide containing pendimethalin (Stomp Aqua, BASF, Cheadle, UK) was applied at a rate of 2.9 L ha^{-1} immediately after sowing, and the plot was hand-weeded as required throughout the growing season. Slug pellets were applied on 16 June 2015. Plants were top-dressed with 87 kg ha^{-1} of Nitram (30 kg ha^{-1} of nitrogen; CF Fertilisers, Ince, UK) at flowering (50 DAS).

Traits for evaluation included the number of plants emerged per RIL, growth habit as assessed 82 DAS (Van Schoonhoven and Pastor-Corrales, 1987), reproductive developmental stage 82 DAS and 105 DAS (Schwartz *et al*., 2004), days to harvest maturity, and yield (total seed weight per harvested plant). The height at which plants set their pods was also evaluated 82 DAS according to a three-point scale, with a score of 0 corresponding to the concentration of pods towards the lower part of the plant, a score of 1 corresponding to the concentration of pods towards the upper part of the plant, and a score of 2 corresponding to a roughly even distribution of pods throughout both the lower and upper parts of the plant. A standard 1–9 scale was used to evaluate the severity of chlorotic mosaic or mottling symptoms observed 82 DAS (as proposed by Schoonhoven and Pastor-Corrales, 1987, for the assessment of common bean disease symptoms in the field), which were likely caused by a virus infecting species in the family Fabaceae (possibly *Cowpea aphid-borne mosaic virus*). In most cases, chlorosis was accompanied by necrotic symptoms, which were evaluated separately on a 1–9 scale 82 DAS. Vegetative adaptation (vigour) and reproductive adaptation (pod load; number and shape of pods, and size and number of seeds per pod) were also evaluated separately 82 DAS on scales of 1–9 as described by Schoonhoven and Pastor-Corrales (1987); a score of 1 corresponds to excellent adaptation and a score of 9 corresponds to very poor adaptation. Green area (an indicator of leaf drop) was similarly assessed on a 1–9 scale, where a score of 1 corresponds to ~ 100% green area and a score of 9 corresponds to ~ 0% green area.

Genetic mapping

QTL analyses were performed as described in Chapter 2. One-dimensional genome scans for single major-effect QTL were performed for all of the traits investigated. However, in view of the provisional nature of the phenotype data obtained from field- and polytunnel-sown RIL plants, and the computationally expensive nature of two-dimensional genome scans for additive and interacting QTL, these searches

were performed only for selected traits. For the polytunnel-collected phenotypic data, these traits were growth habit as assessed 75 DAS, developmental stage 100 DAS, days to harvest maturity, and yield. For the field-collected phenotypic data, these traits were growth habit as assessed 82 DAS, height of pod set 82 DAS, developmental stage 82 DAS and 105 DAS, days to harvest maturity, and yield. QTL models were fitted and refined only for those traits for which loci with LOD scores reaching or approaching significance at the 0.05 probability level were identified by one- or two-dimensional genome scans (Table 4.7). Pearson correlation coefficients were calculated for selected traits of agronomic interest to determine the relationship between phenotypes observed in the polytunnel and in the field for a given trait.

4.3. RESULTS

4.3.1. Water-use efficiency

Differences between Edmund, SOA-BN and Capulet in WUE were not significant at the 0.05 probability level for all but one metric tested (Table 4.1 and Figure 4.2). Comparison and segregation of means by Tukey's HSD test revealed that SOA-BN had a significantly lower mean (less efficient) IWUE than Capulet (HSD-adjusted $P = 0.05$), while parental lines SOA-BN and Edmund did not differ significantly in their IWUE (HSD-adjusted $P = 0.15$). SOA-BN was nonetheless at a significantly more advanced developmental stage than both Edmund and Capulet 44 DAS ($P < 2E{-}16$). These findings suggest that the physiological characteristics of SOA-BN do not have a basis in enhanced WUE under non-stress conditions.

Table 4.1. Summary of analyses of variance comparing *Phaseolus vulgaris* lines SOA-BN, Edmund and Capulet for indicators of water-use efficiency (WUE)

Trait	d.f.	Sum of squares	Mean square	*F*	Probability (*F*)
Dry-weight WUE	2	9.76E-07	4.88E−07	1.951	0.162
Fresh-weight WUE	2	1.47E−05	7.33E−06	0.83	0.447
A (35 DAS)	2	24.585	12.292	2.062	0.147
g_s (35 DAS)	2	546.1	273.03	0.404	0.671
IWUE (35 DAS)	2	7698.5	3849.2	3.373	0.049
g_s (44 DAS)	2	291	145.6	0.066	0.936
Effective dry weight	2	2.967	1.483	1.697	0.202
Effective fresh weight	2	88.58	44.288	0.82	0.451
Effective moisture content	2	101.56	50.778	1.071	0.357
Effective leaf area	2	211369	105685	2.068	0.146
Effective height	2	641	320.48	1.626	0.215
Water transpired	2	89422	44711	2.265	0.123

d.f.: degrees of freedom. *A*: net photosynthetic rate. g_s: stomatal conductance. IWUE: intrinsic water-use efficiency. DAS: days after sowing.

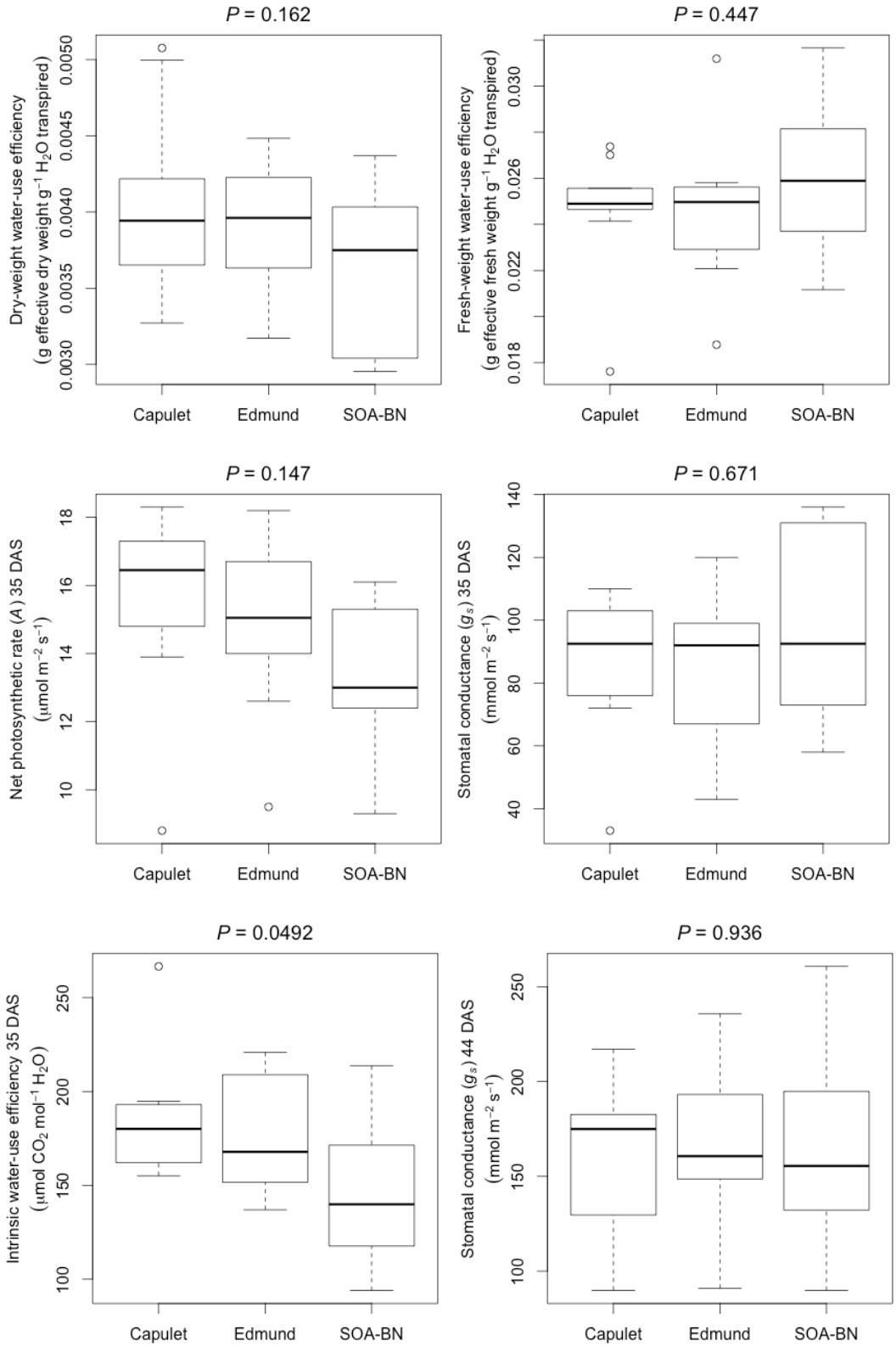


Figure 4.2. Summary of water-use efficiency phenotypes of *Phaseolus vulgaris* lines Capulet, Edmund and SOA-BN. Significance values determined by one-way analyses of variance are indicated above each plot. DAS: days after sowing.

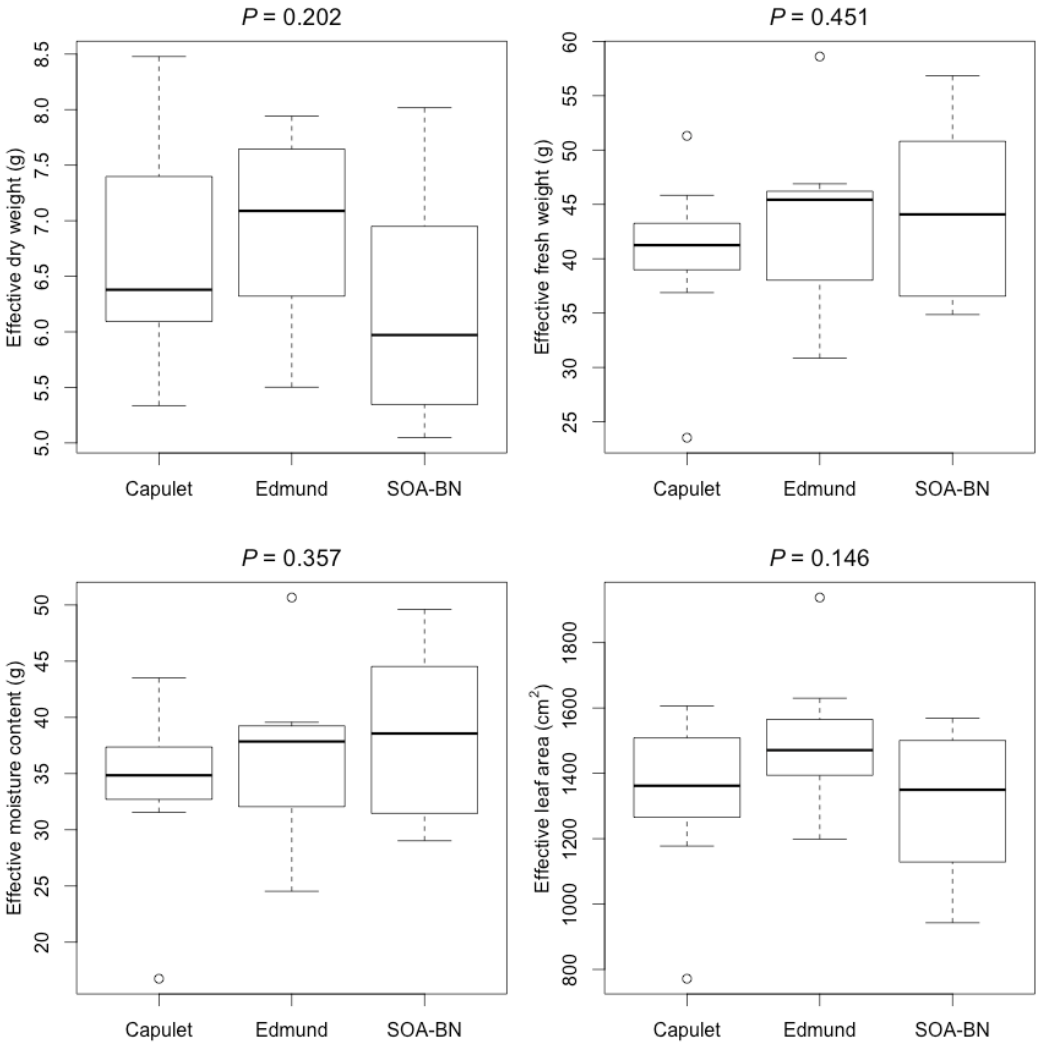


Figure 4.2. Continued.

4.3.2. Root architecture

Contrasting parental-line root architectures

Visible differences were observed between the root architectures of SOA-BN and Edmund (Figures 4.3 and 4.4, respectively). SOA-BN had a significantly larger basal root whorl number (P = 1.22E−05) and, consequently, basal root number (P = 5.99E−05; Table 4.2 and Figure 4.5). Additionally, due to a shallower minimum basal root growth angle (P = 2.21E−05), SOA-BN had a larger range of basal root growth angles than Edmund (P = 2.92E−07). SOA-BN also had a larger taproot diameter at a depth of 1 cm (P = 0.01). However, differences in the lengths of unbroken taproots of SOA-BN and Edmund plants were not significant (P = 0.3). This may be a result of the early developmental stage at which roots were evaluated (21 DAS). Root and shoot dry weights were also significantly greater for SOA-BN compared with Edmund plants (P = 0.008 and P = 0.005).

Figure 4.3. The root system of a SOA-BN plant grown in vermiculite and harvested 21 days after sowing.

Figure 4.4. The root system of an Edmund plant grown in vermiculite and harvested 21 days after sowing.

Table 4.2. Summary of analyses of variance comparing *Phaseolus vulgaris* parental lines SOA-BN and Edmund for seedling root and shoot traits

Trait	d.f.	Sum of squares	Mean square	*F*	Probability (*F*)	LSD
BRN	1	185.79	185.79	36.29	5.99E−05	2.635
BRWN	1	11.161	11.16	50.68	1.22E−05	0.546
BRGA max.	1	44.6	44.64	0.926	0.355	8.087
BRGA min.	1	2187.5	2187.5	44.82	2.21E−05	8.136
BRGA range	1	1607.1	1607.1	103.8	2.92E−07	4.582
BRD	1	0.003	0.003	1	0.337	0.062
BRL min.	1	8.64	8.643	1.901	0.193	2.483
BRL max.	1	0	0	0	1	0.942
BR branch.	1	0.643	0.643	0.529	0.481	1.283
TD	1	0.731	0.731	9.309	0.01	0.326
TL	1	3.71	3.709	1.192	0.298	2.102
T branch.	1	2.57	2.571	0.334	0.574	3.23
ARN	1	4.57	4.571	1.655	0.223	1.935
ARD min.	1	0.001	0.001	0.043	0.839	0.151
ARD med.	1	4E−04	4E−04	0.022	0.884	0.159
ARL min.	1	5.44	5.44	1.176	0.301	2.66
ARL max.	1	15.49	15.49	1.187	0.304	5.299
ARL med.	1	5.95	5.951	0.846	0.377	3.288
Root dry weight	1	100810	100810	10.11	0.008	116.311
Shoot dry weight	1	85645	85645	11.63	0.005	99.93

d.f.: degrees of freedom. LSD: least significant difference. BRN: basal root number. BRWN: mean basal root whorl number per file. BRGA: basal root growth angle. BRD: basal root diameter. BRL: basal root length. BR branch.: basal root branching density. TD: taproot diameter. TL: taproot length. T branch.: taproot branching density. ARN: adventitious root number. ARD: adventitious root diameter. ARL: adventitious root length.

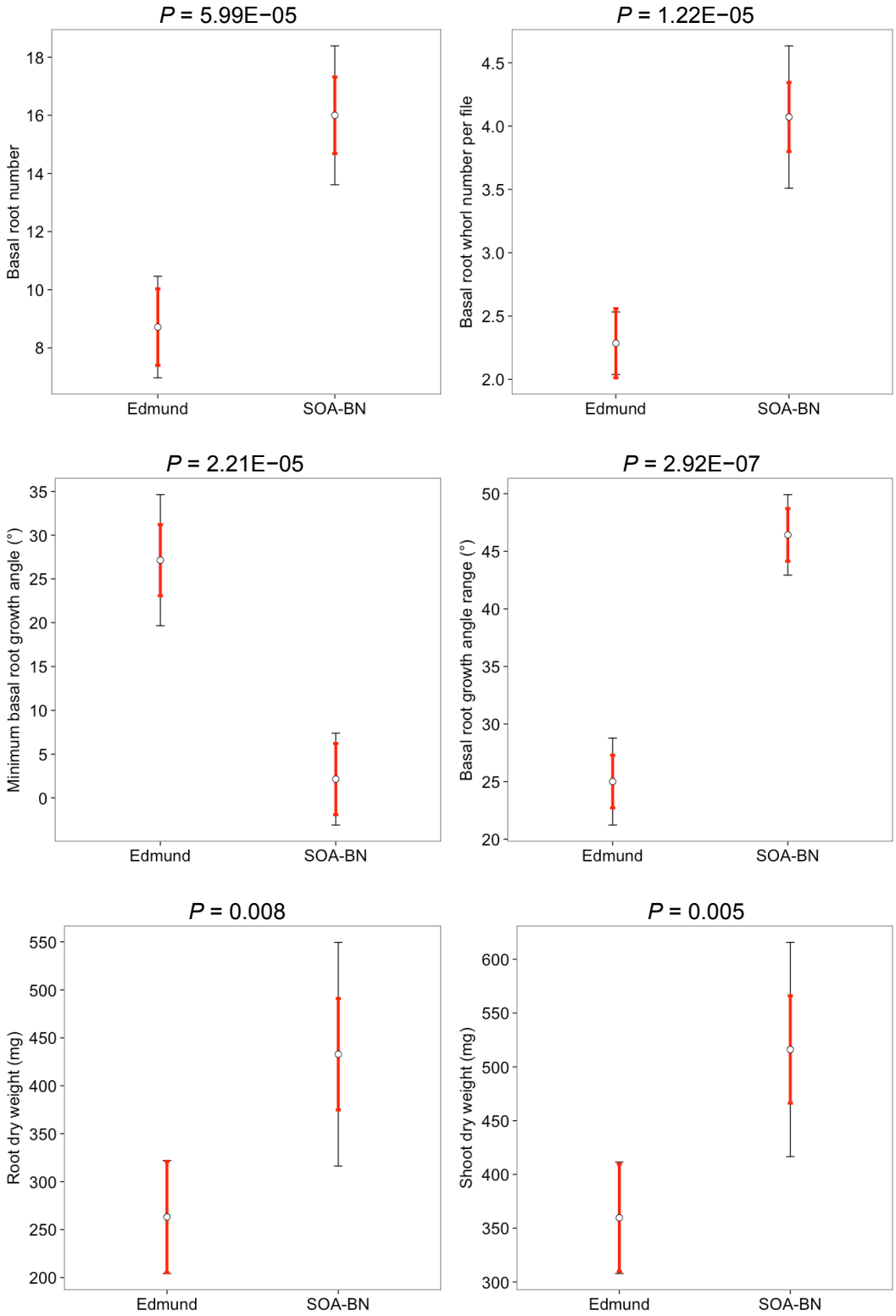


Figure 4.5. Summary of root and shoot phenotypes of *Phaseolus vulgaris* parental lines Edmund and SOA-BN. Means of seven biological replicates (white circles), 95% confidence intervals (thin black lines) and least significant differences (thicker red lines) are shown. Significance values determined by one-way analyses of variance are indicated above each plot.

RIL phenotyping and QTL analysis

Differences between the parental lines were also significant in the RIL phenotyping experiments for all traits investigated (Table 4.3). Unlike in the experiment involving only the parental genotypes, however, maximum basal root growth angle was significantly greater in SOA-BN (mean = 45.0 °; SD = 5.0 °) than in Edmund (mean = 28.3 °; SD = 2.9 °) ($P = 0.007$). This may be a result of growing seedlings in larger (4-L) pots in the RIL phenotyping experiments, and suggests that common bean plants exhibit phenotypic plasticity for basal root growth angles in response to different edaphic conditions. Alternatively, the flaccidity of basal roots at this early developmental stage may have resulted in inconsistent, subjective phenotypic assessments of basal root growth angles between experiments.

Table 4.3. Summary of analyses of variance comparing *Phaseolus vulgaris* parental lines SOA-BN and Edmund for seedling root and shoot traits in RIL assays

Trait	d.f.	Sum of squares	Mean square	*F*	Probability (*F*)	LSD
BRN	1	339.1	339.1	70.63	5.14E−09	1.672
BRWN	1	18.017	18.02	100.2	1.39E−10	0.324
BRGA max.	1	416.7	416.7	25	0.007	9.255
BRGA min.	1	253.5	253.5	20.28	0.012	8.015
BRGA range	1	1320.2	1320.2	79.21	8.81E−04	9.255
Taproot diam.	1	1.373	1.373	51	0.002	0.372
Shoot diam.	1	1.927	1.927	228.9	1E−04	0.208
Root dry weight	1	239201	239201	17.82	0.013	262.675
Shoot dry weight	1	58806	58806	15.29	0.017	140.585

d.f.: degrees of freedom. LSD: least significant difference. BRN: basal root number. BRWN: mean basal root whorl number per file. BRGA: basal root growth angle. Diam.: diameter.

For the SOA-BN × Edmund RIL population phenotypic data, least-squares means were derived for basal root growth angle range, minimum basal root growth angle, and shoot dry weight based on significant random-effects variance components within mixed models (Table A8.1, APPENDIX 8). Replication effects were not detected for other traits. A continuum of phenotypic variation was observed within this population for most of the traits investigated (Figure 4.6). Hypocotyl colour and seed coat colour were characterised by discrete phenotypic categories.

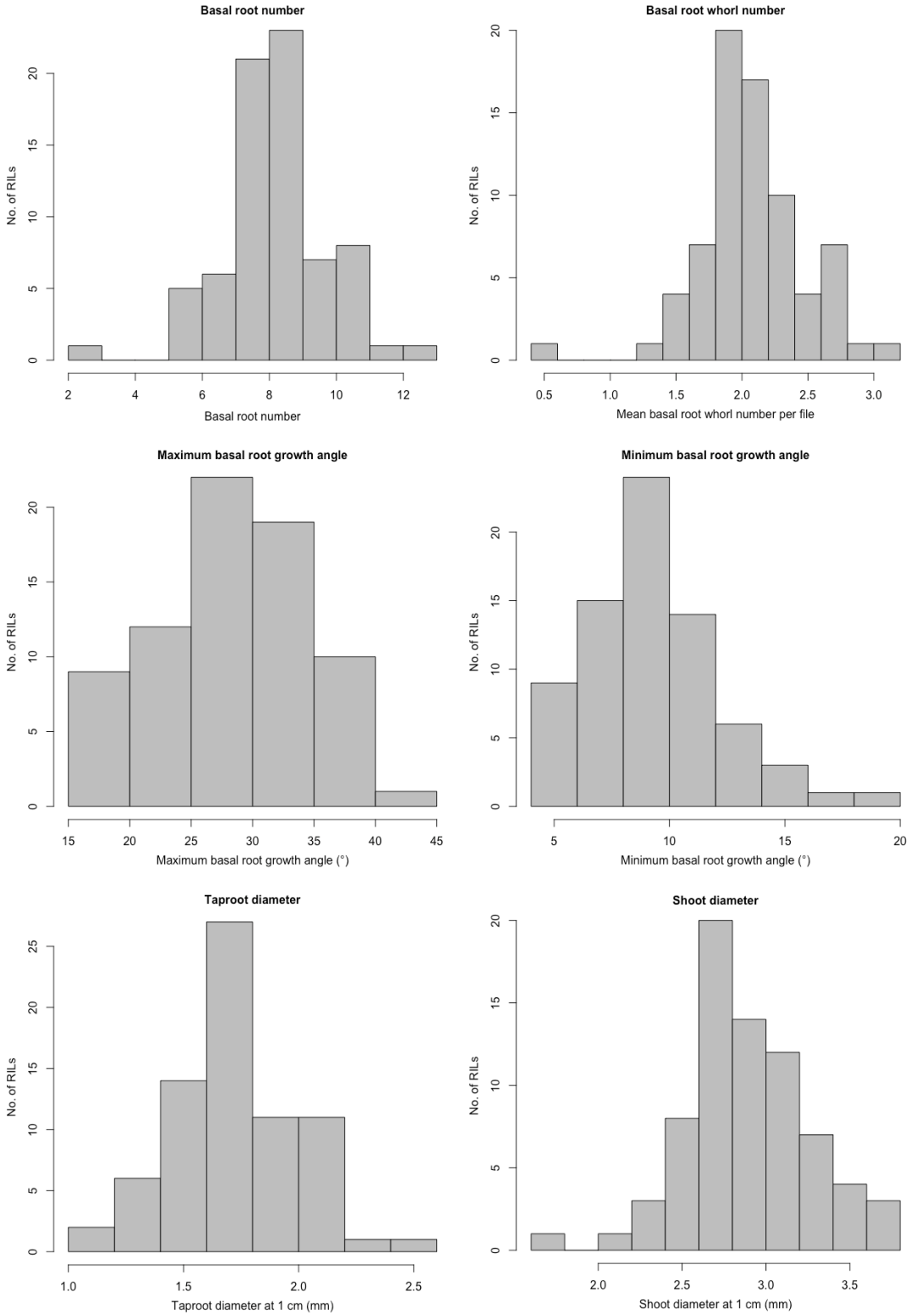


Figure 4.6. Distribution of phenotypic variation for seedling root and shoot traits evaluated 20–22 days after sowing and for seed coat pigmentation in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population.

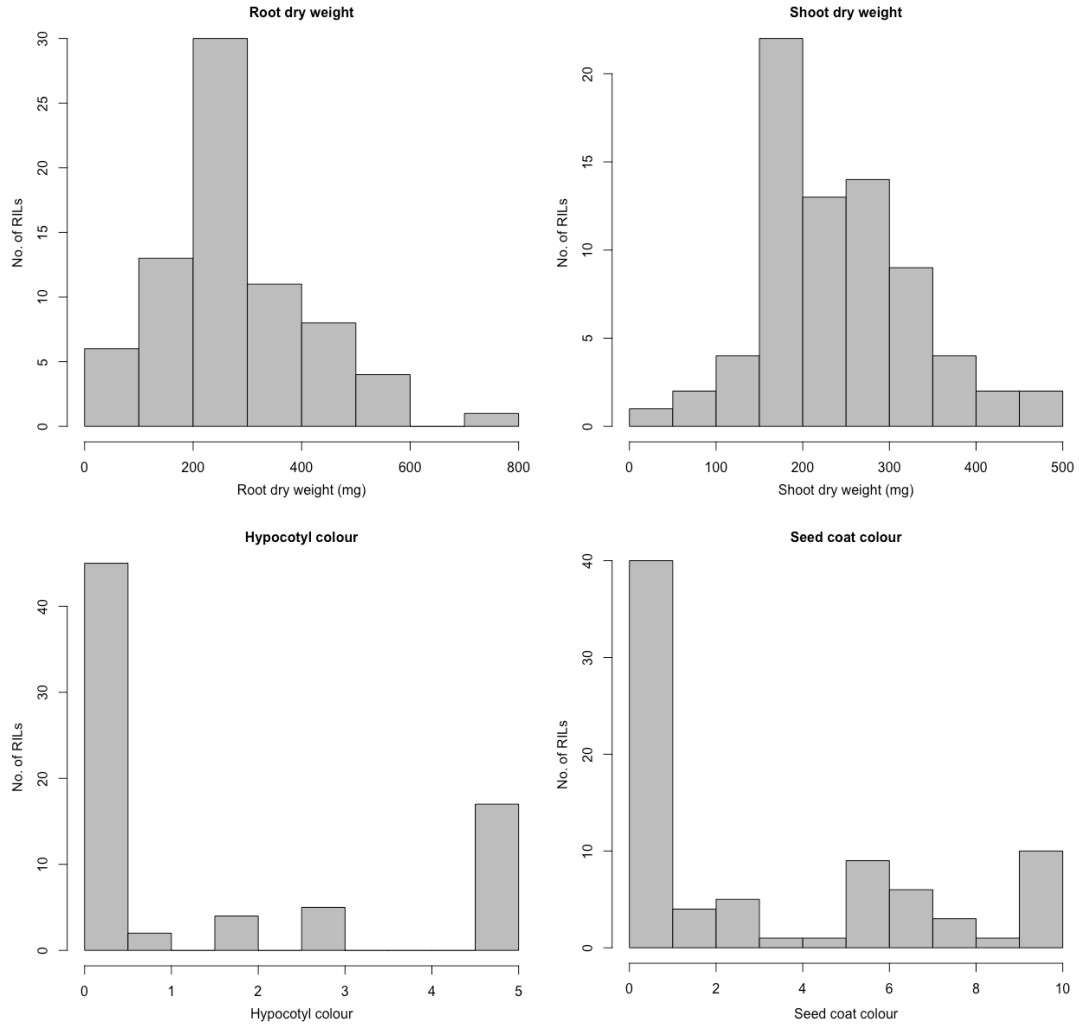


Figure 4.6. Continued.

QTL models for each trait are summarised in Tables 4.4 and 4.5. Models presented in Table 4.5 were fitted using loci with the highest LOD scores not estimated to be significant ($P > 0.05$) by one- and two-dimensional genome scans. One-dimensional genome scans for single major-effect QTL identified loci on chromosomes Pv06 and Pv07 with LOD scores significant at the 0.001 probability level for seed coat colour and hypocotyl colour (Figure 4.7). Two-dimensional genome scans identified additional QTL with additive and interactive effects for these traits (Table 4.4).

Putative QTL partly colocalising on chromosome Pv07 were identified for taproot diameter at 1 cm deep, shoot diameter at 1 cm aboveground, maximum and range of basal root growth angles, basal root number, mean basal root whorl number per file, and root and shoot dry weights (Figure 4.7, and Tables 4.4 and 4.5). In each case, the potentially desirable allele(s) are derived from SOA-BN (denoted by negative values for additive effect) and are linked in coupling phase with dominant *P*

on Pv07, which potentiates pigmentation in the seed coat, flower and hypocotyl (Tables 4.4 and 4.5). With the exception of the QTL on Pv07 for basal root growth angle range, each of these mapping intervals also overlaps the *P* locus. Correlation analysis revealed strong relationships amongst numerous root and shoot traits (Table 4.6).

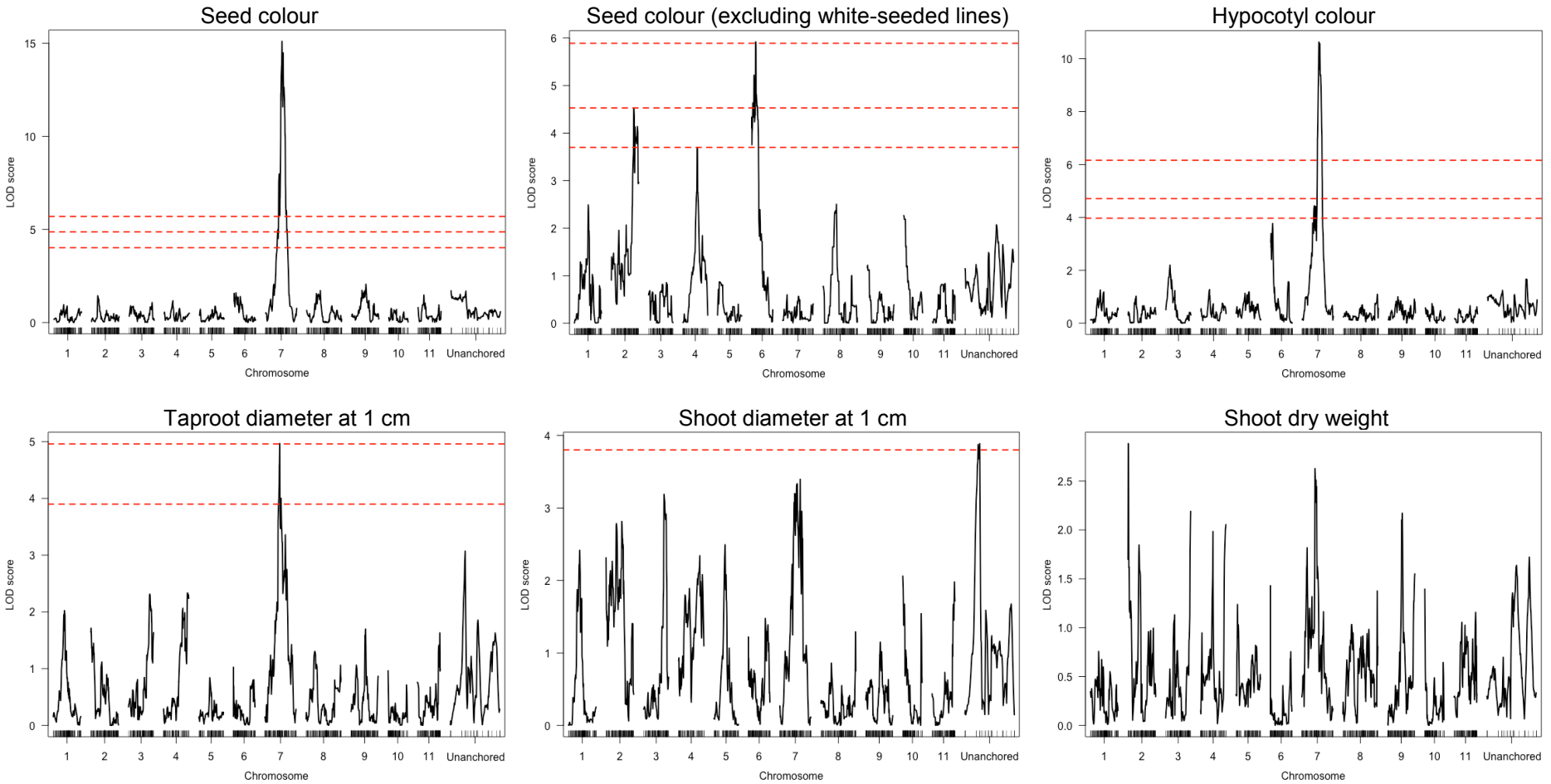


Figure 4.7. LOD profiles obtained by one-dimensional genome scans for seedling root and shoot traits evaluated 20–22 days after sowing and for seed coat pigmentation in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population. Red dashed lines denote significance at the 0.001 (upper of three lines), 0.01 (upper of two lines) and 0.05 (one line) probability levels.

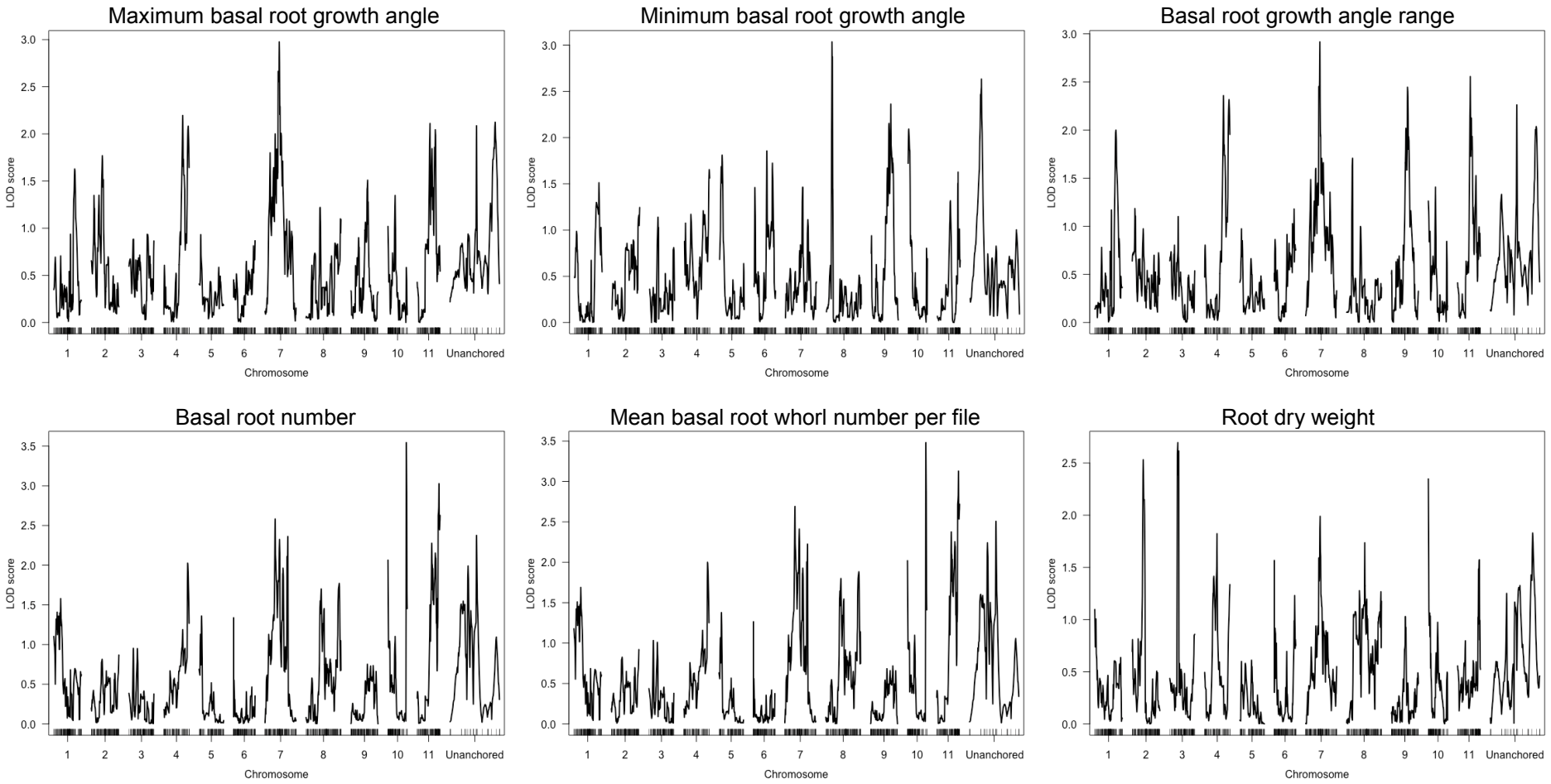


Figure 4.7. Continued.

Table 4.4. Single- and multi-QTL models fitted and refined by multiple imputation for seedling root and shoot traits evaluated 20–22 days after sowing and for seed coat pigmentation in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population

Trait	Chromosome	Start	End	Size (Mb)	d.f.	LOD	R^2 (%)	Probability (*F*)	Additive effect	Dominance deviation
Taproot diameter	7	6,446,596	46,677,962	40.23	2	4.97	26.93	1.70E−05***	−0.14421	0.13568
Hypocotyl colour	7	41,123,382	41,693,662	0.57	10	66.70	79.62	< 2E−16***	−0.8861	−0.4572
	6	1,731,266	15,560,291	13.83	6	51.61	30.00	< 2E−16***	0.7275	0.928
	2	603,969	957,431	0.35	6	38.19	12.18	< 2E−16***	0.2375	−2.0487
	7 × 2	N/A	N/A	N/A	4	37.22	11.38	< 2E−16***	N/A	N/A
	7 × 6	N/A	N/A	N/A	4	44.36	18.55	< 2E−16***	N/A	N/A
Seed colour	7	38,913,940	41,060,962	2.15	10	33.88	77.93	< 2E−16***	−2.9536	−0.4153
	2	603,969	701,581	0.10	6	11.37	11.95	8.17E−08***	0.3879	1.0415
	6	19,094,701	22,908,114	3.81	6	9.13	8.93	4.05E−06***	1.1091	−0.4387
	7 × 2	N/A	N/A	N/A	4	8.26	7.87	2.57E−08***	N/A	N/A
	7 × 6	N/A	N/A	N/A	4	4.50	3.82	1.85E−03**	N/A	N/A
Seed colour (without *pp*)	6	3,699,949	21,618,497	17.92	2	7.38	35.53	3.51E−07***	1.8461	0.5999
	2	42,984,441	48,956,862	5.97	2	5.60	24.04	1.26E−05***	1.3865	−2.284

Abbreviations are as described for Table 2.1 in Chapter 2.

Table 4.5. Provisional single- and multi-QTL models fitted and refined by multiple imputation for root and shoot traits evaluated 20–22 days after sowing, based on loci with the highest LOD scores not estimated to be significant ($P > 0.05$) by one- and two-dimensional genome scans

Trait	Chromosome	Start	End	Size (Mb)	d.f.	LOD	R^2 (%)	Probability (*F*)	Additive effect	Dominance deviation
BRGA range	4	45,237,099	45,760,843	0.52	2	4.72	18.03	5.36E−05***	−2.5454	−11.9978
	11	3,656,216	4,687,504	1.03	2	4.60	17.47	7.00E−05***	−3.1	5.9866
	7	6,446,596	9,119,137	2.67	2	4.07	15.22	2.08E−04***	−2.5617	6.6622
BRGA max.	7	938,189	43,452,259	42.51	2	2.97	17.07	1.43E−03**	−2.4458	3.6903
BRGA min.	8	1,231,860	1,736,363	0.50	2	3.021	11.61	2.25E−03**	0.3256	7.796
	9	17,873,362	36,087,684	18.21	2	2.42	9.13	7.52E−03**	0.8786	−2.1985
	Unanchored	N/A	N/A	N/A	2	1.78	6.58	2.74E−02*	0.1121	1.7088
	10	34,883,408	43,042,087	8.16	2	1.48	5.41	5.06E−02	0.439	2.0914
BRN	10	533,196	42,872,814	42.34	2	3.80	15.69	3.63E−04***	0.0239	−6.1654
	7	1,473,344	49,255,394	47.78	2	2.67	10.63	3.83E−03**	−0.4919	−0.7898
	11	670,527	49,989,392	49.34	2	2.03	7.91	1.45E−02*	−0.4719	0.191
Mean BRWN per file	10	42,784,897	42,872,814	0.09	2	4.31	17.26	1.27E−04***	0.007724	−1.564614
	7	2,281,649	49,255,394	46.97	2	2.77	10.56	3.11E−03**	−0.125636	−0.193863
	11	1,017,267	49,989,392	48.97	2	2.13	7.96	1.18E−02*	−0.121465	0.059913
Root dry weight	10	533,196	40,822,992	40.29	2	2.68	11.91	3.75E−03**	27.652	−157.667
	3	1,978,782	2,802,883	0.82	2	2.68	11.90	3.76E−03**	2.115	263.671
	7	85,550	51,695,284	51.61	2	1.67	7.17	3.09E−02*	−27.613	94.907
Shoot dry weight	2	21,024	48,969,580	48.95	2	2.69	13.23	3.12E−03**	−6.276	113.994
	7	113,683	48,942,564	48.83	2	2.43	11.85	5.43E−03**	−28.962	43.647
Shoot diameter	2	21,024	3,103,787	3.08	2	4.39	16.67	1.07E−04***	−0.15737	0.11932
	Unanchored	N/A	N/A	N/A	2	3.79	14.09	3.75E−04***	−0.20134	0.04366
	7	9,026,655	49,255,394	40.23	2	3.37	12.38	8.91E−04***	−0.11066	−0.40007

BRGA: basal root growth angle. BRN: basal root number. BRWN: mean basal root whorl number per file. Other abbreviations are as described for Table 2.1 in Chapter 2.

Table 4.6. Pearson correlation coefficients for seedling root and shoot traits evaluated 20–22 days after sowing and for seed coat pigmentation in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population

Trait	BRWN	BRGA min.	BRGA max.	BRGA range	Taproot diam.	Shoot diam.	Root dry weight	Shoot dry weight	Hypocotyl colour	Seed colour
BRN	1.00***	−0.57***	0.65***	0.73***	0.55***	0.56***	0.48***	0.55***	0.08^{ns}	0.06^{ns}
BRWN		−0.56***	0.65***	0.72***	0.54***	0.55***	0.47***	0.54***	0.09^{ns}	0.06^{ns}
BRGA min.			−0.36**	−0.64***	−0.22^{ns}	−0.37**	−0.15^{ns}	−0.22^{ns}	0.07^{ns}	0.03^{ns}
BRGA max.				0.93***	0.55***	0.56***	0.56***	0.62***	0.22^{ns}	0.24*
BRGA range					0.54***	0.59***	0.51***	0.60***	0.15^{ns}	0.19^{ns}
Taproot diam.						0.81***	0.62***	0.73***	0.29*	0.39***
Shoot diam.							0.48***	0.51***	0.20^{ns}	0.28*
Root dry weight								0.83***	0.27*	0.27*
Shoot dry weight									0.17^{ns}	0.28*
Hypocotyl colour										0.81***

BRN: basal root number. BRWN: mean basal root whorl number per file. BRGA: basal root growth angle. Diam.: diameter at 1 cm from the base of the hypocotyl.
*, **, ***: significant at the 0.05, 0.01 and 0.001 probability levels. ^{ns}: not significant.

4.3.3. Polytunnel and field evaluation

LOD profiles generated by one-dimensional genome scans for single QTL are summarised in Figures 4.8 (polytunnel) and 4.9 (field). Single- and multi-QTL models for selected traits are summarised in Table 4.7, and include loci with additive and/or interactive effects whose LOD scores were determined to be significant by permutation tests for one- or two-dimensional genome scans.

Plants of 21 of the 80 SOA-BN × Edmund $F_{6:7}$ RILs for which genotyping-by-sequencing data were obtained failed to germinate or emerge in the hand-sown field plot of 2015. Poor germination and emergence rates also resulted in a low number of plants per RIL (mean number of emerged plants per RIL = 1.69; range = 1–4; SD = 0.88). Hand-sown beans were slower to emerge than machine-drilled beans, and growing conditions were not adverse during the period from sowing to emergence. Time to harvest maturity was generally longer for hand-sown RILs (mean = 126.62 days; range = 98–153 days; SD = 12.50 days) compared with machine-drilled SOA-BN and Capulet (105 days). Comparatively shallow planting depth (~ 2–3 cm) and insufficient soil moisture may have caused delayed emergence and maturity of plants in the hand-sown plot. The data collected and mapping results presented should therefore be considered as provisional. This exploratory analysis using pilot field data could inform future well-replicated RIL field trials that would provide a more robust test of adaptation to UK growing conditions. Seed increase achieved in the polytunnel and field in 2015 will enable the design of replicated field trials.

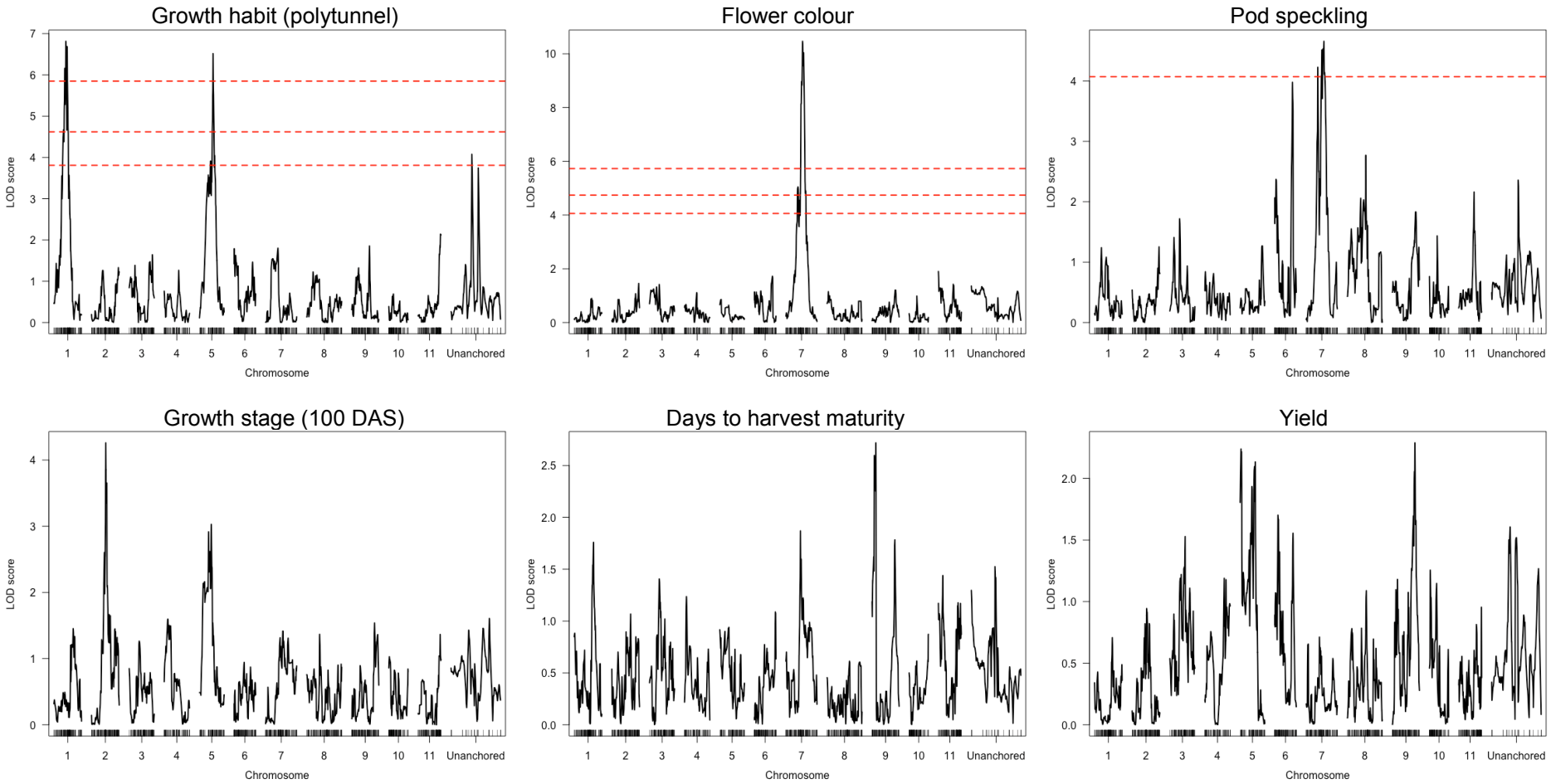


Figure 4.8. LOD profiles obtained by one-dimensional genome scans for traits evaluated in the polytunnel in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population. Red dashed lines denote significance at the 0.001 (upper of three lines), 0.01 (upper of two lines) and 0.05 (one line) probability levels. DAS: days after sowing.

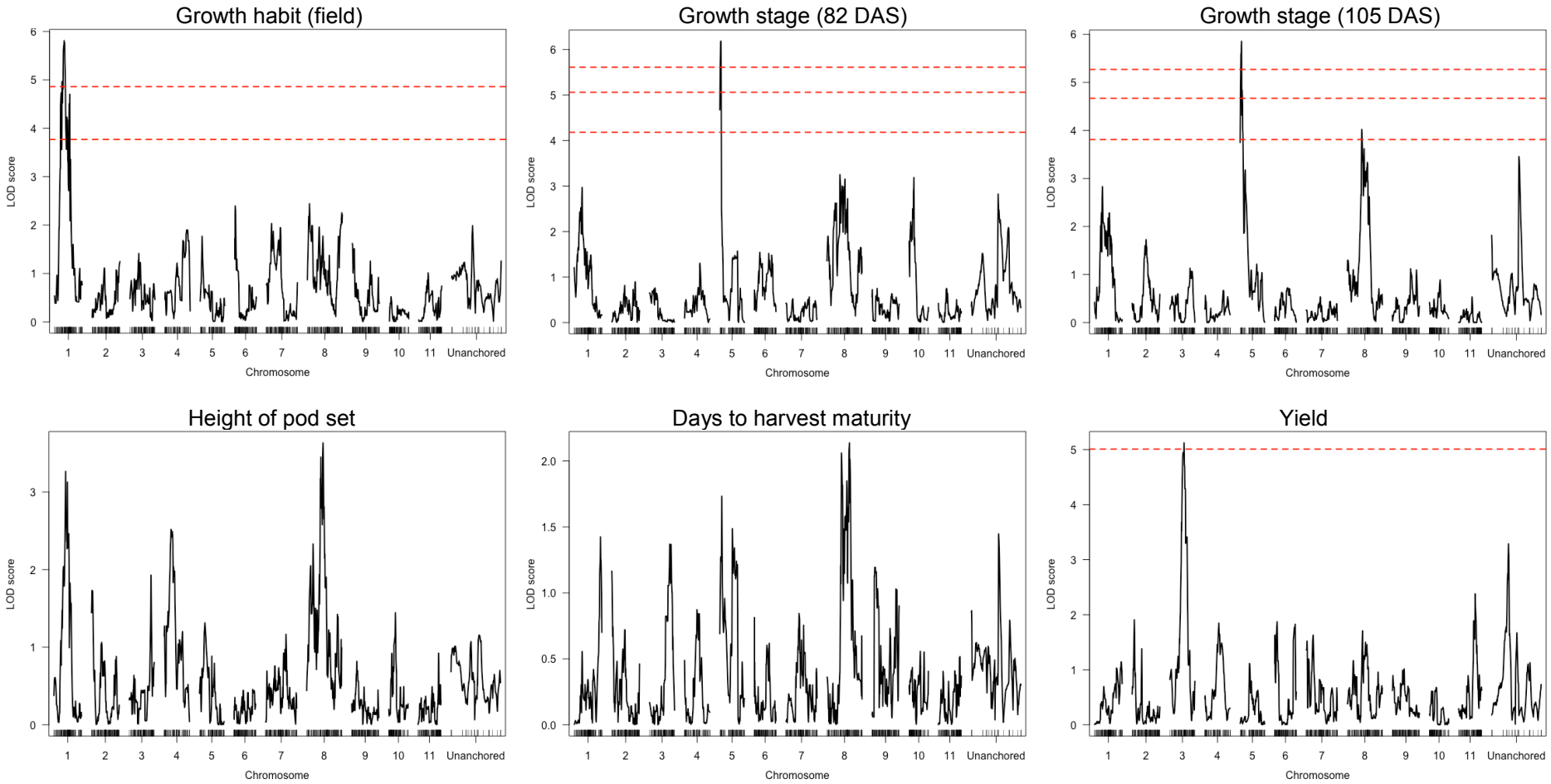


Figure 4.9. LOD profiles obtained by one-dimensional genome scans for traits evaluated in the field in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population. Red dashed lines denote significance at the 0.001 (upper of three lines), 0.01 (upper of two lines) and 0.05 (one line) probability levels. DAS: days after sowing.

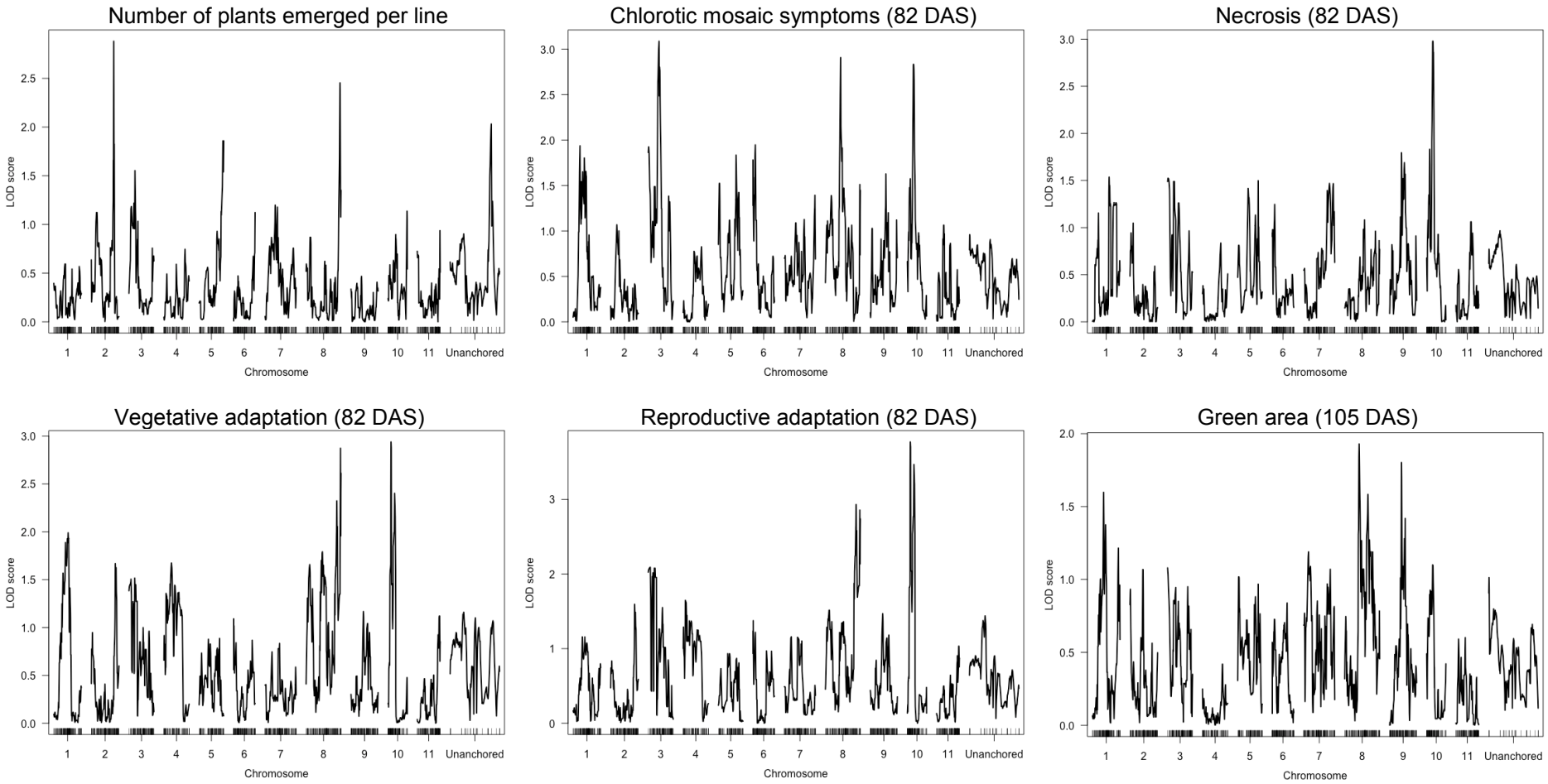


Figure 4.9. Continued.

Table 4.7. Single- and multi-QTL models fitted and refined by multiple imputation for developmental, morphological and reproductive traits evaluated in polytunnel and field experiments in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population

Trait	Chromosome	Start	End	Size (Mb)	d.f.	LOD	R^2 (%)	Probability (*F*)	Additive effect	Dominance deviation
Polytunnel										
Growth habit	1	4,556,729	32,857,505	28.30	2	8.20	27.04	2.39E−08***	0.67266	0.63144
	5	10,469,238	34,622,922	24.15	2	8.02	26.30	3.48E−08***	0.6218	−1.0291
Flower colour	7	37,406,924	43,452,259	6.04	2	10.46	48.79	9.39E−11***	−0.8445	−0.4289
Pod speckling	7	3,509,864	43,452,259	39.94	2	6.10	25.66	2.18E−06***	−0.17313	−0.29527
	6	30,233,555	30,983,282	0.75	2	5.91	24.68	3.28E−06***	0.12432	0.57735
Field										
Growth habit	7	1,750,632	2,679,997	0.93	6	11.33	36.45	7.24E−08***	−0.12896	0.53612
	2	47,436,913	48,956,862	1.52	6	10.87	34.29	1.61E−07***	0.07323	0.42713
	1	3,149,566	6,127,233	2.98	2	5.13	12.62	6.76E−05***	0.37148	1.25125
	2 × 7	N/A	N/A	N/A	4	10.59	33.00	3.49E−08***	N/A	N/A
Pod height	8	1,076,650	11,132,910	10.06	2	3.63	24.70	3.55E−04***	−0.4906	−0.1113
Growth stage (82 DAS)	5	16,742	381,858	0.36	2	5.129	19.57	3.02E−05***	−0.29936	−1.13776
	1	728,612	51,306,251	50.58	2	3.74	13.47	5.07E−04***	−0.33386	−0.60247
	10	533,196	42,954,852	42.42	2	3.12	10.98	1.77E−03**	0.32149	0.49132
Growth stage (105 DAS)	5	16,742	821,451	0.80	2	4.89	23.34	3.38E−05***	−0.33715	−0.69929
	8	3,141,669	53,934,965	50.79	2	3.09	13.65	1.52E−03**	0.29363	−0.29155
Yield	2	523,057	1,043,421	0.52	6	9.69	33.24	1.28E−06**	−6.800	1.422
	Unanchored	N/A	N/A	N/A	6	9.40	31.84	2.11E−06***	ND	ND
	3	3,737,709	33,224,395	29.49	2	5.59	16.07	2.85E−05***	−11.941	21.017
	2 × Unanc.	N/A	N/A	N/A	4	7.73	24.34	6.10E−06***	N/A	N/A

Two-dimensional genome scans were not performed for flower colour and pod speckling data. ND: not determined. Other abbreviations are as described for Table 2.1 in Chapter 2.

Correlation analysis revealed a significant positive relationship between growth habit as evaluated in the polytunnel and as evaluated in the field in the SOA-BN × Edmund RIL population (Table 4.8). However, due to the limited correlation (R = 0.31), field-based selection for upright determinacy and other plant architectural traits amenable to mechanical harvest is advisable. Field–polytunnel correlations for other traits investigated were not significant.

Table 4.8. Pearson correlation coefficients between phenotypes observed in the polytunnel and in the field for selected traits in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population

Polytunnel	Field	Pearson *R*	Probability	d.f.
Growth habit	Growth habit	0.31	0.022*	51
Yield	Yield	0.24	0.089	51
Growth stage (100 DAS)	Growth stage (105 DAS)	0.15	0.290	48
Growth stage (100 DAS)	Growth stage (82 DAS)	0.13	0.365	49
Days to harvest maturity	Days to harvest maturity	0.07	0.601	49

d.f.: degrees of freedom. DAS: days after sowing.
*: significant at the 0.05 probability level.

4.3.4. Candidate-gene variants

A candidate-gene approach to the identification of sequence variants was adopted to enable development of polymorphic markers that may be linked to traits of agronomic importance. This reverse genetics approach is informed by the crop ideotype for adaptation to UK growing conditions and is complementary to the forward genetics approach described above. This research has generated an RNA-seq-derived database of sequence variants occurring within adaptive genes.

A list of candidate genes with previously described putative functions in physiological resilience and morphological traits in common bean and other legumes was compiled (Table A9.1, APPENDIX 9). This includes genes associated with cold and drought tolerance, earliness of maturity, acquisition and use of nutrients, plant architecture, and seed morphological and nutritional characteristics. The physical positions of these genes in the common bean reference genome (Schmutz *et al*., 2014) were determined or predicted based on previously reported coordinates or BLASTN comparisons.

Transcriptome data for the lines SOA-BN, Edmund and Capulet were generated, and single-nucleotide variants (SNVs) and insertions/deletions (INDELs) identified as described in Chapter 2. Variants occurring within candidate genes were annotated with their predicted impacts on gene function using SnpEFF version 4.1 (Cingolani *et al*., 2012). A progression of custom Perl scripts was used to identify variant alleles within candidate genes that are polymorphic between the three lines and to generate files amenable to conversion into polymorphic markers for genotyping. Specifically, reference-derived sequences flanking each candidate-gene variant (200 bp on either side) were retrieved to generate these Assay Development Format (ADF) files. International Union of Pure and Applied Chemistry (IUPAC) ambiguity codes were included in ADF files where variant sites occurred within flanking sequences. ADF file sequences were compared with the reference genome by BLASTN analysis to ensure only one match per variant-containing sequence, using a 'word size' of 11 bp to detect regions of similarity. This step helps to ensure that primers designed based on variant-flanking sequences are specific to the intended target and generate only one amplicon.

4.4. DISCUSSION

This research has revealed putative QTL for physiological and morphological traits required for adaptation of common bean to UK growing conditions. It has demonstrated the utility of genotyping-by-sequencing for high-resolution linkage mapping of adaptive traits in a bi-parental recombinant inbred population of common bean. Water-use efficiency was not found to differ between the parental lines under non-stress conditions. However, evaluation of phenotypic variation within the inbred population with regard to contrasting parental root traits enabled detection of QTL that may be implicated in vital resilience characteristics. Preliminary phenotypic and mapping data from pilot polytunnel and field evaluations provide the basis for further field-based investigation of heritable developmental, morphological and reproductive traits, with ultimate application in marker-assisted breeding.

Root architecture

Correlation analysis detected strong relationships amongst seedling root and shoot architectural traits (Table 4.6). Thus, a useful breeding strategy would be to apply marker-assisted selection for those traits with the greatest heritability. For example, selection for larger taproot diameter, with 26.93% of phenotypic variation explained by the QTL on chromosome Pv07, may enable the accumulation of functional alleles on Pv07 governing other potentially desirable root traits. Similarly, selection for alleles derived from SOA-BN at the putative QTL on Pv11 for range of basal root growth angles (accounting for 17.47% of variation) may enable incorporation of linked alleles governing basal root whorl number and, by extension, basal root number.

Colocalisation of putative QTL for different root and shoot traits indicates that genes on chromosome Pv07 may be important determinants of root architectural phenotypes in common bean (Tables 4.4 and 4.5). These findings are consistent with those presented by Miguel *et al*. (2009, unpublished data), who reported a QTL on Pv07 for basal root whorl number (*Rwn*7.1) that accounts for 24.75% and 58.72% of phenotypic variation observed in DOR364 × G19833 and G2333 × G19839 RIL populations, respectively. They concluded that basal root whorl number is therefore likely to be governed by few genes. Similarly, they mapped a QTL for basal root

number to approximately the same location on Pv07 (*Rno*7.1), which explains 14.25% and 61.4% of variation observed in the two RIL populations. In the DOR364 × G19833 population, the desirable allele(s) on Pv07 were derived from the reference-genome genotype G19833, an inbred Andean landrace with low phosphorus (P) tolerance, a brown seed coat and an indeterminate growth habit (Cichy *et al*., 2009; Schmutz *et al*., 2014). Castaño (2013) also reported QTL on Pv07 for basal root whorl number and basal root number in the DOR364 × G19833 population. Miguel *et al*. (2009, unpublished data) proposed that basal root whorl number could be used for selection of lines with improved performance under low P conditions. Common bean cultivars typically possess one to four and, exceptionally, five basal root whorls (Miguel *et al*., 2013). Three to six whorls were observed on SOA-BN roots, indicating the potential value of this genotype as a breeding line.

Miguel *et al*. (2013) found that common bean genotypes with a larger basal root whorl number had shallower root systems with a larger range of basal root growth angles. This is consistent with the strong positive correlation between basal root whorl number and basal root growth angle range observed in the SOA-BN × Edmund RIL population ($R = 0.72$) (Table 4.6). Miguel *et al*. (2013) also reported a strong positive correlation between basal root whorl number and P acquisition in low P soil in the glasshouse ($R^2 = 0.64$) and in the field ($R^2 = 0.88$). P acquisition is improved by shallow basal root growth angles due to an enhanced capacity for foraging of the topsoil, where P is concentrated (Miguel *et al*., 2013, 2015). Miguel *et al*. (2015) subsequently reported a synergistic effect upon P acquisition when the architectural phene of shallower basal root growth is combined with the anatomical phene of greater root hair length and density. Cichy *et al*. (2009) reported QTL on Pv07 and Pv11 for P acquisition under low P conditions in a AND696 × G19833 RIL population. It would therefore be informative to determine whether SOA-BN exhibits enhanced P uptake capability in low P soil compared with Edmund, with a view to QTL mapping based on phenotypic variation observed amongst RILs.

The major-effect QTL on Pv07 for taproot diameter, which was significant in analyses using both one-dimensional and two-dimensional genome scans, may be particularly useful for marker-assisted breeding for resilience traits (Table 4.4 and Figure 4.7). Taproot diameter is often used as an indicator of rooting depth in the field, where accurate measurements of length are not possible due to taproot breakage during root excavation (Burridge, 2012). The assumption underlying this

approach is that a taproot with a larger diameter is able to extend deeper to conduct water that is unavailable in the upper soil layers, which is a desirable trait under drought (Burridge, 2012).

Putative QTL for seedling root architecture traits evaluated under glasshouse conditions require validation through field evaluations, particularly in view of phenotypic plasticity in response to different edaphic conditions. Assessment at different developmental stages would also be valuable for determining whether phenotypes observed during the vegetative stages of growth are predictive of the extensiveness of root architectures observed during the reproductive growth stages. This can be achieved through application of the high-throughput "shovelomics" phenotyping methodology for visual evaluation of root crowns excavated from the field developed by Lynch and Brown (c2016; also see Burridge, 2012; and Burridge *et al*., 2016). Field evaluations during the vegetative growth stages will nonetheless be important for enabling selection for root architectural phenes that may enhance tolerance of stress conditions occurring during the early stages of growth, such as low temperatures.

Root QTL linked in coupling phase with pigmentation factor *P*

A UK breeding programme that aims to introgress genes governing physiological resilience into a white-seeded bean with haricot characteristics may require that linkage of potentially desirable root traits in coupling phase with dominant *P* on chromosome Pv07 is broken. Identification of white-seeded lines with alleles from SOA-BN at putative QTL for root architecture traits on Pv07 could prove valuable for the development of backcross populations in which desirable alleles are accumulated.

The 'ground factor' gene *P* governing seed coat pigmentation maps to the same locus as flower colour and hypocotyl colour on chromosome Pv07 (Figures 4.7 and 4.8, and Tables 4.4 and 4.7). The dominant allele potentiates coloured seed coats and flowers (Hosfield, 2001) and has been previously mapped to Pv07 (see McClean *et al*., 2002, and references therein; and Reinprecht *et al*., 2013). This is consistent with an observed increasing proportion of white-seeded (*pp*) SOA-BN × Edmund RILs with each generation of inbreeding. Removing all RILs homozygous for Edmund-derived *p* from the genetic map enabled localisation of the 'violet factor' gene *V* on chromosome Pv06 (Figure 4.7 and Table 4.4), which agrees with previous

mappings (McClean *et al*., 2002, and references therein; Reinprecht *et al*., 2013). In the presence of dominant *P*, the dominant allele at *V* confers a blue/violet to black seed coat (Hosfield, 2001; McClean *et al*., 2002). Parental line SOA-BN (*PP*, *vv*) has a brown seed coat and pink flowers, whereas Edmund (*pp*, *VV*) has white seeds and flowers.

Field evaluation

Pilot data and provisional QTL for traits evaluated in the field and polytunnel provide the basis for the investigation of agronomic traits that appear to exhibit useful levels of heritability for marker-assisted breeding. A lack of colocalisation of QTL potentially governing traits that were evaluated in the field with putative QTL for the same traits evaluated in the polytunnel (Table 4.7, and Figures 4.8 and 4.9), and a lack of correlation between field- and polytunnel-collected phenotype data (Table 4.8), highlight the requirement for field evaluations of characteristics of agronomic importance.

Nonetheless, the major-effect QTL for growth habit detected on chromosome Pv01 using both field and polytunnel evaluations may correspond to the *fin* locus for determinacy in common bean (Kwak *et al*., 2008). *PvTFL1y* encoding the common bean orthologue of Arabidopsis TERMINAL FLOWER 1 was previously reported to cosegregate with *fin* (Kwak *et al*., 2008; Repinksi *et al*., 2013). *PvTFL1y* was predicted by BLASTN comparison with the common bean reference genome to be located on Pv01 from 45,561,512 bp to 45,563,326 bp. However, the 1.8-LOD support intervals of the two QTL for growth habit mapped to Pv01 in the current research do not overlap these coordinates (Table 4.7).

Further research

A further assay that may yield more insight into the physiological basis of cold tolerance in SOA-BN could be carried out in a thermal gradient tunnel. For instance, plants of each parental and RIL genotype would be grown under different temperature regimes (within a range spanning ‘suboptimal’ and ‘optimal’ for bean growth), and assessed with regard to time to emergence, emergence rate, vegetative and reproductive development, yield metrics, and WUE indicators evaluated at key developmental stages (vegetative, pollination, flowering and grain filling).

De Ron *et al.* (2016) described methods for the phenotypic evaluation of common bean genotypes under different temperatures in controlled-environment chambers and in the field. They highlighted that more reliable assessments of cold tolerance were obtained under field conditions, citing inconsistencies between results derived from the two growing environments. In field trials conducted in Pontevedra, Spain, De Ron *et al.* (2016) used three staggered planting dates in April, May and June to evaluate differences between 28 common bean genotypes in their phenotypic responses to low, moderate and warm temperature conditions. They measured time to emergence, emergence rate, and several vegetative growth, reproductive growth and yield traits. These methods could be applied to confirm the reported superior cold tolerance of SOA-BN compared with other cultivars (Dodd and Taylor, 1991, unpublished data), and to then evaluate the potentially segregating SOA-BN × Edmund RIL population to enable genetic mapping and marker-assisted breeding for cold tolerance.

An experiment to evaluate the effects of water stress and rehydration on these genotypes with regard to WUE indicators might also provide more insight into the basis of potential drought tolerance in SOA-BN. This would involve limiting water to plants at key developmental stages and rehydrating when the pre-dawn leaf water potential (Ψ_1) approaches the minimum required for full recovery in beans (−1.5 MPa; Boyer, 1978). These experiments would require replication of both treated and non-stressed control plants and, consequently, would have substantial labour and time requirements if scaled up to a large phenotyping study of recombinant inbred populations to enable genetic mapping.

Chapter 5

General discussion

5.1. CONCLUSIONS

The advent of next-generation sequencing has facilitated remarkable progress in plant and pathogen genetics and genomics for crop improvement. In the current research, genotyping-by-sequencing of a common bean recombinant inbred population and whole-genome sequencing of pathogen isolates were exploited for high-resolution linkage and association mapping of genes encoding candidate plant disease resistance proteins and bacterial effectors.

Marker-assisted breeding strategy for halo blight resistance

This research has identified polymorphic markers that delimit a 500-kb mapping interval for quantitative, potentially race-nonspecific, halo blight resistance effective against *Psph* race 6 derived from the El Salvadoran red dry bean landrace PI 150414. It is anticipated that these markers will be tested on additional populations as part of African breeding programmes to determine their accuracy and applicability (G. Ascough, personal communication; B. Raatz, personal communication). In addition to target enrichment sequencing of Pv04 candidate genes in halo blight-resistant and susceptible lines, seed increase of SOA-BN × Edmund RILs that are heterozygous at markers across or within the Pv04 mapping interval (3.1%) will enable potential segregation for halo blight resistance to be tracked, thereby aiding efforts to resolve this locus. Residual heterozygosity of 5.6% at the 1,248 genome-wide markers is consistent with proportions expected at this stage of inbreeding based on theoretical probabilities (6.3% at F_5, 3.1% at F_6, and 1.6% at F_7), with slight deviations due to random assortment of homologous chromosomes. This study demonstrates the utility of retaining small amounts of heterozygosity within RIL populations for fine mapping. These fine-scale mapping efforts may improve the accuracy of marker-assisted breeding for potentially durable quantitative resistance to an important disease constraint on global common bean production.

Taylor *et al.* (1996b) proposed that quantitative, possibly race-nonspecific, resistance (such as that derived from PI 150414) be combined with race-specific resistance effective against *Psph* races prevalent in particular regions as part of a breeding strategy to establish more durable halo blight resistance. For example, they recommended the combined deployment of quantitative resistance and *Pse-3* in East

and Central Africa to counter *Psph* races 3 and 4, which were found to be largely confined to these regions (Taylor *et al.*, 1996a). Cosegregating genetic markers for quantitative resistance, such as polymorphisms identified within a gene predicted to encode an RNA-binding protein (this work), and for race-specific resistance (Miklas *et al*., 2009; 2011; 2014; Tock *et al.*, manuscript in preparation) provide molecular breeding tools for implementing this strategy. Marker-linked alleles from PI 150414 delineating the Pv04 mapping interval will assist breeding for *Psph* race 6 resistance in common bean lines of Mesoamerican origin.

Breeding targets linked in repulsion

Major *R* genes and QTL governing resistance to the pathogens causing anthracnose (*Co-3*, *Co-9*, *Co-y*, *Co-z*, $ANT_{BJ,DG}$), halo blight (*Pse-1*, *Pse-6*, *Pse-Race 1*, *Pse-Race 7*), rust (*Ur-5*, *Ur-Dorado*), *Bean golden yellow mosaic virus* ($BGYMV_{DX,DG}$) and ashy stem blight (ASB_{DX}) have been mapped to the same region at the telomeric end of the short arm of chromosome Pv04 (Fourie *et al*., 2004; Miklas *et al*., 2006, and references therein). Accordingly, efforts to introgress new, potentially recessive, nearby targets will need to avoid displacement of *R* genes that are in repulsion phase linkage with these breeding targets (Miklas *et al*., 2014). The potentially recessive nature of resistance derived from PI 150414 presents obstacles to approaches using conventional breeding alone. New breeding techniques utilizing genome editing (e.g., via the CRISPR–Cas9 system; Jinek *et al.*, 2012) could enable targeted disruption of hypothesized dominant susceptibility alleles within the Pv04 mapping interval. Combined with marker-assisted breeding, this could enable more expedient deployment of this resistance in lines adapted to local conditions and combining marker-linked major *R* genes governing resistance to halo blight and other regionally important diseases.

Recent efforts by others to map *Psph* race 6 resistance derived from potentially novel sources have identified the same region of Pv04 to which quantitative resistance from PI 150414 was mapped in the current research (Ghising *et al*., 2016), as summarised in the General introduction. Ghising *et al*. (2016) inoculated the second trifoliolate leaf using a multiple-needle method, rather than primary leaves using the spray method that corresponds to the five-point scoring system of Innes *et al*. (1984). Lines with disease severity scores of ≥ 2.0 on this scale were classified as susceptible. The reportedly high levels of resistance

exhibited by ten accessions (with a mean score of ≤ 1.0) in the USDA common bean core collection might suggest that sources of qualitative, potentially race-specific, resistance to race 6 were identified. If so, these might be used as novel halo blight host differentials for the possible subdivision of *Psph* 'race 6', as proposed in Chapter 3. Evaluation of these lines for reaction to other *Psph* races was not reported.

Receptor-like kinases

Evidence to suggest that a mechanism for quantitative resistance to halo blight derived from PI 150414 might be governed by nucleotide-binding site–leucine-rich repeat (NLR) proteins involved in effector-triggered immunity (ETI) was presented in Chapter 2 (Table 2.2). However, a transmembrane basal disease resistance mechanism enabling recognition of extracellular pathogen proteins, as defined by Jones and Dangl (2006), cannot be ruled out. Probabilistic functional gene network prediction (Figure 2.9) nonetheless suggests that any possible role in this mechanism for proteins with functions in extracellular pathogen perception (pattern recognition receptors, PRRs) is likely to occur within a wider signalling pathway that mediates the coordinated activation of defence-related proteins, potentially including NLRs.

Within the 500-kb mapping interval defined for quantitative resistance to halo blight on chromosome Pv04 of the common bean reference genome (Schmutz *et al*., 2014), a predicted cysteine-rich receptor-like protein kinase (CRK; Phvul.004G011000) and two plant self-incompatibility proteins (Phvul.004G010200 and Phvul.004G010300) also have candidate protein potential (Table 2.2). Cysteine-rich proteins have diverse developmental and defence-related functions in plants, and include defensins that confer broad-spectrum antimicrobial activity against plant pathogens (Templeton *et al*., 1994; Silverstein *et al*., 2005; Marshall *et al*., 2011). A number of functionally characterised defensins in common bean exhibit activity against plant and animal fungi, bacteria and viruses (Wong *et al*., 2006; Wu *et al*., 2011; de O Mello *et al*., 2014). Similarly, pathogen-responsive defensin-like (*DEFL*) genes have been identified in *Medicago truncatula*, although none with homology to Phvul.004G011000 (Tesfaye *et al*., 2013). This common bean CRK shares 45%, 44%, 34% and 34% amino acid identity with *Arabidopsis thaliana* proteins CRK4, CRK5, CRK19 and CRK20, respectively. The genes encoding these four CRKs are induced by infection with *P. syringae* pv. *tomato* DC3000 (Chen *et*

al., 2003, 2004). Additionally, induced expression of these four genes triggers a hypersensitive cell death response in transgenic plants (Chen *et al*., 2003, 2004). CRKs also have developmental functions in plant self-incompatibility in the Brassicaceae through receptor–ligand-based recognition and rejection of 'self'-pollen. In species that are not affected by self-incompatibility, these S-domain receptor-like kinases are induced by pathogen recognition, and may have defence functions similar to those of transmembrane PRRs (Sanabria *et al*., 2008; Boller and Felix, 2009).

The CRK in the Pv04 mapping interval shares 40% amino acid identity with PBS1 (AVRPPHB SUSCEPTIBLE 1; AT5G13160), a serine/threonine kinase in *A. thaliana*. The *A. thaliana* NLR protein RPS5 is activated by cleavage of PBS1 by the *Pseudomonas syringae* pv. *phaseolicola* (*Psph*) cysteine protease type III effector AvrPphB (syn. HopAR1; Shao *et al*., 2003). The CRK may similarly be required for activation of NLR-mediated ETI. Interestingly, *Pse-3* in common bean confers race-specific resistance to *Psph* isolates possessing *avrPphB* (races 3 and 4).

The candidate CRK on Pv04 also shares 43% amino acid identity with an LRR transmembrane protein kinase in *A. thaliana* (AT1G56145). The quantitative nature of the halo blight resistance locus on Pv04 might therefore have a basis in an innate immune response mediated by PRR(s) that governs broad-based resistance to more than one *Psph* race. Susceptibility haplotypes at this locus may represent PRR-encoding allele(s) that have been subverted by the pathogen. In the absence of qualitative or other quantitative resistance to the broadly virulent *Psph* race 6, susceptible genotypes may have allowed the invading pathogen to inject virulence effectors into the cell cytoplasm and to proliferate. The resistance from PI 150414 has been deployed extensively in common bean breeding programmes for at least two decades without the emergence of any reported resistance-breaking strains (J. D. Taylor, personal communication; Silbernagel and Hannan, 1992). The potentially durable and race-nonspecific nature of this resistance might therefore have a basis in a PRR-mediated, rather than NLR-mediated, mechanism.

Testing for enhanced susceptibility in *Arabidopsis thaliana*

Assessment of non-host interaction phenotypes of *A. thaliana* T-DNA insertion lines mutated at selected potential orthologues may provide more evidence for a candidate defence response pathway. The alternative hypothesis that motivates further

experimental investigation is that disruption of a non-host, species-level resistance mechanism in *A. thaliana* (potentially orthologous to a broad-spectrum halo blight resistance mechanism in *P. vulgaris*) will cause enhanced susceptibility following inoculation with *Psph* (Forsyth *et al*., 2010).

Pathovar- and pathotype-level candidate virulence determinants

Two computational approaches to GWA genetics were exploited for the identification of candidate molecular targets in *Psph* for hypothesised race-nonspecific and race-specific host resistance to halo blight. One approach, based on TBLASTN searches of genomic contigs for sequences with homology to previously characterised or predicted type III secretion system (T3SS) proteins (Bart *et al*., 2012), enabled detection of presence/absence and allelic sequence variation across a pathovar and a species (Tables 3.2 and A5.1). Manual inspection of multiple sequence alignments permitted convenient visualisation of the variable conservation within *P. syringae* of predicted type III effectors (T3Es) and hypersensitive response and pathogenicity (Hrp) proteins required for the synthesis of the T3SS apparatus. This provided the basis for the provisional definition of protein groups that serve at different diagnostic levels as potential determinants of host specialisation (Table 3.2). Candidate pathovar-level (race-nonspecific) and pathotype-level (race-specific) determinants of pathogenic capability were identified using this approach.

A statistical phylogenomic approach to microbial GWA genetics provided a robust method for the identification of five high-probability candidate determinants of the broad virulence of *Psph* race 6, with pan-genome single-nucleotide resolution (Table 3.3 and Figure 3.1). SNPs were identified by PPFS2 (Hall, 2014a,b) as potentially causally related to the race 6 virulence phenotype based on low χ^2 probabilities their allelic state changed randomly along the same phylogenomic branches along which the phenotype changed.

The most intriguing candidate identified is AvrD (Table 3.4), a protein in *P. syringae* that directs the production of syringolide elicitors that, in turn, trigger a hypersensitive resistant reaction in soybean plants encoding receptor protein *Rpg4*. HopX1 is another high-probability candidate potentially causal of race 6 virulence. However, a diagnostic premature stop codon predicted to be encoded by eight of the 12 isolates originally designated as race 6 may render these alleles non-functional in encoding the type III (a)virulence effector HopX1 (Table 3.5). Cleaved amplified

polymorphic sequence (CAPS) markers targeting diagnostic substitutions within alleles *avrD*^{p6} and *hopX1*^{p6} were designed to enable pathotype predictions for newly collected *Psph* isolates (Table 3.6). If validated, these molecular diagnostic tools will permit rapid detection of a broadly virulent pathogen, *Psph* race 6, which could be useful for identifying and deploying novel and previously identified sources of race-specific and race-nonspecific resistance to halo blight.

The putative presence of HopF1 (corresponding to resistance conferred by *Pse-1* in common bean) uniquely in isolates from *Psph* races 1, 5, 7 and 9, and HopAR1 (corresponding to *Pse-3*) uniquely in those from races 3 and 4 (Table 3.2) is consistent with previous genetic predictions (Jenner *et al.*, 1991; Tsiamis *et al.*, 2000). This demonstrates the utility of the GWA genetics approach adopted in this research for identifying determinants of pathogen avirulence. It could therefore provide a useful tool with which to investigate the effector complements of other plant pathogens, with a view to developing host resistance strategies that are appropriate for controlling regionally important pathogen races (Bart *et al*., 2012).

Candidate targets for syringolide elicitors

A mapping interval with physical coordinates on the soybean genome has not yet been defined for *Rpg4* (Farhatullah *et al.*, 2010), although findings from biochemical assays suggest that thiol-protease-like P34 is the syringolide receptor that mediates *avrD–Rpg4* complementarity (Ji *et al.*, 1997, 1998). P42-2, which is 97% identical to a soybean nicotinamide adenine dinucleotide phosphate (NADP)-dependent hydroxypyruvate reductase, was subsequently identified as a potential messenger that may be involved in defence signalling following P34–syringolide binding (Okinaka *et al.*, 2002). Selote *et al*. (2014) showed that silencing orthologues of genes encoding Arabidopsis NON RACE-SPECIFIC DISEASE RESISTANCE 1 (NDR1) in soybean (*GmNDR1a* and *b*) compromises *Rpg4*-mediated resistance to *P. syringae* pv. *glycinea* expressing *avrD*. Other candidate genes have been implicated in the *Rpg4* defence response pathway due to their induction by syringolide elicitors in expression profiling experiments. These include sequences with homology to a putative defence-associated acid phosphatase (GenBank accession AY055218) and an NADP-dependent malic enzyme (X80051) in *Phaseolus vulgaris* (Hagihara *et al*., 2004).

Intriguingly, Slaymaker and Keen (2004) reported that syringolide treatment of soybean cells expressing *Rpg4* induced phosphorylation (a known plant defence signalling mechanism) of a glycine-rich RNA-binding protein (GmGRP, sharing 69.8% amino acid identity with Arabidopsis GRP7 targeted by *P. syringae* T3E HopUI) and an EF-hand calcium-binding protein (GmEFH). These were thus identified as candidate plant substrates for syringolide elicitors. However, none of the above candidates shares notable amino acid or nucleotide identity with any of the candidates located within the mapping interval defined for quantitative halo blight resistance on common bean chromosome Pv04.

Breeding beans for British growing conditions

Selection for traits that are genetically linked and/or phenotypically correlated with other desirable characteristics (e.g., taproot and basal root traits associated with improved water and phosphorus acquisition; Miguel *et al*., 2013, 2015) could provide an expedient breeding strategy for the accumulation of beneficial alleles in locally adapted genotypes. For example, a QTL detected on chromosome Pv07 for taproot diameter exhibited the greatest heritability amongst the putative QTL reported for root architecture traits, accounting for almost 27% of phenotypic variation observed in the SOA-BN × Edmund recombinant inbred population (Table 4.4). Colocalisation of QTL for other root traits on Pv07 (Table 4.5), and strong phenotypic correlations between these traits (Table 4.6), present an opportunity for marker-assisted accumulation of alleles with possible functions in physiological resilience. Coupling phase linkage of potentially desirable alleles with seed coat pigmentation factor *P* on Pv07 (Table 4.4) may require the identification of recombinants for the development of backcross populations within which desired market class and resilience characteristics are accumulated.

Provisional QTL analyses using RIL phenotypic data collected in pilot field and polytunnel experiments demonstrate the requirement for field evaluation of agronomic traits important for adaptation to UK growing conditions. However, the major-effect QTL detected on Pv01 for growth habit under both field and polytunnel conditions, which likely corresponds to the *fin* locus for determinacy, may prove useful for discarding material predicted to have an indeterminate, climbing growth habit that is unsuitable for field production (Figures 4.8 and 4.9, and Table 4.7). Colocalisation of major-effect QTL on Pv05 for reproductive developmental stage

using data from field evaluations at 82 DAS and 105 DAS (Figure 4.9 and Table 4.7) indicates that polymorphisms in this chromosome region may provide valuable markers for selection for earliness of maturity. These QTL were highly significant in the context of both one-dimensional genome scans (for single QTL) and multi-QTL models, with potentially useful levels of heritability (Figure 4.9 and Table 4.7). Although this locus was not significant in the search for single QTL using data from polytunnel evaluation at 100 DAS, a higher-than-average LOD score was detected on Pv05 (Figure 4.8).

A serendipitous encounter

My entry into crop science was unconventional. A background in human rights law and the economic and policy dimensions of food security motivated me to undertake postgraduate study related to food production at the University of Warwick. My aim was to develop scientific skills that would equip me to make a practical contribution to improving smallholder farming. Serendipitous timing and generous mentoring from a likeminded professor enabled me to continue this pursuit with PhD training in the genetic improvement of a globally important staple food crop. Efforts to map disease resistance and abiotic stress tolerance loci in common bean have recently intensified thanks to timely advances in DNA sequencing and computational biology. These developments make this a thrilling period within which to learn and apply plant genetics for crop improvement.

References

Afanador, L. K., Haley, S. D. and Kelly, J. D. (1993). Adoption of a "mini-prep" DNA extraction method for RAPD marker analysis in common bean (*Phaseolus vulgaris* L.). *Annu. Rep. Bean Improv. Coop.* **35**, 10–11.

Afzal, A. J., Srour, A., Saini, N., Hemmati, N., El Shemy, H. A. and Lightfoot, D. A. (2012). Recombination suppression at the dominant *Rhg1/Rfs2* locus underlying soybean resistance to the cyst nematode. *Theor. Appl. Genet.* **124**, 1027–1039.

Ahmad, S., Van Hulten, M., Martin, J., Pieterse, C. M. J., Van Wees, S. C. M. and Ton, J. (2011). Genetic dissection of basal defence responsiveness in accessions of *Arabidopsis thaliana*. *Plant Cell Environ.* **34**, 1191–1206.

Alfano, J. R., Charkowski, A. O., Deng, W. L., Badel, J. L., Petnicki-Ocwieja, T., van Dijk, K. and Collmer, A. (2000). The *Pseudomonas syringae* Hrp pathogenicity island has a tripartite mosaic structure composed of a cluster of type III secretion genes bounded by exchangeable effector and conserved effector loci that contribute to parasitic fitness and pathogenicity in plants. *Proc. Natl. Acad. Sci. USA* **97**, 4856–4861.

Algorithmic Biology Lab (undated). SPAdes 3.0 on GAGE-B data sets. Algorithmic Biology Lab, St. Petersburg Academic University of the Russian Academy of Sciences, St. Petersburg, Russia. http://bioinf.spbau.ru/en/content/spades-30-gage-b-data-sets (accessed 5 Aug. 2016).

Alikhan, N. F., Petty, N. K., Ben Zakour, N. L. and Beatson, S. A. (2011). BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* **12**, 402.

Andersson, M. X., Kourtchenko, O., Dangl, J. L., Mackey, D. and Ellerström, M. (2006). Phospholipase-dependent signalling during the AvrRpm1- and AvrRpt2-induced disease resistance responses in *Arabidopsis thaliana*. *Plant J.* **47**, 947–959.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 18 Jul. 2016).

Ariel, F., Brault-Hernandez, M., Laffont, C., Huault, E., Brault, M., Plet, J., Moison, M., Blanchet, S., Ichanté, J. L., Chabaud, M., Carrere, S., Crespi, M., Chan, R. L. and Frugier, F. (2012). Two direct targets of cytokinin signaling regulate symbiotic nodulation in *Medicago truncatula*. *Plant Cell* **24**, 3838–3852.

Ariyarathne, H. M., Coyne, D. P., Jung, G., Skroch, P. W., Vidaver, A. K., Steadman, J. R., Miklas, P. N. and Bassett, M. J. (1999). Molecular

mapping of disease resistance genes for halo blight, common bacterial blight, and bean common mosaic virus in a segregating population of common bean. *J. Am. Soc. Hortic. Sci.* **124**, 654–662.

Arnold, D. L., Godfrey, S. A. C. and Jackson, R. W. (2009). *Pseudomonas syringae* genomics provides important insights to secretion systems, effector genes and the evolution of virulence. In *Plant Pathogenic Bacteria: Genomics and Molecular Biology*, pp. 203–226. Edited by R. W. Jackson. Norfolk, UK: Caister Academic Press.

Arnold, D. L., Lovell, H. C., Jackson, R. W. and Mansfield, J. W. (2011). *Pseudomonas syringae* pv. *phaseolicola*: from 'has bean' to supermodel. *Mol. Plant Pathol.* **12**, 617–627.

Asensio-S.-Manzanera, M. C., Asensio, C. and Singh, S. P. (2006). Gamete selection for resistance to common and halo bacterial blights in dry bean intergene pool populations. *Crop Sci.* **46**, 131–135.

Ashfield, T., Bocian, A., Held, D., Henk, A. D., Marek, L. F., Danesh, D., Peñuela, S., Meksem, K., Lightfoot, D. A., Young, N. D., Shoemaker, R. C. and Innes, R. W. (2003). Genetic and physical localization of the soybean *Rpg1-b* disease resistance gene reveals a complex locus containing several tightly linked families of NBS–LRR genes. *Mol. Plant–Microbe Interact.* **16**, 817–826.

Austin, R. B., Hardwick, R. C. and MacLean, M. S. M (1970). Plant studies. *Rep. Natn. Veg. Res. Stn. for 1969*, 64.

Austin, R. B. and MacLean, M. S. M. (1972a). Some effects of temperature on the rates of photosynthesis and respiration of *Phaseolus vulgaris* L. *Photosynthetica* **6**, 41–50.

Austin, R. B. and MacLean, M. S. M. (1972b). A method for screening *Phaseolus* genotypes for tolerance to low temperatures. *J. Hort. Sci.* **47**, 279–290.

Axtell, M. J., Chisholm, S. T., Dahlbeck, D. and Staskawicz, B. J. (2003). Genetic and molecular evidence that the *Pseudomonas syringae* type III effector protein AvrRpt2 is a cysteine protease. *Mol. Microbiol.* **49**, 1537–1546.

Baltrus, D. A., Nishimura, M. T., Dougherty, K. M., Biswas, S., Mukhtar, M. S., Vicente, J., Holub, E. B. and Dangl, J. L. (2012). The molecular basis of host specialization in bean pathovars of *Pseudomonas syringae*. *Mol. Plant–Microbe Interact.* **25**, 877–888.

Bankevich, A., Nurk, S., Antipov, D., Gurevich, A. A., Dvorkin, M., Kulikov, A. S., Lesin, V. M., Nikolenko, S. I., Pham, S., Prjibelski, A. D., Pyshkin, A. V., Sirotkin, A. V., Vyahhi, N., Tesler, G., Alekseyev, M. A. and Pevzner, P. A. (2012). SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J. Comp. Biol.* **19**, 455–477.

Barrera-Figueroa, B. E., Peña-Castro, J. M., Acosta-Gallegos, J. A., Ruiz-Medrano, R. and Xoconostle-Cázares, B. (2007). Isolation of dehydration-responsive genes in a drought tolerant common bean cultivar and expression of a group 3 late embryogenesis abundant mRNA in tolerant and susceptible bean cultivars. *Funct. Plant Biol.* **34**, 368–381.

Bart, R., Cohn, M., Kassen, A., McCallum, E. J., Shybut, M., Petriello, A., Krasileva, K., Dahlbeck, D., Medina, C., Alicai, T., Kumar, L., Moreira, L. M., Neto, J. R., Verdier, V., Santana, M. A., Kositcharoenkul, N., Vanderschuren, H., Gruissem, W., Bernal, A. and Staskawicz, B. J. (2012). High-throughput genomic sequencing of cassava bacterial blight strains identifies conserved effectors to target for durable resistance. *Proc. Natl. Acad. Sci. USA* **109**, 1972–1979.

Becerra-Velásquez, V. L. and Gepts, P. (1994). RFLP diversity of common bean (*Phaseolus vulgaris*) in its centres of origin. *Genome* **37**, 256–263.

Beebe, S. E., Rao, I. M., Blair, M. W. and Acosta, J. A. (2013). Phenotyping common beans for adaptation to drought. *Front. Physiol.* **4**, 35.

Beebe, S., Rao, I., Blair, M. W. and Butare, L. (2009). Breeding for abiotic stress tolerance in common bean: present and future challenges. In *Australasian Plant Breeding: SABRAO Conference*. Cairns, Queensland, Australia. https://www.researchgate.net/publication/228612601_Breeding_for_abiotic_stress_tolerance_in_common_bean_Present_and_future_challenges (accessed 23 Sept. 2016).

Beebe, S. E., Rao, I. M., Cajiao, I. and Grajales, M. (2008). Selection for drought resistance in common bean also improves yield in phosphorus limited and favorable environments. *Crop Sci.* **48**, 582–592.

Beebe, S., Rengifo, J., Gaitan, E., Duque, M. C. and Tohme, J. (2001). Diversity and origin of Andean landraces of common bean. *Crop Sci.* **41**, 854–862.

Beebe, S., Skroch, P. W., Tohme, J., Duque, M. C., Pedraza, F. and Nienhuis, J. (2000). Structure of genetic diversity among common bean landraces of Middle American origin based on correspondence analysis of RAPD. *Crop Sci.* **40**, 264–273.

Bello, M. H., Moghaddam, S. M., Massoudi, M., McClean, P. E., Cregan, P. B. and Miklas, P. N. (2014). Application of *in silico* bulked segregant analysis for rapid development of markers linked to *Bean common mosaic virus* resistance in common bean. *BMC Genomics* **15**, 903.

Berardini, T. Z., Reiser, L., Li, D., Mezheritsky, Y., Muller, R., Strait, E. and Huala, E. (2015). The Arabidopsis Information Resource: making and mining the "gold standard" annotated reference plant genome. *Genesis* **53**: 474–485.

Berr, A., McCallum, E. J., Alioua, A., Heintz, D., Heitz, T. and Shen, W.-H. (2010). Arabidopsis histone methyltransferase SET DOMAIN GROUP8 mediates induction of the jasmonate/ethylene pathway genes in plant defense response to necrotrophic fungi. *Plant Physiol.* **154**, 1403–1414.

Biffen, R. H. (1907). Studies in the inheritance of disease resistance. *J. Agric. Sci.* **2**, 109–128.

Biffen, R. H. (1912). Studies in the inheritance of disease resistance. II. *J. Agric. Sci.* **4**, 421–429.

Bijtenhoorn, P., Mayerhofer, H., Müller-Dieckmann, J., Utpatel, C., Schipper, C., Hornung, C., Szesny, M., Grond, S., Thürmer, A., Brzuszkiewicz, E., Daniel, R., Dierking, K., Schulenburg, H. and Streit, W. R. (2011). A novel metagenomic short-chain dehydrogenase/reductase attenuates *Pseudomonas aeruginosa* biofilm formation and virulence on *Caenorhabditis elegans*. *PLoS ONE* **6**, E26278.

Blair, M. W., Cortés, A. J., Penmetsa, R. V., Farmer, A., Carrasquilla-Garcia, N. and Cook, D. R. (2013). A high-throughput SNP marker system for parental polymorphism screening, and diversity analysis in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **126**, 535–548.

Blair, M. W., Galeano, C. H., Tovar, E., Munoz Torres, M. C., Castrillon, A. V., Beebe, S. E. and Rao, I. M. (2012). Development of a Mesoamerican intra-genepool genetic map for quantitative trait loci detection in a drought tolerant × susceptible common bean (*Phaseolus vulgaris* L.) cross. *Mol. Breed.* **29**, 71–88.

Blair, M. W., Giraldo, M. C., Buendia, H. F., Tovar, E., Duque, M. C. and Beebe, S. E. (2006). Microsatellite marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **113**, 100–109.

Blair, M. W., Hurtado, N., Chavarro, C., Munoz-Torres, M., Giraldo, M., Pedraza, F., Tomkins, J. and Wing, R. (2011). Gene-based SSR markers for common bean (*Phaseolus vulgaris* L.) derived from root and leaf tissue ESTs: an integration of the BMc series. *BMC Plant Biol.* **11**, 50.

Blair, M. W., Pedraza, F., Buendia, H. F., Gaitan-Solis, E., Beebe, S. E., Gepts, P. and Tohme, J. (2003). Development of a genome-wide anchored microsatellite map for common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **107**, 1362–1374.

Bolger, A. M., Lohse, M. and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120.

Boller, T. and Felix, G. (2009). A renaissance of elicitors: perception of microbe-associated molecular patterns and danger signals by pattern-recognition receptors. *Annu. Rev. Plant Biol.* **60**, 379–406.

Bolser, D. (2016). SNP calling pipeline. Figshare. doi: https://dx.doi.org/10.6084/m9.figshare.3847275.v1 (accessed 22 Sept. 2016).

Borhan, M. H., Gunn, N., Cooper, A., Gulden, S., Tör, M., Rimmer, S. R. and Holub, E. B. (2008). *WRR4* encodes a TIR–NB–LRR protein that confers broad-spectrum white rust resistance in *Arabidopsis thaliana* to four physiological races of *Albugo candida*. *Mol. Plant–Microbe Interact.* **21**, 757–768.

Borhan, M. H., Holub, E. B., Kindrachuk, C., Omidi, M., Bozorgmanesh-Frad, G. and Rimmer, S. R. (2010). *WRR4*, a broad-spectrum TIR–NB–LRR gene from *Arabidopsis thaliana* that confers white rust resistance in transgenic oilseed brassica crops. *Mol. Plant Pathol.* **11**, 283–291.

Boyd, L. A., Ridout, C., O'Sullivan, D. M., Leach, J. E. and Leung, H. (2013). Plant–pathogen interactions: disease resistance in modern agriculture. *Trends Genet.* **29**, 233–240.

Boyer, J. S. (1978). Water deficits and photosynthesis. In *Water Deficits and Plant Growth IV*, pp. 154–191. Edited by T. T. Kozloviski. New York: Academic Press.

Bradbury, P. J., Zhang, Z., Kroon, D. E., Casstevens, T. M., Ramdoss, Y. and Buckler, E. S. (2007). TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics* **23**, 2633–2635.

Bressani, R. (1983). Research needs to up-grade the nutritional quality of common beans (*Phaseolus vulgaris*). *Plant Food. Hum. Nutr.* **32**, 101–110.

Brisson, N., Gate, P., Gouache, D., Charmet, G., Oury, F.-X. and Huard, F. (2010). Why are wheat yields stagnating in Europe? A comprehensive data analysis for France. *Field Crops Res.* **119**, 201–212.

Broman, K. W. and Sen, Ś. (2009). *A Guide to QTL Mapping with R/qtl*. London: Springer.

Broman, K. W., Wu, H., Sen, Ś. and Churchill, G. A. (2003). R/qtl: QTL mapping in experimental crosses. *Bioinformatics* **19**, 889–890.

Broughton, W. J., Hernández, G., Blair, M., Beebe, S., Gepts, P. and Vanderleyden, J. (2003). Beans (*Phaseolus* spp.) – model food legumes. *Plant Soil* **252**, 55–128.

Burridge, J. (2012). Shovelomics: high-throughput phenotyping of root system architecture [video]. VCell Productions. Pennsylvania State University and United States Department of Agriculture, National Institute of Food and Agriculture. Published 14 Jun. 2012: http://plantscience.psu.edu/research/labs/roots/methods/field/shovelomics (accessed 15 Sept. 2016).

Burridge, J., Jochua, C. N., Bucksch, A. and Lynch, J. P. (2016). Legume shovelomics: high-throughput phenotyping of common bean (*Phaseolus vulgaris* L.) and cowpea (*Vigna unguiculata* subsp. *unguiculata*) root architecture in the field. *Field Crops Res.* **192**, 21–32.

Burt, A. J., William, H. M., Perry, G., Khanal, R., Pauls, K. P., Kelly, J. D. and Navabi, A. (2015). Candidate gene identification with SNP marker-based fine mapping of anthracnose resistance gene *Co-4* in common bean. *PLoS ONE* **10**, E0139450.

CABI (2016a). *Pseudomonas savastanoi* pv. *phaseolicola* (halo blight (of beans)). In *Crop Protection Compendium*. Wallingford, UK: CAB International. www.cabi.org/cpc (accessed 26 Sept. 2016).

CABI (2016b). Plantwise Technical Factsheet: Halo blight (of beans) (*Pseudomonas savastanoi* pv. *phaseolicola*). In *Plantwise Knowledge Bank*. Wallingford, UK: CAB International. http://www.plantwise.org/knowledgebank/datasheet.aspx?dsid=44987 (accessed 29 Jul. 2016).

Castaño, N. H. (2013). Identification of candidate genes for basal root whorl number and basal root number in common bean (*Phaseolus vulgaris* L.). Master's thesis. Universidad Nacional de Colombia, Palmira, Colombia.

Catanzariti, A.-M., Dodds, P. N., Ve, T., Kobe, B., Ellis, J. G. and Staskawicz, B. J. (2010). The AvrM effector from flax rust has a structured C-terminal domain and interacts directly with the M resistance protein. *Mol. Plant–Microbe Interact.* **23**, 49–57.

Cesari, S., Thilliez, G., Ribot, C., Chalvon, V., Michel, C., Jauneau, A., Rivas, S., Alaux, L., Kanzaki, H., Okuyama, Y., Morel, J. B., Fournier, E., Tharreau, D., Terauchi, R. and Kroj, T. (2013). The rice resistance protein pair RGA4/RGA5 recognizes the *Magnaporthe oryzae* effectors AVR-Pia and AVR1-CO39 by direct binding. *Plant Cell* **25**, 1463–1481.

Chang, J. H., Urbach, J. M., Law, T. F., Arnold, L. W., Hu, A., Gombar, S., Grant, S. R., Ausubel, F. M. and Dangl, J. L. (2005). A high-throughput, near-saturating screen for type III effector genes from *Pseudomonas syringae*. *Proc. Natl. Acad. Sci. USA* **102**, 2549–2554.

Chataika, B. Y. E., Bokosi, J. M., Chirwa, R. M. and Kwapata, M. B. (2011). Inheritance of halo blight resistance in common bean. *Afr. Crop Sci. J.* **19**, 325–333.

Chataika, B. Y. E., Bokosi, J. M., Kwapata, M. B., Chirwa, R. M., Mwale, V. M., Mnyenyembe, P. and Myers, J. R. (2010). Performance of parental genotypes and inheritance of Angular Leaf Spot (*Phaeosariopsis griseola*) resistance in the common bean (*Phaseolus vulgaris*). *Afr. J. Biotechnol.* **9**, 4398–4406.

Chen, K., Du, L. and Chen, Z. (2003). Sensitization of defense responses and activation of programmed cell death by a pathogen-induced receptor-like protein kinase in *Arabidopsis*. *Plant Mol. Biol.* **53**, 61–74.

Chen. K., Fan, B., Du, L. and Chen, Z. (2004). Activation of hypersensitive cell death by pathogen-induced receptor-like protein kinases from *Arabidopsis*. *Plant Mol. Biol.* **56**, 271–283.

Chen, L., Zheng, D., Liu, B., Yang, J. and Jin, Q. (2016). VFDB 2016: hierarchical and refined dataset for big data analysis—10 years on. *Nucleic Acids Res.* **44**, D694–D697.

Chen, N. W. G., Sévignac, M., Thareau, V., Magdelenat, G., David, P., Ashfield, T., Innes, R. W. and Geffroy, V. (2010). Specific resistances against *Pseudomonas syringae* effectors AvrB and AvrRpm1 have evolved differently in common bean (*Phaseolus vulgaris*), soybean (*Glycine max*), and *Arabidopsis thaliana*. *New Phytol.* **187**, 941–956.

Chen, W., Mingus, J., Mammadov, J., Backlund, J. E., Greene, T., Thompson, S. and Kumpatla, S. (2010). KASPar: a simple and cost-effective system for SNP genotyping. Abstract. Plant & Animal Genome Conference XVIII. San Diego, USA.

Chen, Y., Liu, Z. and Halterman, D. A. (2012). Molecular determinants of resistance activation and suppression by *Phytophthora infestans* effector IPI-O. *PLoS Pathog.* **8**, E1002595.

Choi, K., Reinhard, C., Serra, H., Ziolkowski, P. A., Underwood, C. J., Zhao, X., Hardcastle, T. J., Yelina, N. E., Griffin, C., Jackson, M., Mézard, C., McVean, G., Copenhaver, G. P. and Henderson, I. R. (2016). Recombination rate heterogeneity within Arabidopsis disease resistance genes. *PLoS Genet.* **12**, E1006179.

Christie, P. J. (1997). *Agrobacterium tumefaciens* T-complex transport apparatus: a paradigm for a new family of multifunctional transporters in eubacteria. *J. Bacteriol.* **179**, 3085–3094.

Cichy, K. A., Blair, M. W., Mendoza, C. H. G., Snapp, S. S. and Kelly, J. D. (2009). QTL analysis of root architecture traits and low phosphorus tolerance in an Andean bean population. *Crop Sci.* **49**, 59–68.

Cichy, K. A., Porch, T., Beaver, J. S., Cregan, P. B., Fourie, D., Glahn, R. P., Grusak, M. A., Kamfwa, K., Katuuramu, D. N., McClean, P., Mndolwa, E., Nchimbi-Msolla, S., Pastor-Corrales, M. A. and Miklas, P. N. (2015). A *Phaseolus vulgaris* diversity panel for Andean bean improvement. *Crop Sci.* **55**, 2149–2160.

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., Land, S.J., Lu, X. and Ruden, D. M. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in

the genome of *Drosophila melanogaster* strain w^{1118}; *iso*-2; *iso*-3. *Fly* **6**, 80–92.

Collier, R. (2014). Introduction of novel crops in the UK: a workshop. Crop-Innovations and Warwick Crop Centre, Wellesbourne, UK.

Collmer, C. W., Marston, M. F., Taylor, J. C. and Jahn, M. (2000). The *I* gene of bean: a dosage-dependent allele conferring extreme resistance, hypersensitive resistance, or spreading vascular necrosis in response to the potyvirus *Bean common mosaic virus*. *Mol. Plant–Microbe Interact.* **13**, 1266-1270.

Comai, L., Young, K., Till, B. J., Reynolds, S. H., Greene, E. A., Codomo, C. A., Enns, L. C., Johnson, J. E., Burtner, C., Odden, A. R. and Henikoff, S. (2004). Efficient discovery of DNA polymorphisms in natural populations by Ecotilling. *Plant J.* **37**, 778–786.

Conway, J., Hardwick, R. C., Innes, N. L., Taylor, J. D. and Walkey, D. G. A. (1982). White-seeded beans (*Phaseolus vulgaris*) resistant to halo blight (*Pseudomonas phaseolicola*), to bean common mosaic virus, and to anthracnose (*Colletotrichum lindemuthianum*). *J. Agric. Sci.* **99**, 555–560.

Coyne, D. P., Schuster, M. L. and Gallegos, C. C. (1971). Inheritance and linkage of the halo blight systemic chlorosis and leaf water soaked reaction in *Phaseolus vulgaris* variety crosses. *Plant Dis. Report.* **55**, 203–207.

Cortés, A. J., Chavarro, M. C. and Blair, M. W. (2011). SNP marker diversity in common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **123**, 827–845.

Cortés, A. J., Chavarro, M. C., Madrinan, S., This, D. and Blair, M. W. (2012b). Molecular ecology and selection in the drought-related *Asr* gene polymorphisms in wild and cultivated common bean (*Phaseolus vulgaris* L.). *BMC Genet.* **13**, 58.

Cortés, A. J., This, D., Chavarro, C., Madrinan, S. and Blair, M. W. (2012a). Nucleotide diversity patterns at the drought-related *DREB2* encoding genes in wild and cultivated common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **125**, 1069–1085.

Cortés, A. J., Monserrate, F. A., Ramirez-Villegas, J., Madrinan, S. and Blair, M. W. (2013). Drought tolerance in wild plant populations: the case of common beans (*Phaseolus vulgaris* L.). *PLoS ONE* **8**, E62898.

Cummings, M. P., Handley, S. A., Myers, D. S., Reed, D. L., Rokas, A. and Winka, K. (2003). Comparing bootstrap and posterior probability values in the four-taxon case. *Syst. Biol.* **52**, 477–487.

Cuppen, E. (2007). Genotyping by allele-specific amplification (KASPar). *Cold Spring Harb. Protoc.* **2007**, pdb.prot4841.

Dakora, F. D. and Keya, S. O. (1997). Contribution of legume nitrogen fixation to sustainable agriculture in sub-Saharan Africa. *Soil Biol. Biochem.* **29**, 809–817.

Dangl, J. L., Horvath, D. M. and Staskawicz, B. J. (2013). Pivoting the plant immune system from dissection to deployment. *Science* **341**, 746–751.

Dangl, J. L. and Jones, J. D. G. (2001). Plant pathogens and integrated defence responses to infection. *Nature* **411**, 826–833.

Danielsen, S. and Kelly, P. (2010). A novel approach to quality assessment of plant health clinics. *Int. J. Agr. Sust.* **8**, 257–269.

de O Mello, E., dos Santos, I. S., de O Carvalho, A., de Souza, L. S., de Souza-Filho, G. A., do Nascimento, V. V., Machado, O. L. T., Zottich, U. and Gomes, V. M. (2014). Functional expression and activity of the recombinant antifungal defensin *Pv*D$_1$r from *Phaseolus vulgaris* L. (common bean) seeds. *BMC Biochem.* **15**, 7.

De Ron, A. M., Rodiño, A. P., Santalla, M., González, A. M., Lema, M. J., Martín, I. and Kigel, J. (2016). Seedling emergence and phenotypic response of common bean germplasm to different temperatures under controlled conditions and in open field. *Front. Plant Sci.* **7**, 1087.

de Vienne, D. (Ed.) (2003a). *Molecular Markers in Plant Genetics and Biotechnology.* Plymouth, UK: Science Publishers, Inc.

de Vienne, D., Santoni, S. and Falque, M. (2003b). Principal sources of molecular markers. In *Molecular Markers in Plant Genetics and Biotechnology*, 3–46. Edited by D. de Vienne. Plymouth, UK: Science Publishers, Inc.

Deslandes, L., Olivier, J., Peeters, N., Feng, D. X., Khounlotham, M., Boucher, C., Somssich, I., Genin, S. and Marco, Y. (2003). Physical interaction between RRS1-R, a protein conferring resistance to bacterial wilt, and PopP2, a type III effector targeted to the plant nucleus. *Proc. Natl. Acad. Sci. USA* **100**, 8024–8029.

Dickson, M. H. and Natti, J. J. (1966). Breeding for halo-blight and virus resistance in snap beans. *Farm Res.* **32**, 4–5.

Diévart, A. and Clark, S. E. (2004). LRR-containing receptors regulating plant development and defense. *Development* **131**, 251–261.

Dóczi, R., Brader, G., Pettkó-Szandtner, A., Rajh, I., Djamei, A., Pitzschke, A., Teige, M. and Hirt, H. (2007). The *Arabidopsis* mitogen-activated protein kinase kinase MKK3 is upstream of group C mitogen-activated protein kinases and participates in pathogen signaling. *Plant Cell* **19**, 3266–3279.

Dodd, M. and Taylor, J. D. (1991, unpublished data). Evaluation of navy bean (*Phaseolus vulgaris*) advanced breeding lines for adaptation to UK conditions. [On file with the author.]

Dodds, P. N., Lawrence, G. J., Catanzariti, A. M., Teh, T., Wang, C. I., Ayliffe, M. A., Kobe, B. and Ellis, J. G. (2006). Direct protein interaction underlies gene-for-gene specificity and coevolution of the flax resistance genes and flax rust avirulence genes. *Proc. Natl. Acad. Sci. USA* **103**, 8888–8893.

Dodds, P. N. and Rathjen, J. P. (2010). Plant immunity: towards an integrated view of plant–pathogen interactions. *Nat. Rev. Genet.* **11**, 539–548.

Drijfhout, E., van Oeveren, J. C. and Jansen, R. C. (1991). A non-destructive selection method for faster growth at suboptimal temperature in common bean (*Phaseolus vulgaris*). *Euphytica* **58**, 65–70.

Duncan, R. W., Gilberston, R. L., Lema, M. and Singh, S. P. (2014b). Inheritance of resistance to the widely distributed race 6 of *Pseudomonas syringae* pv. *phaseolicola* in common bean pinto US14HBR6. *Can. J. Plant Sci.* **94**, 923–928.

Duncan, R. W., Lema, M., Gilbertson, R. L. and Singh, S. P. (2014a). Registration of common bean pinto US14HBR6 resistant to race 6 of the halo blight pathogen, *Pseudomonas syringae* pv. *phaseolicola*. *J. Plant Regist.* **8**, 53–56.

Duncan, R. W., Singh, S. P. and Gilbertson, R. L. (2008). Identification and inheritance of a new source of halo blight resistance in common bean. *Annu. Rep. Bean Improv. Coop.* **51**, 48–49.

Duplan, V. and Rivas, S. (2014). E3 ubiquitin-ligases and their target proteins during the regulation of plant innate immunity. *Front. Plant Sci.* **5**, 42.

Edgar, R. C. (2010). Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**, 2460–2461.

Eichinger, V., Nussbaumer, T., Platzer, A., Jehl, M. A., Arnold, R. and Rattei, T. (2016). EffectiveDB—updates and novel features for a better annotation of bacterial secreted proteins and Type III, IV, VI secretion systems. *Nucleic. Acids Res* **44**, D669–D674.

Elshire, R. J., Glaubitz, J. C., Sun, Q., Poland, J. A., Kawamoto, K., Buckler, E. S. and Mitchell, S. E. (2011). A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS ONE* **6**, E19379.

Engledow, A. S., Medrano, E. G., Mahenthiralingam, E., LiPuma, J. J. and Gonzalez, C. F. (2004). Involvement of a plasmid-encoded type IV secretion system in the plant tissue watersoaking phenotype of *Burkholderia cenocepacia*. *J. Bacteriol.* **186**, 6015–6024.

Farhatullah, Groose, R. W., Raziuddin, Akmal, M. and Inayatullah, M. (2010). Mapping gene for bacterial blight (*Rpg4* locus) in soybean. *Pak. J Bot.* **42**, 2145–2149.

Federation of British Plant Pathologists (1973). A guide to the use of terms in plant pathology. *Phytopathological Paper* No. 17.

Félix-Gastélum, R., Maldonado-Mendoza, I. E., Navarrete-Maya, R., Olivas-Peraza, N. G., Brito-Vega, H. and Acosta-Gallegos, J. A. (2016). Identification of *Pseudomonas syringae* pv. *phaseolicola* as the causal agent of halo blight in yellow beans in northern Sinaloa, Mexico. *Phytoparasitica* **online 2016**, doi:10.1007/s12600-016-0530-5.

Flor, H. H. (1942). Inheritance of pathogenicity in *Melampsora lini*. *Phytopathology* **32**, 653–669.

Flor, H. H. (1971). Current status of the gene-for-gene concept. *Annu. Rev. Phytopathol.* **9**, 275–296.

Food and Agriculture Organization of the United Nations (FAO) (c2013). FAOSTAT Statistical Database: Trade. Rome: FAO. http://faostat.fao.org/DesktopModules/Faostat/WATFDetailed2/watf.aspx?PageID=536 (accessed 23 Sept. 2016).

Food and Agriculture Organization of the United Nations (FAO) (c2015). FAOSTAT Statistical Database: Production. Rome: FAO. http://faostat3.fao.org/browse/Q/QC/E (accessed 23 Sept. 2016).

Forsyth, A., Mansfield, J. W., Grabov, N., de Torres, M., Sinapidou, E. and Grant, M. R. (2010). Genetic dissection of basal resistance to *Pseudomonas syringae* pv. *phaseolicola* in accessions of *Arabidopsis*. *Mol. Plant–Microbe Interact.* **12**, 1545–1552.

Fouilloux, G. and Bannerot, H. (1977). RH13, a four disease resistant line. *Annu. Rep. Bean Improv. Coop.* **20**, 59.

Fourie, D., Miklas, P. N. and Ariyarathne, H. M. (2004). Genes conditioning halo blight resistance to races 1, 7, and 9 occur in a tight cluster. *Annu. Rep. Bean Improv. Coop.* **47**, 103–104.

Fraser, R. S. S. (1986). Genes for resistance to plant viruses. *CRC Crit. Rev. Plant Sci.* **3**, 257–294.

Fraser, R. S. S. (1990). The genetics of resistance to plant viruses. *Annu. Rev. Phytopathol.* **28**, 179–200.

Fu, Z. Q., Guo, M., Jeong, B. R., Tian, F., Elthon, T. E., Cerny, R. L., Staiger, D. and Alfano, J. R. (2007). A type III effector ADP-ribosylates RNA-binding proteins and quells plant immunity. *Nature* **447**, 284–288.

Galeano, C. H., Cortés, A. J., Fernandez, A. C., Soler, A., Franco-Herrera, N., Makunde, G., Vanderleyden, J. and Blair, M. W. (2012). Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. *BMC Genet.* **13**, 48.

Galeano, C. H., Fernandez, A. C., Franco-Herrera, N., Cichy, K. A., McClean, P. E., Vanderleyden, J. and Blair, M. W. (2011). Saturation of an intra-gene pool linkage map: towards a unified consensus linkage map for fine mapping and synteny analysis in common bean. *PLoS ONE* **6**, E28135.

Galeano, C. H., Fernandez, A. C., Gomez, M. and Blair, M. W. (2009b). Single strand conformation polymorphism based SNP and Indel markers for genetic mapping and synteny analysis of common bean (*Phaseolus vulgaris* L.). *BMC Genomics* **10**, 629.

Galeano, C. H., Gomez, M., Rodriguez, L. M. and Blair, M. W. (2009a). CEL I nuclease digestion for SNP discovery and marker development in common bean (*Phaseolus vulgaris* L.). *Crop Sci.* **49**, 381–394.

Gardner, S. N. and Hall, B. G. (2013). When whole-genome alignments just won't work: kSNP v2 software for alignment-free SNP discovery and phylogenetics of hundreds of microbial genomes. *PLoS ONE* **8**, E81760.

Gardner, S. N. and Hall, B. G. (2016). kSNP3 User Guide. https://sourceforge.net/projects/ksnp/files/ (accessed 5 Aug. 2016).

Gardner, S. N., Slezak, T. and Hall, B. G. (2015). kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. *Bioinformatics* **31**, 2877–2878.

Gepts, P., Aragão, F. J. L., Barros, E., Blair, M. W., Brondani, R., Broughton, W., Galasso, I., Hernández, G., Kami, J., Lariguet, P., McClean, P., Melotto, M., Miklas, P., Pauls, P., Pedrosa-Harand, A., Porch, T., Sánchez, F., Sparvoli, F. and Yu, K. (2008). Genomics of *Phaseolus* beans, a major source of dietary protein and micronutrients in the tropics. In *Genomics of Tropical Crop Plants*, Vol. 1, 113–143. Edited by P. H. Moore and R. Ming. New York: Springer.

Ghising, K., Pasche, J. S., Lamppa, R., Schroder, S., Vasquez-Guzman, J., Soltani, A., Moghaddam, S. M., Mamidi, S., McPhee, K., McClean, P. E. and Osorno, J. M. (2016). Identifying genomic regions associated with halo blight resistance within the USDA core collection of common bean. *Annu. Rep. Bean Improv. Coop.* **59**, 111–112.

Gilbert, B. M. and Wolpert, T. J. (2013). Characterization of the *LOV1*-mediated, victorin-induced, cell-death response with virus-induced gene silencing. *Mol. Plant–Microbe Interact.* **26**, 903–917.

Gimenez-Ibanez, S., Boter, M., Fernández-Barbero, G., Chini, A., Rathjen, J. P. and Solano, R. (2014). The bacterial effector HopX1 targets JAZ

transcriptional repressors to activate jasmonate signaling and promote infection in *Arabidopsis*. *PLoS Biol.* **12**, E1001792.

Glaubitz, J. C., Casstevens, T. M., Lu, F., Harriman, J., Elshire, R. J., Sun, Q. and Buckler, E. S. (2014). TASSEL-GBS: a high capacity genotyping by sequencing analysis pipeline. *PLoS ONE* **9**, E90346. [See TASSEL 5.0 GBSv2 Discovery/Production Pipeline Wiki for an overview of the plugins used in the current research: https://bitbucket.org/tasseladmin/tassel-5-source/wiki/Tassel5GBSv2Pipeline (accessed 5 Jun. 2016).]

Glazebrook, J. (2005). Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**, 205–227.

Goel, S., Chen, Z., Conner, J. A., Akiyama, Y., Hanna, W. W. and Ozias-Akins, P. (2003). Delineation by fluorescence *in situ* hybridization of a single hemizygous chromosomal region associated with aposporous embryo sac formation in *Pennisetum squamulatum* and *Cenchrus ciliaris*. *Genetics* **163**, 1069–1082.

González, A. J., Landeras, E. and Mendoza, M. C. (2000). Pathovars of *Pseudomonas syringae* causing bacterial brown spot and halo blight in *Phaseolus vulgaris* L. are distinguishable by ribotyping. *Appl. Environ. Microbiol.* **66**, 850–854.

González, A. M., Yuste-Lisbona, F. J., Godoy, L., Fernández-Lozano, A., Rodiño, A. P., De Ron, A. M., Lozano, R. and Santalla, M. (2016). Exploring the quantitative resistance to *Pseudomonas syringae* pv. *phaseolicola* in common bean (*Phaseolus vulgaris* L.). *Mol. Breed.* **36**, 166.

Goodstein, D. M., Shu, S., Howson, R., Neupane, R., Hayes, R. D., Fazo, J., Mitros, T., Dirks, W., Hellsten, U., Putnam, N. and Rokhsar, D. S. (2012). Phytozome: a comparative platform for green plant genomics. *Nucleic Acids Res.* **40**, D1178–D1186.

Graham, M. A., Ramírez, M., Valdés-López, O., Lara, M., Tesfaye, M., Vance, C. P. and Hernandez, G. (2006). Identification of candidate phosphorus stress induced genes in *Phaseolus vulgaris* through clustering analysis across several plant species. *Funct. Plant Biol.* **33**, 789–797.

Grant, M. R., Godiard, L., Straube, E., Ashfield, T., Lewald, J., Sattler, A., Innes, R. W. and Dangl, J. L. (1995). Structure of the *Arabidopsis RPM1* gene enabling dual specificity disease resistance. *Science* **269**, 843–846.

Gurevich, A., Saveliev, V., Vyahhi, N. and Tesler, G. (2013). QUAST: quality assessment tool for genome assemblies. *Bioinformatics* **29**, 1072–1075.

Guy, C. L. (2003). Freezing tolerance of plants: current understanding and selected emerging concepts. *Can. J. Bot.* **81**, 1216–1223.

H. J. Heinz Co. Ltd. (c2016). About Heinz: Trivia. http://www.heinz.co.uk/en/our-

company/about-heinz/trivia (accessed 23 Sept. 2016).

Hagedorn, D. J., Walker, J. C. and Rand, R. E. (1974). Wis. HBR 40 and Wis. HBR 72 bean germplasm. *HortScience* **9**, 402.

Hagihara, T., Hashi, M., Takeuchi, Y. and Yamaoka, N. (2004). Cloning of soybean genes induced during hypersensitive cell death caused by syringolide elicitor. *Planta* **218**, 606–614.

Hall, B. G. (2014a). SNP-associations and phenotype predictions from hundreds of microbial genomes without genome alignments. *PLoS ONE* **9**, E90490.

Hall, B. G (2014b). PPFS2: The Predicting Phenotypes from SNPs Package User Guide. https://sourceforge.net/projects/ppfs/ (accessed 7 Aug. 2016).

Hardwick, R. C. (1988). Review of recent research on navy beans (*Phaseolus vulgaris*) in the United Kingdom. *Ann. Appl. Biol.* **113**, 205–227.

Hardwick, R. C. and Andrews, D . J. (1980). A method of measuring differences between bean varieties in tolerance to sub-optimal temperatures. *Ann. Appl. Biol.* **95**, 235–247.

Hardwick, R. C. and Andrews, D. J. (1981). Seedling growth in *Phaseolus vulgaris* L. The early growth of cultivars selected for seedling cold tolerance. *Ann. Bot.* **47**, 203–214.

Hardwick, R. C., Hardaker, J. M. and Innes, N. L. (1978). Yields and components of yield of dry beans (*Phaseolus vulgaris* L.) in the United Kingdom. *J. Agric. Sci.* **90**, 291–297.

Hart, J. P. and Griffiths, P. D. (2015). Genotyping-by-sequencing enabled mapping and marker development for the *By-2* potyvirus resistance allele in common bean. *Plant Genome* **8**, doi: 10.3835/plantgenome2014.09.0058.

Harveson, R. M. (2009). Extension publication G1958: Halo blight of dry beans in Nebraska. University of Nebraska, Lincoln, USA. http://extensionpublications.unl.edu/assets/pdf/g1958.pdf (accessed 22 Sept. 2016).

Henderson, I. R. (2012). Control of meiotic recombination frequency in plant genomes. *Curr. Opin. Plant Biol.* **15**, 556–561.

Hill, K., Coyne, D. P. and Schuster, K. L. (1972). Leaf, pod and systemic chlorosis reactions in *Phaseolus vulgaris* to halo blight controlled by different genes. *J. Am. Soc. Hortic. Sci.* **97**, 494–498.

Hosfield, G. L. (2001). Seed coat color in *Phaseolus vulgaris* L., its chemistry and associated health related benefits. *Annu. Rep. Bean Improv. Coop.* **44**, 1–6.

Hougaard, B. K., Madsen, L. H., Sandal, N., de Carvalho Moretzsohn, M., Fredslund, J., Schauser, L., Nielsen, A. M., Rohde, T., Sato, S., Tabata, S., Bertioli, D. J. and Stougaard, J. (2008). Legume anchor markers link syntenic regions between *Phaseolus vulgaris*, *Lotus japonicus*, *Medicago truncatula* and Arachis. *Genetics* **179**, 2299–2312.

Hu, W. W., Gong, H. and Pua, E. C. (2005). The pivotal roles of the plant *S*-adenosylmethionine decarboxylase 5′ untranslated leader sequence in regulation of gene expression at the transcriptional and posttranscriptional levels. *Plant Physiol.* **138**, 276–286.

Hulbert, S. H., Webb, C. A., Smith, S. M. and Sun, Q. (2001). Resistance gene complexes: evolution and utilization. *Annu. Rev. Phytopathol.* **39**, 285–312.

Hyten, D., Song, Q., Fickus, E., Quigley, C., Lim, J.-S., Choi, I.-Y., Hwang, E.-Y., Pastor-Corrales, M. and Cregan, P. (2010). High-throughput SNP discovery and assay development in common bean. *BMC Genomics* **11**, 475.

Innes, N. L., Conway, J. and Taylor, J. D. (1984). Resistance to halo-blight in the Cambridge accessions V4606 and V4058 of *Phaseolus* beans. *Ann. Appl. Biol.* **104**, 307–314.

Inoue, Y. and Takikawa, Y. (1999). Investigation of repeating sequences in *hrpL* neighbouring region of *Pseudomonas syringae* strains. *Ann. Phytopathol. Soc. Jpn.* **65**, 100–109.

Iyer-Pascuzzi, A. S. and McCouch, S. R. (2007). Recessive resistance genes and the *Oryza sativa–Xanthomonas oryzae* pv. *oryzae* pathosystem. *Mol. Plant–Microbe Interact.* **20**, 731–739.

Jackson, R. W., Athanassopoulos, E., Tsiamis, G., Mansfield, J. W., Sesma, A., Arnold, D. L., Gibbon, M. J., Murillo, J., Taylor, J. D. and Vivian, A. (1999). Identification of a pathogenicity island, which contains genes for virulence and avirulence, on a large native plasmid in the bean pathogen *Pseudomonas syringae* pathovar phaseolicola. *Proc. Natl. Acad. Sci. USA* **96**, 10875–10880.

Jackson, R. W., Mansfield, J. W., Arnold, D. L., Sesma, A., Paynter, C. D., Murillo, J., Taylor, J. D. and Vivian, A. (2000). Excision from tRNA genes of a large chromosomal region, carrying *avrPphB*, associated with race change in the bean pathogen, *Pseudomonas syringae* pv. *phaseolicola*. *Mol. Microbiol.* **38**, 186–197.

Jenner, C., Hitchin, E., Mansfield, J., Walters, K., Betteridge, P., Teverson, D. and Taylor, J. (1991). Gene-for-gene interactions between *Pseudomonas syringae* pv. *phaseolicola* and *Phaseolus*. *Mol. Pant–Microbe Interact.* **4**, 553–562.

Jeong, B. R., Lin, Y., Joe, A., Guo, M., Korneli, C., Yang, H., Wang, P., Yu, M., Cerny, R. L., Staiger, D., Alfano, J. R. and Xu, Y. (2011). Structure

function analysis of an ADP-ribosyltransferase type III effector and its RNA-binding target in plant immunity. *J. Biol. Chem.* **286**, 43272–43281.

Ji, C., Boyd, C., Slaymaker, D., Okinaka, Y., Takeuchi, Y., Midland, S. L., Sims, J. J., Herman, E. and Keen N. (1998). Characterization of a 34-kDa soybean binding protein for the syringolide elicitors. *Proc. Natl. Acad. Sci. USA* **95**, 3306–3311.

Ji, C., Okinaka, Y., Takeuchi, Y., Tsurushima, T., Buzzell, R. I., Sims, J. J., Midland, S. L., Slaymaker, D., Yoshikawa, M., Yamaoka, N. and Keen, N. T. (1997). Specific binding of the syringolide elicitors to a soluble protein fraction from soybean leaves. *Plant Cell* **9**, 1425–1433.

Jia, Y., McAdams, S. A., Bryan, G. T., Hershey, H. P. and Valent, B. (2000). Direct interaction of resistance gene and avirulence gene products confers rice blast resistance. *EMBO J.* **19**, 4004–4014.

Jinek, M., Chylinski, K., Fonfara, I., Hauer, M., Doudna, J. A. and Charpentier, E. (2012). A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* **337**, 816–821.

Joardar, V., Lindeberg, M., Jackson, R. W., Selengut, J., Dodson, R., Brinkac, L. M., Daugherty, S. C., Deboy, R., Durkin, A. S., Giglio, M. G., Madupu, R., Nelson, W. C., Rosovitz, M. J., Sullivan, S., Crabtree, J., Creasy, T., Davidsen, T., Haft, D. H., Zafar, N., Zhou, L., Halpin, R., Holley, T., Khouri, H., Feldblyum, T., White, O., Fraser, C. M., Chatterjee, A. K., Cartinhour, S., Schneider, D. J., Mansfield, J., Collmer, A. and Buell, C. R. (2005). Whole-genome sequence analysis of *Pseudomonas syringae* pv. phaseolicola 1448A reveals divergence among pathovars in genes involved in virulence and transposition. *J. Bacteriol.* **187**, 6488–6498.

Jones, D. A. and Jones, J. D. G. (1997). The role of leucine-rich repeat proteins in plant defences. *Adv. Bot. Res.* **24**, 89–167.

Jones, E. S., Sullivan, H., Bhattramakki, D. and Smith, J. S. C. (2007). A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). *Theor. Appl. Genet.* **115**, 361–371.

Jones, J. D. G. and Dangl, J. L. (2006). The plant immune system. *Nature* **444**, 323–329.

Kamfwa, K., Cichy, K. A. and Kelly, J. D. (2015). Genome-wide association study of agronomic traits in common bean. *Plant Genome* **8**, doi: 0.3835/plantgenome2014.09.0059.

Kang, B.-C., Yeam, I. and Jahn, M. M. (2005). Genetics of plant virus resistance. *Annu. Rev. Phytopathol.* **43**, 581–621.

Kanzaki, H., Yoshida, K., Saitoh, H., Fujisaki, K., Hirabuchi, A., Alaux, L., Fournier, E., Tharreau, D. and Terauchi, R. (2012). Arms race co-evolution of *Magnaporthe oryzae AVR-Pik* and rice *Pik* genes driven by their physical interactions. *Plant J.* **72**, 894–907.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P. and Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649.

Keen, N. T., Tamaki, S., Kobayashi, D., Gerhold, D., Stayton, M., Shen, H., Gold, S., Lorang, J., Thordal-Christensen, H., Dahlbeck, D. and Staskawicz, B. (1990). Bacteria expressing avirulence gene D produce a specific elicitor of the soybean hypersensitive reaction. *Mol. Plant–Microbe Interact.* **3**, 112–121.

Keith, L. W., Boyd, C., Keen, N. T. and Partridge, J. E. (1997). Comparison of *avrD* alleles from *Pseudomonas syringae* pv. *glycinea*. *Mol. Plant–Microbe Interact.* **10**, 416–422.

Kelly, J. D., Gepts, P., Miklas, P. N. and Coyne, D. P. (2003). Tagging and mapping of genes and QTL and molecular marker-assisted selection for traits of economic importance in bean and cowpea. *Field Crops Res.* **82**, 135–154.

King, E. O., Ward, M. K. and Raney, D. E. (1954). Two simple media for the demonstration of pyocyanin and fluorescin. *J. Lab. Clin. Med.* **44**, 301–307.

Krasileva, K. V., Dahlbeck, D. and Staskawicz, B. J. (2010). Activation of an Arabidopsis resistance protein is specified by the in planta association of its leucine-rich repeat domain with the cognate oomycete effector. *Plant Cell* **22**, 2444–2258.

Kumar, S., Stecher, G. and Tamura, K. (2016). MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874.

Kuruvadi, S. and Aguilera, D. M. (1990). Patrones del sistema radicular en frijol común (*Phaseolus vulgaris* L.). *Turrialba* **40**, 491–498.

Kuzin, A. P., Sun, T., Jorczak-Baillass, J., Healy, V. L., Walsh, C. T. and Knox, J. R. (2000). Enzymes of vancomycin resistance: the structure of D-alanine–D-lactate ligase of naturally resistant *Leuconostoc mesenteroides*. *Structure* **8**, 463–470.

Kwak, M., Velasco, D. and Gepts, P. (2008). Mapping homologous sequences for determinacy and photoperiod sensitivity in common bean (*Phaseolus vulgaris*). *J. Hered.* **99**, 283–291.

Laffray, D. and Louguet, P. (1990). Stomatal responses and drought resistance. *B. Soc. Bot. Fr.–Actual.* **137**, 47–60.

Lamppa, R. S., Gross, P. L. and del Río, L. E. (2002). Identification of races of *Pseudomonas syringae* pv. *phaseolicola* present in North Dakota. *Phytopathology* **92**(Suppl.), S139.

Langmead, B. and Salzberg S. (2012). Fast gapped-read alignment with Bowtie 2. *Nat. Methods* **9**, 357–359.

Leach, J. E., Vera Cruz, C. M., Bai, J. and Leung, H. (2001). Pathogen fitness penalty as a predictor of durability of disease resistance genes. *Annu. Rev. Phytopathol.* **39**, 187–224.

Lee, T., Yang, S., Kim, E., Ko, Y., Hwang, S., Shin, J., Shim, J. E., Shim, H., Kim, H., Kim, C. and Lee I. (2015). AraNet v2: an improved database of co-functional gene networks for the study of *Arabidopsis thaliana* and 27 other nonmodel plant species. *Nucleic Acids Res.* **43**, D996–1002.

Lessard, I. A. D. and Walsh, C. T. (1999). VanX, a bacterial D-alanyl–D-alanine dipeptidase: resistance, immunity, or survival function? *Proc. Natl. Acad. Sci. USA* **96**, 11028–11032.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., Durbin, R. and 1000 Genome Project Data Processing Subgroup (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079.

Liao, H., Yan, X., Rubio, G., Pedraza, F., Beebe, S. and Lynch, J. P. (2004). Genetic mapping of basal root gravitropism and phosphorus acquisition efficiency in common bean. *Funct. Plant Biol.* **31**, 1–12.

Lindeberg, M., Cunnac, S. and Collmer, A. (2012). *Pseudomonas syringae* type III effector repertoires: last words in endless arguments. *Trends Microbiol.* **20**, 199–208.

Lindeberg, M., Stavrinides, J., Chang, J. H., Alfano, J. R., Collmer, A., Dangl, J. L., Greenberg, J. T., Mansfield, J. W. and Guttman, D. S. (2005). Proposed guidelines for a unified nomenclature and phylogenetic analysis of type III Hop effector proteins in the plant pathogen *Pseudomonas syringae*. *Mol. Plant–Microbe Interact.* **18**, 275–282.

Liu, B., Zhang, S., Zhu, X., Yang, Q., Wu, S., Mei, M., Mauleon, R., Leach, J., Mew, T. and Leung, H. (2004). Candidate defense genes as predictors of quantitative blast resistance in rice. *Mol. Plant–Microbe Interact.* **17**, 1146–1152.

Lorang, J., Kidarsa, T., Bradford, C. S., Gilbert, B., Curtis, M., Tzeng, S. C., Maier, C. S. and Wolpert, T. J. (2012). Tricking the guard: exploiting plant defense for disease susceptibility. *Science* **338**, 659–662.

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

Lynch, J. P. and Brown, K. (c2016). Common Bean Shovelomics. Roots Lab, Pennsylvania State University. http://plantscience.psu.edu/research/labs/roots/methods/field/shovelomics/intensive-bean-crown-phenotyping (accessed 15 Sept. 2016).

Mackey, D., Holt III, B. F., Wiig, A. and Dangl, J. L. (2002). RIN4 interacts with *Pseudomonas syringae* type III effector molecules and is required for RPM1-mediated resistance in *Arabidopsis*. *Cell* **108**, 743–754.

Magoc, T., Pabinger, S., Canzar, S., Liu, X., Su, Q., Puiu, D., Tallon, L. J. and Salzberg, S. L. (2013). GAGE-B: an evaluation of genome assemblers for bacterial organisms. *Bioinformatics* **29**, 1718–1725.

Mansfield, J., Jenner, C., Hockenhull, R., Bennett, M. A. and Stewart, R. (1994). Characterization of *avrPphE*, a gene for cultivar-specific avirulence from *Pseudomonas syringae* pv. *phaseolicola* which is physically linked to *hrpY*, a new *hrp* gene identified in the halo-blight bacterium. *Mol. Plant–Microbe Interact.* **7**, 726–739.

Maqbool, A., Saitoh, H., Franceschetti, M., Stevenson, C. E., Uemura, A., Kanzaki, H., Kamoun, S., Terauchi, R. and Banfield, M. J. (2015). Structural basis of pathogen recognition by an integrated HMA domain in a plant NLR immune receptor. *Elife* **4**, E08709.

Marshall, E., Costa, L. M. and Gutierrez-Marcos, J. (2011). Cysteine-rich peptides (CRPs) mediate diverse aspects of cell–cell communication in plant reproduction and development. *J. Exp. Bot.* **62**, 1677–1686.

McFarland, J. (1907). The nephelometer: an instrument for estimating the number of bacteria in suspensions used for calculating the opsonic index and for vaccines. *JAMA* **14**, 1176–1178.

McHale, L., Tan, X., Koehl, P. and and Michelmore, R. W. (2006). Plant NBS–LRR proteins: adaptable guards. *Genome Biol.* **7**, 212.

McClean, P. E., Burridge, J., Beebe, S., Rao, I. M. and Porch, T. G. (2011). Crop improvement in the era of climate change: an integrated, multi-disciplinary approach for common bean (*Phaseolus vulgaris*). *Funct. Plant Biol.* **38**, 927–933.

McClean, P. E., Lee, R. K., Otto, C., Gepts, P. and Bassett, M. J. (2002). Molecular and phenotypic mapping of genes controlling seed coat pattern and color in common bean (*Phaseolus vulgaris* L.). *J. Hered.* **93**, 148–152.

McConnell, M., Mamidi, S., Lee, R., Chikara, S., Rossi, M., Papa, R. and McClean, P. (2010). Syntenic relationships among legumes revealed using a gene-based genetic linkage map of common bean (*Phaseolus vulgaris* L.). *Theor. Appl. Genet.* **121**, 1103–1116.

Mengiste, T. (2012). Plant immunity to necrotrophs. *Annu. Rev. Phytopathol.* **50**, 267–294.

Meschini, E. P., Blanco, F. A., Zanetti, M. E., Beker, M. P., Küster, H., Pühler, A. and Aguilar, O. M. (2008). Host genes involved in nodulation preference in common bean (*Phaseolus vulgaris*)–*Rhizobium etli* symbiosis revealed by suppressive subtractive hybridization. *Mol. Plant–Microbe Interact.* **21**, 459–468.

Micheletto, S., Rodriguez-Uribe, L., Hernandez, R., Richins, R. D., Curry, J. and O'Connell, M. A. (2007). Comparative transcript profiling in roots of *Phaseolus acutifolius* and *P. vulgaris* under water deficit stress. *Plant Sci.* **173**, 510–520.

Midland, S. L., Keen, N. T. and Sims, J. J. (1995). Secosyrins 1 and 2 and syributins 1 and 2: novel structures produced by bacteria expressing the *avrD* gene. *J. Org. Chem.* **60**, 1118–1119.

Midland, S. L., Keen, N. T., Sims, J. J., Midland, M. M., Stayton, M. M., Burton, V., Smith, M. J., Mazzola, E. P., Graham, J. K. and Clardy, J. (1993). The structures of syringolides 1 and 2, novel C-glycosidic elicitors from *Pseudomonas syringae* pv. *tomato*. *J. Org. Chem.* **58**, 2940–2945.

Miedaner, T., Risser, P., Paillard, S., Schnurbusch, T., Keller, B., Hartl, L., Holzapfel, J., Korzun, V., Ebmeyer, E. and Utz, H. F. (2012). Broad-spectrum resistance loci for three quantitatively inherited diseases in two winter wheat populations. *Mol. Breed.* **29**, 731–742.

Miguel, M. A., Brown, K., Blair, M. and Lynch, J. P. (2009). QTL mapping of basal root whorl number in common bean (*Phaseolus vulgaris* L.). Poster presentation. Footprints in the Landscape: Sustainability through Plant and Soil Sciences. Pittsburgh, USA. https://scisoc.confex.com/crops/2009am/webprogram/Paper55496.html (accessed 24 Feb. 2016).

Miguel, M. A., Postma, J. A. and Lynch, J. P. (2015). Phene synergism between root hair length and basal root growth angle for phosphorous acquisition. *Plant Physiol.* **167**, 1430–1439.

Miguel, M. A., Widrig, A., Vieira, R. F., Brown, K. M. and Lynch, J. P. (2013). Basal root whorl number: a modulator of phosphorus acquisition in common bean (*Phaseolus vulgaris*). *Ann. Bot.* **112**, 973–982.

Miklas, P. N., Coyne, D. P., Grafton, K. F., Mutlu, N., Reiser, J., Lindgren, D.

and Singh, S. P. (2003). A major QTL for common bacterial blight resistance derives from the common bean great northern landrace cultivar Montana No. 5. *Euphytica* **131**, 137–146.

Miklas, P. N., Fourie, D., Trapp, J., Davis, J. and Myers, J. R. (2014). New loci including *Pse-6* conferring resistance to halo bacterial blight on chromosome Pv04 in common bean. *Crop Sci.* **54**, 2099–2108.

Miklas, P. N., Fourie, D., Trapp, J., Larsen, R. C., Chavarro, C., Blair, M. W. and Gepts, P. (2011). Mapping *Pse-2* gene for resistance to halo blight in common bean. *Crop Sci.* **51**, 2439–2448.

Miklas, P. N., Fourie, D., Wagner, J., Larsen, R. C. and Mienie, C. M. S. (2009). Tagging and mapping *Pse-1* gene for resistance to halo blight in common bean differential cultivar UI-3. *Crop Sci.* **49**, 41–48.

Miklas, P. N., Kelly, J. D., Beebe, S. E. and Blair, M. W. (2006). Common bean breeding for resistance against biotic and abiotic stresses: from classical to MAS breeding. *Euphytica* **147**, 105–131.

Miller, C., Ochoa, I., Nielsen, K. L., Beck, D. and Lynch, J. P. (2003). Genetic variation for adventitious rooting in response to low phosphorus availability: potential utility for phosphorus acquisition from stratified soils. *Funct. Plant Biol.* **30**, 973–985.

Miller, J. C., Chezem, W. R. and Clay, N. K. (2016). Ternary WD40 repeat-containing protein complexes: evolution, composition and roles in plant immunity. *Front. Plant Sci.* **6**, 1108.

Miller, P. R., Waddington, J., McDonald, C. L. and Derksen, D. A. (2002). Cropping sequence affects wheat productivity on the semiarid northern Great Plains. *Can. J. Plant Sci.* **82**, 307–318.

Mindrinos, M., Katagiri, F., Yu, G. L. and Ausubel, F. M. (1994). The *A. thaliana* disease resistance gene *RPS2* encodes a protein containing a nucleotide-binding site and leucine-rich repeats. *Cell* **78**, 1089–1099.

Moghaddam, S. M., Song, Q., Mamidi, S., Schmutz, J., Lee, R., Cregan, P., Osorno, J. M. and McClean, P. E. (2014). Developing market class specific InDel markers from next generation sequence data in *Phaseolus vulgaris* L. *Front. Plant Sci.* **5**, 185.

Murillo I., Á., Falconí, E., Marzón, N. and Peralta I., E. (2006). Resistance sources for rust, angular leaf spot, and common bacterial blight in common bean for Ecuador. *Annu. Rep. Bean Improv. Coop.* **49**, 229–230.

Narusaka, Y., Narusaka, M., Seki, M., Umezawa, T., Ishida, J., Nakajima, M., Enju, A. and Shinozaki, K. (2004). Crosstalk in the responses to abiotic and biotic stresses in *Arabidopsis*: analysis of gene expression in *cytochrome P450* gene superfamily by cDNA microarray. *Plant Mol. Biol.* **55**, 327–342.

Nicaise, V., Joe, A., Jeong, B. R., Korneli, C., Boutrot, F., Westedt, I., Staiger, D., Alfano, J. R. and Zipfel, C. (2013). *Pseudomonas* HopU1 modulates plant immune receptor levels by blocking the interaction of their mRNAs with GRP7. *EMBO J.* **32**, 701–712.

Noël, L., Moores, T. L., van der Biezen, E. A., Parniske, M., Daniels, M. J., Parker, J. E. and Jones, J. D. G. (1999). Pronounced intraspecific haplotype divergence at the *RPP5* complex disease resistance locus of Arabidopsis. *Plant Cell* **11**, 2099–2111.

Norman, M. J. T., Pearson, C. J. and Searle, P. G. E. (1995). *The Ecology of Tropical Food Crops*. Cambridge, UK: Cambridge University Press.

Notredame, C., Higgins, D. G. and Heringa J. (2000). T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217.

Oblessuc, P. R., Baroni, R. M., Garcia, A. A., Chioratto, A. F., Carbonell, S. A., Camargo, L. E. and Benchimol, L. L. (2012). Mapping of angular leaf spot resistance QTL in common bean (*Phaseolus vulgaris* L.) under different environments. *BMC Genomics* **13**, 50.

Okinaka, Y., Yang, C. H., Herman, E., Kinney, A. and Keen, N. T. (2002). The P34 syringolide elicitor receptor interacts with a soybean photorespiration enzyme, NADH-dependent hydroxypyruvate reductase. *Mol. Plant–Microbe Interact.* **15**, 1213–1218.

Oliphant, A., Barker, D. L., Stuelpnagel, J. R. and Chee, M. S. (2002). BeadArray^{TM} technology: enabling an accurate, cost-effective approach to high-throughput genotyping. *Biotechniques* **32**, S56–S61.

Orjuela, J., Deless, E. F. T., Kolade, O., Cheron, S., Ghesquiere, A. and Albar, L. (2013). A recessive resistance to *Rice yellow mottle virus* is associated with a rice homolog of the *CPR5* gene, a regulator of active defense mechanisms. *Mol. Plant–Microbe Interact.* **26**, 1455-1463.

Ortiz-Martín, I., Thwaites, R., Macho, A. P., Mansfield, J. W. and Beuzón, C. R. (2010a). Positive regulation of the Hrp type III secretion system in *Pseudomonas syringae* pv. *phaseolicola*. *Mol. Plant–Microbe Interact.* **23**, 665–681.

Ortiz-Martín, I., Thwaites, R., Mansfield, J. W. and Beuzón, C. R. (2010b). Negative regulation of the Hrp type III secretion system in *Pseudomonas syringae* pv. *phaseolicola*. *Mol. Plant–Microbe Interact.* **23**, 682–701.

Osmond, C. B., Winter, K. and Powles, S. B. (1980). Adaptive significance of carbon dioxide cycling during photosynthesis in water-stressed plants. In *Adaptation of Plants to Water and High Temperature Stress*, pp. 139–154. Edited by N. C. Turner and P. J. Kramer. New York: John Wiley and Sons.

Ozias-Akins, P., Roche, D. and Hanna, W. W. (1998). Tight clustering and hemizygosity of apomixis-linked molecular markers in *Pennisetum squamulatum* implies genetic control of apospory by a divergent locus that may have no allelic form in sexual genotypes. *Proc. Natl. Acad. Sci. USA* **95**, 5127–5132.

Parker, J. E., Coleman, M. J., Szabò, V., Frost, L. N., Schmidt, R., van der Biezen, E. A., Moores, T., Dean, C., Daniels, M. J., and Jones, J. D. G. (1997). The Arabidopsis downy mildew resistance gene *RPP5* shares similarity to the Toll and interleukin-1 receptors with *N* and *L6*. *Plant Cell* **9**, 879–894.

Passioura, J. (2006). Increasing crop productivity when water is scarce—from breeding to field management. *Agr. Water Manage.* **80**, 176–196.

Patel, P. N. and Walker, J. C. (1965). Resistance in *Phaseolus* to halo blight. *Phytopathology* **55**, 889–894.

Patel, P. N. and Walker, J. C. (1966). Inheritance of tolerance to halo blight in bean. *Phytopathology* **56**, 681–682.

Perseguini, J. M. K. C., Oblessuc, P. R., Rosa, J. R. B. F., Gomes, K. A., Chiorato, A. F., Carbonell, S. A. M., Garcia, A. A. F., Vianello, R. P. and Benchimol-Reis, L. L. (2016). Genome-wide association studies of anthracnose and angular leaf spot resistance in common bean (*Phaseolus vulgaris* L.). *PLoS ONE* **11**, E0150506.

Pimentel, C., Laffray, D. and Louguet, P. (1999). Intrinsic water use efficiency at the pollination stage as a parameter for drought tolerance selection in *Phaseolus vulgaris*. *Physiol. Plant.* **106**, 184–189.

Pitman, A. R., Jackson, R. W., Mansfield, J. W., Kaitell, V., Thwaites, R. and Arnold, D. L. (2005). Exposure to host resistance mechanisms drives evolution of bacterial virulence in plants. *Curr. Biol.* **15**, 2230–2235.

Prosen, D., Hatziloukas, E., Schaad, N. W. and Panopoulos, N. J. (1993). Specific detection of *Pseudomonas syringae* pv. *phaseolicola* DNA in bean seed by polymerase chain reaction-based amplification of a phaseolotoxin gene region. *Phytopathology* **83**, 965–970.

Qi, Y., Tsuda, K., Joe, A., Sato, M., Nguyen, L. V., Glazebrook, J., Alfano, J. R., Cohen, J. D. and Katagiri, F. (2010). A putative RNA-binding protein positively regulates salycillic acid-mediated immunity in *Arabidopsis*. *Mol. Plant–Microbe Interact.* **23**, 1573–1583.

Rambaut, A. (2014). FigTree version 1.4.2. http://tree.bio.ed.ac.uk/software/figtree/ (accessed 7 Aug. 2016).

Ramírez, M., Graham, M. A., Blanco-López, L., Silvente, S., Medrano-Soto, A., Blair, M. W., Hernández, G., Vance, C. P. and Lara, M. (2005). Sequencing and analysis of common bean ESTs. Building a foundation for functional genomics. *Plant Physiol.* **137**, 1211–1227.

R Core Team (2016). *R: A Language and Environment for Statistical Computing, Reference Index Version 3.3.0*. Vienna, Austria: R Foundation for Statistical Computing. https://cran.r-project.org/doc/manuals/r-release/fullrefman.pdf (accessed 28 Jun. 2016).

Read, T. D. and Massey, R. C. (2014). Characterizing the genetic basis of bacterial phenotypes using genome-wide association studies: a new direction for bacteriology. *Genome Med.* **6**, 109.

Reinprecht, Y., Yadegari, Z., Perry, G. E., Siddiqua, M., Wright, L. C., McClean, P. E. and Pauls, K. P. (2013). *In silico* comparison of genomic regions containing genes coding for enzymes and transcription factors for the phenylpropanoid pathway in *Phaseolus vulgaris* L. and *Glycine max* L. Merr. *Front. Plant Sci.* **4**, 317.

Repinski, S. L., Kwak, M. and Gepts, P. (2012). The common bean growth habit gene *PvTFL1y* is a functional homolog of *Arabidopsis TFL1*. *Theor. Appl. Genet.* **124**, 1539–1547.

Richter, T. E., Pryor, T. J., Bennetzen, J. L. and Hulbert, S. H. (1995). New rust resistance specificities associated with recombination in the *Rp1* complex in maize. *Genetics* **141**, 373–381.

Rico, A., López, R., Asensio, C., Aizpún, M. T, Asensio-S-Manzanera, M. C and Murillo, J. (2003). Nontoxigenic strains of *Pseudomonas syringae* pv. *phaseolicola* are a main cause of halo blight of beans in Spain and escape current detection methods. *Phytopathology* **93**, 1553–1559.

Rivas, L. A., Mansfield, J., Tsiamis, G., Jackson, R. W. and Murillo, J. (2005). Changes in race-specific virulence in *Pseudomonas syringae* pv. *phaseolicola* are associated with a chimeric transposable element and rare deletion events in a plasmid-borne pathogenicity island. *Appl. Environ. Microbiol.* **71**, 3778–3785.

Rodiño, A. P., Lema, M., Pérez-Barbeito, M., Santalla, M. and De Ron, A. M. (2007). Assessment of runner bean (*Phaseolus coccineus* L.) germplasm for tolerance to low temperature during early seedling growth. *Euphytica* **155**, 63–70.

Sanabria, N., Goring, D., Nürnberger, T. and Dubery, I. (2008). Self/nonself perception and recognition mechanisms in plants: a comparison of self-incompatibility and innate immunity. *New Phytol.* **178**, 503–514.

Sarris, P. F., Cevik, V., Dagdas, G., Jones, J. D. and Krasileva, K. V. (2016). Comparative analysis of plant immune receptor architectures uncovers host proteins likely targeted by pathogens. *BMC Biol.* **14**, 8.

Sarris, P. F., Duxbury, Z., Huh, S. U., Ma, Y., Segonzac, C., Sklenar, J., Derbyshire, P., Cevik, V., Rallapalli, G., Saucet, S. B., Wirthmueller, L., Menke, F. L., Sohn, K. H. and Jones, J. D. (2015). A plant immune receptor detects pathogen effectors that target WRKY transcription factors. *Cell* **161**, 1089–1100.

Saucet, S. B., Ma, Y., Sarris, P. F., Furzer, O. J., Sohn, K. H. and Jones, J. D. (2015). Two linked pairs of Arabidopsis TNL resistance genes independently confer recognition of bacterial effector AvrRps4. *Nat. Commun.* **6**, 6338.

Schmutz, J., McClean, P. E., Mamidi, S., Wu, G. A., Cannon, S. B., Grimwood, J., Jenkins, J., Shu, S., Song, Q., Chavarro, C., Torres-Torres, M., Geffroy, V., Moghaddam, S. M., Gao, D., Abernathy, B., Barry, K., Blair, M., Brick, M. A., Chovatia, M., Gepts, P., Goodstein, D. M., Gonzales, M., Hellsten, U., Hyten, D. L., Jia, G., Kelly, J. D., Kudrna, D., Lee, R., Richard, M. M., Miklas, P. N., Osorno, J. M., Rodrigues, J., Thareau, V., Urrea, C. A., Wang, M., Yu, Y., Zhang, M., Wing, R. A., Cregan, P. B., Rokhsar, D. S. and Jackson, S. A. (2014). A reference genome for common bean and genome-wide analysis of dual domestications. *Nat. Genet.* **46**, 707–713. Sequence data are available at http://www.phytozome.net/commonbean (accessed 23 Sept. 2016).

Schneider, C. A., Rasband, W. S. and Eliceiri, K. W. (2012). NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675.

Schwartz, H. F. (2016). Halo blight (*Pseudomonas savastanoi* pv. *phaseolicola*) image numbers 5444724 and 5361428. Howard F. Schwartz, Colorado State University, Bugwood.org. http://www.ipmimages.org/browse/detail.cfm?imgnum=5444724 and http://www.ipmimages.org/browse/detail.cfm?imgnum=5361428 (accessed 27 Sept. 2016). Creative Commons Attribution 3.0 licence: https://creativecommons.org/licenses/by/3.0/us/ (accessed 27 Sept. 2016).

Schwartz, H. F., Brick, M. A., Harveson, R. M. and Franc, G. D. (Eds.) (2004). *Dry Bean Production and Integrated Pest Management*, 2nd edn. Colorado State University, University of Nebraska and University of Wyoming, USA.

Seki, M., Kamei, A., Yamaguchi-Shinozaki, K. and Shinozaki, K. (2003). Molecular responses to drought, salinity and frost: common and different paths for plant protection. *Curr. Opin. Biotechnol.* **14**, 194–199.

Seki, M., Narusaka, M., Abe, H., Kasuga, M., Yamaguchi-Shinozaki, K., Carninci, P., Hayashizaki, Y. and Shinozaki, K. (2001). Monitoring the expression pattern of 1300 Arabidopsis genes under drought and cold stresses by using a full-length cDNA microarray. *Plant Cell* **13**, 61–72.

Seki, M., Narusaka, M., Ishida, J., Nanjo, T., Fujita, M., Oono, Y., Kamiya, A., Nakajima, M., Enju, A., Sakurai, T., Satou, M., Akiyama, K., Taji, T., Yamaguchi-Shinozaki, K., Carninci, P., Kawai, J., Hayashizaki, Y. and Shinozaki, K. (2002). Monitoring the expression profiles of 7000 Arabidopsis genes under drought, cold and high-salinity stresses using a full-length cDNA microarray. *Plant J.* **31**, 279–292.

Selote, D., Shine, M. B., Robin, G. P. and Kachroo, A. (2014). Soybean NDR1-like proteins bind pathogen effectors and regulate resistance signaling. *New Phytol.* 202, 485–498.

Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* **13**, 2498–2504.

Shao, F., Goldstein, C., Ade, J., Stoutemyer, M., Dixon, J. E. and Innes, R. W. (2003). Cleavage of *Arabidopsis* PBS1 by a bacterial type III effector. *Science* **301**, 1230–1233.

Shao, Z. Q., Zhang, Y. M., Hang, Y. Y., Xue, J. Y., Zhou, G. C., Wu, P., Wu, X. Y., Wu X. Z., Wang, Q., Wang, B. and Chen, J. Q. (2014). Long-term evolution of nucleotide-binding site–leucine-rich repeat genes: understanding gained from and beyond the legume family. *Plant Physiol*. **166**, 217–234.

Sharma, M., Pandey, A. and Pandey, G. K. (2014). β-catenin in plants and animals: common players but different pathways. *Front. Plant Sci.* **5**, 143.

Sheng, J., D'Ovidio, R. and Mehdy, M. C. (1991). Negative and positive regulation of a novel proline-rich protein mRNA by fungal elicitor and wounding. *Plant J.* **1**, 345–354.

Shinozaki, K. and Yamaguchi-Shinozaki, K. (2006). Global analysis of gene networks to solve complex abiotic stress responses. In *Cold Hardiness in Plants: Molecular Genetics, Cell Biology and Physiology*, pp. 1–10. Edited by T. H. H. Chen, M. Uemura and S. Fujikawa. Wallingford, UK: CABI Publishing.

Shisanya, C. A. and Gitonga, N. M. (2007). Evaluation of nitrogen fixation using 15N dilution methods and economy of a maize–tepary bean intercrop farming system in semi-arid SE-Kenya. In *Advances in Integrated Soil Fertility Management in sub-Saharan Africa: Challenges and Opportunities*, pp. 389–400. Edited by A. Bationo, B. S. Waswa, J. Kihara and J. Kimetu. Berlin: Springer.

Silbernagel, M. J. (1977). Seed quality index as an indicator of crop production potential and a selection tool for the genetic improvement of snap been seed quality. *Annu. Rep. Bean Improv. Coop.* **20**, 40–42.

Silbernagel, M. J. and Hannan, R. M. (1992). Use of plant introductions to develop U.S. bean cultivars. In *Use of Plant Introductions in Cultivar Development Part 2*, pp. 1–8. Edited by H. L. Shands and L. E. Weisner. CSSA Special Publication 20. Madison, WI: Crop Science Society of America.

Silverstein, K. A. T., Graham, M. A., Paape, T. D. and VandenBosch, K. A. (2005). Genome organization of more than 300 defensin-like genes in Arabidopsis. *Plant Physiol.* **138**, 600–610.

Simmons, M. P., Pickett, K. M. and Miya, M. (2004). How meaningful are Bayesian support values? *Mol. Biol. Evol.* **21**, 188–199.

Singh, S. P. and Schwartz, H. F. (2010). Breeding common bean for resistance to diseases: a review. *Crop Sci.* **50**, 2199–2223.

Smith, M. J., Mazzola, E. P., Simms, J. J., Midland, S. L., Keen, N. T., Burton, V. and Stayton, M. M. (1993). The syringolides: bacterial C-glycosyl lipids that trigger plant disease resistance. *Tetrahedron Lett.* **34**, 223–226.

Sohn, K. H., Segonzac, C., Rallapalli, G., Sarris, P. F., Woo, J. Y., Williams, S. J., Newman, T. E., Paek, K. H., Kobe, B. and Jones, J. D. (2014). The nuclear immune receptor RPS4 is required for RRS1SLH1-dependent constitutive defense activation in *Arabidopsis thaliana*. *PLoS Genet.* **10**, E1004655.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., Normandeau, É., Laroche, J., Larose, S., Jean, M. and Belzile, F. (2013). An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS ONE* **8**, E54603.

Song, Q., Jia, G., Hyten, D. L., Jenkins, J., Hwang, E. Y., Schroeder, S. G., Osorno, J. M., Schmutz, J., Jackson, S. A., McClean, P. E. and Cregan, P. B. (2015). SNP assay development for linkage map construction, anchoring whole-genome sequence, and other genetic and genomic applications in common bean. *G3 (Bethesda)* **5**, 2285–2290.

Staiger, D., Korneli, C., Lummer, M. and Navarro, L. (2013). Emerging role for RNA-based regulation in plant immunity. *New Phtyol.* **197**, 394–404.

Stamatakis, A. (2014). RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313.

Stavrinides, J. (2009). Origin and evolution of phytopathogenic bacteria. In *Plant Pathogenic Bacteria: Genomics and Molecular Biology*, pp. 1–35. Edited by R. W. Jackson. Norfolk, UK: Caister Academic Press.

Stevens, C., Bennett, M. A., Athanassopoulos, E., Tsiamis, G., Taylor, J. D. and Mansfield, J. W. (1998). Sequence variations in alleles of the avirulence

gene *avrPphE.R2* from *Pseudomonas syringae* pv. *phaseolicola* lead to loss of recognition of the AvrPphE protein within bean cells and a gain in cultivar-specific virulence. *Mol. Microbiol.* **29**, 165–177.

Sweat, T. A., Lorang, J. M., Bakker, E. G. and Wolpert, T. J. (2008). Characterization of natural and induced variation in the *LOV1* gene, a CC-NB-LRR gene conferring victorin sensitivity and disease susceptibility in *Arabidopsis*. *Mol. Plant–Microbe Interact.* **21**, 7–19.

Tatusova, T., Ciufo, S., Fedorov, B., O'Neill, K. and Tolstoy I. (2014). RefSeq microbial genomes database: new representation and annotation strategy. *Nucleic Acids Res.* **42**, D553–D559.

Taylor, J. D., Innes, N. L., Dudley, C. L. and Griffiths, W. A. (1978). Sources and inheritance of resistance to halo-blight of *Phaseolus* beans. *Ann. Appl. Biol.* **90**, 101–110.

Taylor, J. D., Teverson, D. M., Allen, D. J. and Pastor-Corrales, M. A. (1996a). Identification and origin of races of *Pseudomonas syringae* pv. *phaseolicola* from Africa and other bean growing areas. *Plant Pathol.* **45**, 469–478.

Taylor, J. D., Teverson, D. M. and Davis, J. H. C. (1996b). Sources of resistance to *Pseudomonas syringae* pv. *phaseolicola* races in *Phaseolus vulgaris*. *Plant Pathol.* **45**, 479–485.

Templeton, M. D., Rikkerink, E. H. A. and Beever, R. E. (1994). Small, cysteine-rich proteins and recognition in fungal–plant interactions. *Mol. Plant–Microbe Interact.* **7**, 320–325.

Tesfaye, M., Silverstein, K. A. T., Nallu, S., Wang, L., Botanga, C. J., Gomez, S. K., Costa, L. M., Harrison, M. J., Samac, D. A., Glazebrook, J., Katagiri, F., Gutierrez-Marcos, J. F., VandenBosch, K. A. (2013). Spatio-temporal expression patterns of *Arabidopsis thaliana* and *Medicago truncatula* defensin-like genes. *PLoS ONE* **8**, E58992.

Teverson, D. M. (1991). Genetics of pathogenicity and resistance in the halo-blight disease of beans in Africa. Ph.D. thesis. University of Birmingham, Birmingham, UK.

Teverson, D. M. (2003). Participatory Promotion of Disease Resistant and Farmer Acceptable *Phaseolus* Beans in the Southern Highlands of Tanzania. Final Technical Report R7569 (ZA 0374). Crop Protection Programme, Natural Resources Institute.

Thiel, T., Kota, R., Grosse, I., Stein, N. and Graner, A. (2004). SNP2CAPS: a SNP and INDEL analysis tool for CAPS marker development. *Nucleic Acids Res.* **32**, E5.

Thring, O. (2011). Consider baked beans. *Guardian* (22 Feb.). http://www.guardian.co.uk/lifeandstyle/wordofmouth/2011/feb/22/consider-baked-beans (accessed 23 Sept. 2016).

Timme, R. E., Pettengill, J. B., Allard, M. W., Strain E, Barrangou R, Wehnes C, Van Kessel, J. S., Karns, J. S., Musser, S. M. and Brown, E. W. (2013). Phylogenetic diversity of the enteric pathogen *Salmonella enterica* subsp. *enterica* inferred from genome-wide reference-free SNP characters. *Genome Biol. Evol.* **5**, 2109–2123.

Tornero, P., Chao, R. A., Luthin, W. N., Goff, S. A. and Dangl, J. L. (2002). Large-scale structure-function analysis of the Arabidopsis RPM1 disease resistance protein. *Plant Cell* **14**, 435–450.

Toruño, T. Y., Stergiopoulos, I. and Coaker, G. (2016). Plant pathogen effectors: cellular probes interfering with plant defenses in a spatial and temporal manner. *Annu. Rev. Phytopathol.* **54**, doi: 10.1146/annurev-phyto-080615-100204.

Trabanco, N., Asensio-Manzanera, M. C., Pérez-Vega, E., Ibeas, A., Campa, A. and Ferreira, J. J. (2014). Identification of quantitative trait loci involved in the response of common bean to *Pseudomonas syringae* pv. *phaseolicola*. *Mol. Breed.* **33**, 577–588.

Tsiamis, G., Mansfield, J. W., Hockenhull, R., Jackson, R. W., Sesma, A., Athanassopoulos, E., Bennett, M. A., Stevens, C., Vivian, A., Taylor, J. D. and Murillo, J. (2000). Cultivar-specific avirulence and virulence functions assigned to *avrPphF* in *Pseudomonas syringae* pv. *phaseolicola*, the cause of bean halo-blight disease. *EMBO J.* **19**, 3204–3214.

Ueda, H., Yamaguchi, Y. and Sano, H. (2006). Direct interaction between the tobacco mosaic virus helicase domain and the ATP-bound resistance protein, N factor during the hypersensitive response in tobacco plants. *Plant Mol. Biol.* **61**, 31–45.

Vallejos, C. E., Astua-Monge, G., Jones, V., Plyler, T. R., Sakiyama, N. S. and Mackenzie, S. A. (2006). Genetic and molecular characterization of the *I* locus of *Phaseolus vulgaris*. *Genetics* **172**, 1229–1242.

Valdés-López, O., Thibivilliers, S., Qiu, J., Xu, W. W., Nguyen, T. H., Libault, M., Le, B. H., Goldberg, R. B., Hill, C. B., Hartman, G. L., Diers, B. and Stacey, G. (2011). Identification of quantitative trait loci controlling gene expression during the innate immunity response of soybean. *Plant Physiol.* **157**, 1975–1986.

van der Hoorn, R. A. and Kamoun, S. (2008). From Guard to Decoy: a new model for perception of plant pathogen effectors. *Plant Cell* **20**, 2009–2017.

Van Schoonhoven, A. and Pastor-Corrales, M. A. (1987). *Standard System for the Evaluation of Bean Germplasm*. Cali, Colombia: Centro Internacional de

Agricultura Tropical (CIAT).

Vencato, M., Tian, F., Alfano, J. R., Buell, C. R., Cartinhour, S., DeClerck, G. A., Guttman, D. S., Stavrinides, J., Joardar, V., Lindeberg, M., Bronstein, P. A., Mansfield, J. W., Myers, C. R., Collmer, A. and Schneider, D. J. (2006). Bioinformatics-enabled identification of the HrpL regulon and type III secretion system effector proteins of *Pseudomonas syringae* pv. *phaseolicola* 1448A. *Mol. Plant–Microbe Interact.* **19**, 1193–1206.

Vicente, J. M. G. N. (2000). Diversidade da *Xanthomonas campestres* pv. *Campestres* (Pammel) Dowson e caracterização da resistência genética em *Brassica* spp. Ph.D. thesis. Doutoramento em Engenharia Agronómica, Lisboa.

Vicente, J. G., Gunn, N. D., Bailey, L., Pink, D. A. C. and Holub, E. B. (2012). Genetics of resistance to downy mildew in *Brassica oleracea* and breeding towards durable disease control for UK vegetable production. *Plant Pathol.* **61**, 600–609.

Vijayan, P., Parkin, I. A. P., Karcz, S. R., McGowan, K., Vijayan, K., Vandenberg, A. and Bett, K. E. (2011). Capturing cold-stress-related sequence diversity from a wild relative of common bean (*Phaseolus angustissimus*). *Genome* **54**, 620–628.

Walker, J. C. and Patel, P. N. (1964). Inheritance of resistance to halo blight of bean. *Phytopathology* **54**, 952–954.

Wang, X., Richards, J., Gross, T., Druka, A., Kleinhofs, A., Steffenson, B., Acevedo, M. and Brueggeman, R. (2013). The *rpg3*-mediated resistance to wheat stem rust (*Puccinia graminis*) in barley (*Hordeum vulgare*) requires *Rpg5*, a second NBS-LRR gene, and an actin depolymerization factor. *Mol. Plant–Microbe Interact.* **26**, 407–418.

Wang, Y., Wei, X., Bao, H. and Liu, S. L. (2014). Prediction of bacterial type IV secreted effectors by C-terminal features. *BMC Genomics* **15**, 50.

Wei, F., Wing, R. A. and Wise, R. P. (2002). Genome dynamics and evolution of the *Mla* (powdery mildew) resistance locus in barley. *Plant Cell* **14**, 1903–1917.

White, J. W., Castillo, J. A. and Ehleringer, J. (1990). Association between productivity, root growth and carbon isotope discrimination in *Phaseolus vulgaris* under water stress. *Aust. J. Plant Physiol.* **17**, 189–198.

White, J. W., Castillo, J. A., Ehleringer, J. R., Garcia, J. A. C. and Singh, S. P. (1994b). Relations of carbon isotope discrimination and other physiological traits to yield in common bean (*Phaseolus vulgaris*) under rainfed conditions. *J. Agric. Sci.* **122**, 275–284.

White, J. W., Ochoa, R. M., Ibarra, F. P. and Singh, S. P. (1994a). Inheritance of seed yield, maturity and seed weight of common bean (*Phaseolus vulgaris*) under semi-arid rainfed conditions. *J. Agric. Sci.* **122**, 265–273.

Wiesner-Hanks, T. and Nelson, R. (2016). Multiple disease resistance in plants. *Annu. Rev. Phytopathol.* **54**, doi: 10.1146/annurev-phyto-080615-100037.

Winsor, G. L., Griffiths, E. J., Lo, R., Dhillon, B. K., Shay, J. A. and Brinkman, F. S. (2016). Enhanced annotations and features for comparing thousands of *Pseudomonas* genomes in the Pseudomonas genome database. *Nucleic Acids Res.* **44**, D646–D653.

Woloshen, V., Huang, S. and Li, X. (2011). RNA-binding proteins in plant immunity. *J. Pathog.* **2011**, doi:10.4061/2011/278697.

Wong, J. H., Zhang, X. Q., Wang, H. X. and Ng, T. B. (2006). A mitogenic defensin from white cloud beans (*Phaseolus vulgaris*). *Peptides* **27**, 2075–2081.

Woronuk, G., Vijayan, P., Laberge, S., Vandenberg, B. and Bett, K. (2010). Transcriptomic analysis of chilling stress in *Phaseolus* spp. *Environ. Exp. Bot.* **69**, 95–104.

Wortmann, C. S., Kirkby, R. A., Eledu, C. A. and Allen, D. J. (c2004). *Atlas of Common Bean (*Phaseolus vulgaris *L.) Production in Africa*. Kampala, Uganda: International Center for Tropical Agriculture (CIAT). http://ciat-library.ciat.cgiar.org/Articulos_CIAT/Atlas_common_bean_Africa.pdf (accessed 29 Sept. 2016).

Wu, J. L., Sinha, P. K., Variar, M., Zheng, K. L., Leach, J. E., Courtois, B. and Leung, H. (2004). Association between molecular markers and blast resistance in an advanced backcross population of rice. *Theor. Appl. Genet.* **108**, 1024–1032.

Wu, X., Sun, J., Zhang, G., Wang, H. and Ng, T. B. (2011). An antifungal defensin from *Phaseolus vulgaris* cv. 'Cloud Bean'. *Phytomedicine* **18**, 104–109.

Yaish, M. W. F., Sosa, D., Vences, F. J. and Vaquero, F. (2006). Genetic mapping of quantitative resistance to race 5 of *Pseudomonas syringae* pv. *phaseolicola* in common bean. *Euphytica* **152**, 397–404.

Yang, B., Sugio, A. and White, F. F. (2006). *Os8N3* is a host disease-susceptibility gene for bacterial blight of rice. *Proc. Natl. Acad. Sci. USA* **103**, 10503–10508.

Young, J. M., Dye, D. W., Bradbury, J. F., Panagopoulos, C. G. and Robbs, C. F. (1978). A proposed nomenclature and classification for plant pathogenic bacteria. *New Zeal. J. Agr. Res.* **21**, 153–177.

Yucel, I., Boyd, C., Debnam, Q. and Keen, N. T. (1994a). Two different classes of *avrD* alleles occur in pathovars of *Pseudomonas syringae*. *Molecular Plant–Microbe Interact.* **7**, 131–139.

Yucel, I. and Keen, N. T. (1994). Amino acid residues required for the activity of *avrD* alleles. *Mol. Plant–Microbe Interact.* **7**, 140–147.

Yucel, I., Midland, S. L., Sims, J. J. and Keen, N. T. (1994b). Class I and class II *avrD* allles direct the production of different products in Gram-negative bacteria. *Mol. Plant–Microbe Interact.* **7**, 148–150.

Zhang, J., Lu, H., Li, X., Li, Y., Cui, H., Wen, C. K., Tang, X., Su, Z. and Zhou, J. M. (2010). Effector-triggered and pathogen-associated molecular pattern-triggered immunity differentially contribute to basal resistance to *Pseudomonas syringae*. *Mol. Plant–Microbe Interact.* **23**, 940–948.

Zhang, S. and Mehdy, M. C. (1994). Binding of a 50-kD protein to a U-rich sequence in an mRNA encoding a proline-rich protein that is destabilized by fungal elicitor. *Plant Cell* **6**, 135–145.

Zhang, S., Sheng, J., Liu, Y. and Mehdy, M. C. (1993). Fungal elicitor-induced bean proline-rich protein mRNA down-regulation is due to destabilization that is transcription and translation dependent. *Plant Cell* **5**, 1089–1099.

Zhao, B., Dahlbeck, D., Krasileva, K. V., Fong, R. W. and Staskawicz, B. J. (2011). Computational and biochemical analysis of the *Xanthomonas* effector AvrBs2 and its role in the modulation of *Xanthomonas* type three effector delivery. *PLoS Pathog.* **7**, E1002408.

Zuiderveen, G. H., Padder, B. A., Kamfwa, K., Song, Q. and Kelly, J. D. (2016). Genome-wide association study of anthracnose resistance in Andean beans (*Phaseolus vulgaris*). *PLoS ONE* **11**, E0156391.

Zumaquero, A., Macho, A. P., Rufián, J. S. and Beuzón, C. R. (2010). Analysis of the role of the type III effector inventory of *Pseudomonas syringae* pv. phaseolicola 1448a in interaction with the plant. *J. Bacteriol.* **192**, 4474–4488.

Appendices

APPENDIX 1. Development and maintenance of plant and pathogen genetic resources

Plant material

The SOA-BN × Edmund recombinant inbred population of 80 lines used in this work was initiated and developed up to generations F_3 and F_4 by Dr John Taylor and colleagues at Warwick Horticultural Research Institute (HRI) near Wellesbourne, UK. The author of this thesis subsequently advanced and curated the population up to F_7 and F_8 over four growing seasons from 2012 to 2015. All plant material used in this work was obtained or developed from the Phaseolus collection maintained at the University of Warwick Crop Centre, Wellesbourne, UK.

Bacterial isolates

Thirty-two isolates of *Psph* were selected for whole-genome sequencing in this research. The isolates were selected from a collection assembled by John Taylor and colleagues over two decades of collecting from 17 countries across four continents (Table A1.1), and maintained at the University of Warwick Crop Centre, Wellesbourne, UK. They include isolates from each of the nine described pathotypic races (Taylor *et al*., 1996a), with 11 of the selected isolates from race 6.

Table A1.1. Thirty-two isolates of *Pseudomonas syringae* pv. *phaseolicola* selected for whole-genome sequencing from a pathogenically and geographically diverse collection assembled by Dr John Taylor and colleagues

Isolate	Race	Host	Country and date of collection	Continent	Collector/ collection no.
1281A	1	*Phaseolus coccineus*	UK (1984)	Europe	J. D. Taylor
1816A	1	*Phaseolus lunatus*	Madagascar (1987)	Africa	D. M. Teverson
725A	1	*Phaseolus vulgaris*	UK (1973)	Europe	J. D. Taylor
1502A	2	*Phaseolus vulgaris*	Mauritius (1986)	Africa	Unknown
1678	2	*Phaseolus vulgaris*	Sweden (1956)	Europe	NCPPB 380
2698A	2	*Phaseolus vulgaris*	Zimbabwe (1990)	Africa	J. D. Taylor
882	2	*Phaseolus vulgaris*	USA (1975)	North America	J. D. Taylor
1301A	3	*Phaseolus vulgaris*	Tanzania (1984)	Africa	J. H. C. Davis
1567A	3	*Phaseolus vulgaris*	Burundi (1986)	Africa	D. M. Teverson
2475A	3	*Phaseolus vulgaris*	Colombia (1989)	South America	ICA
1302A	4	*Phaseolus vulgaris*	Rwanda (1984)	Africa	J. H. C. Davis
1334A	4	*Phaseolus vulgaris*	Uganda (1984)	Africa	T. N. Seegooba
2240A	4	*Phaseolus vulgaris*	Burundi (1988)	Africa	D. M. Teverson
11	5	*Phaseolus vulgaris*	Canada (1941)	North America	NCPPB 52
1516A	5	*Neonotonia wightii*	Tanzania (1986)	Africa	D. J. Allen
1010	6	*Phaseolus vulgaris*	USA (1979)	North America	M. L. Schuster
1294	6	*Phaseolus vulgaris*	Colombia (1984)	South America	M. A. Pastor-Corrales
1299A	6	*Phaseolus vulgaris*	Tanzania (1984)	Africa	J. H. C. Davis
1308	6*	*Phaseolus vulgaris*	Colombia (1984)	South America	M. A. Pastor-Corrales
1599A	6	*Phaseolus coccineus*	UK (1986)	Europe	J. D. Taylor
1715A	6	*Phaseolus vulgaris*	Ethiopia (1986)	Africa	D. J. Allen
1769A	6	*Phaseolus vulgaris*	Bulgaria (1987)	Europe	I. Poryazov
2109	6	*Phaseolus vulgaris*	Spain (1988)	Europe	C. Asensio-Vegas
2654A	6	*Phaseolus vulgaris*	Lesotho (1990)	Africa	J. D. Taylor
2693A	6	*Phaseolus vulgaris*	Zimbabwe (1990)	Africa	J. D. Taylor
716B	6	*Phaseolus vulgaris*	UK (1973)	Europe	J. D. Taylor
1314	7	*Phaseolus vulgaris*	Colombia (1984)	South America	CIAT
1354A	7	*Phaseolus vulgaris*	Kenya (1985)	Africa	D. J. Allen
1549A	7	*Phaseolus vulgaris*	UK (1986)	Europe	J. H. C. Davis
1645	8	*Dolichos* sp.	Tanzania (1964)	Africa	NCPPB 1647
2654C	8	*Phaseolus vulgaris*	Lesotho (1990)	Africa	J. D. Taylor
2732E	9	*Phaseolus vulgaris*	Colombia (1990)	South America	J. D. Taylor

The collection of bacterial isolates is maintained at the University of Warwick Crop Centre, Wellesbourne, United Kingdom.
ICA: Instituto de Ciencias Agrarias. CIAT: International Center for Tropical Agriculture.
*Isolate 1308 was originally designated race 1. Whole-genome sequencing revealed that this isolate lacks *hopF1*, which is present in all other race 1 isolates. This isolate was virulent on each of the host differential lines tested and was therefore re-classified as race 6.

APPENDIX 2. Determination of the inoculum concentration

To enable determination and adjustment of the inoculum concentration, a spectrophotometer was calibrated according to the number of viable colonies counted on plates following serial dilutions of an original bacterial suspension, using a method adapted from Vicente (2000). For example, *Pseudomonas syringae* pv. *phaseolicola* race 6 isolate 716B was cultured on King's Medium B (King *et al.*, 1954) for 24–48 h at 25 °C. Bacteria were harvested and suspended in 50 ml of sterile reverse-osmosis-fed deionised (RO-DI) water, thus obtaining the original suspension from which serial dilutions were made (Table A2.1).

Table A2.1. Serial dilutions used for the calibration of a spectrophotometer

Suspension	Dilution factor (×)	Original suspension (ml)	Buffer (ml)
A	0	10	0
B	1.25	8	2
C	1.67	6	4
D	2.5	4	6
E	5	2	8
F	10	1	9

The optical density (OD) at 640 nm of the six suspensions was measured with a spectrophotometer (Table A2.2). The spectrophotometer was calibrated with sterile RO-DI water. From each of the six suspensions (A to F), nine successive 10× dilutions (100 µl of the preceding suspension and 900 µl of RO-DI water) were made in 2-ml microcentrifuge tubes, thus obtaining the series X, X^{-1}, X^{-2}, … X^{-9}. An aliquot of 100 µl of each suspension in each series was dispensed and spread onto sterile 9-cm-diameter Petri dishes containing 20 ml of King's Medium B. Plates were incubated inverted at 25 °C for 48 h. Only plates with between 30 and 300 colonies were counted. The number of bacteria was given by the formula:

Colony-forming units (cfu) ml^{-1} = average number of colonies × dilution × 10

Table A2.2 shows the average number of viable colonies for each suspension as determined by the plate counting method.

Table A2.2. Optical density at 640 nm of six suspensions and the corresponding average number of viable colonies determined by the plate counting method

Suspension	Optical density	Number of viable colonies (cfu ml^{-1})
A	0.257	4.61×10^{11}
B	0.209	5.97×10^{10}
C	0.151	5.38×10^{8}
D	0.094	3.14×10^{8}
E	0.047	1.95×10^{8}
F	0.032	4.93×10^{7}

Colony-number data for the six suspensions were $\log_{10}$-transformed. The calibration line obtained using a simple linear regression (Figure A2.1) is:

$\text{Log}_{10}(\text{cfu ml}^{-1}) = 7.101 + 16.61$ OD (correlation coefficient = 0.96)

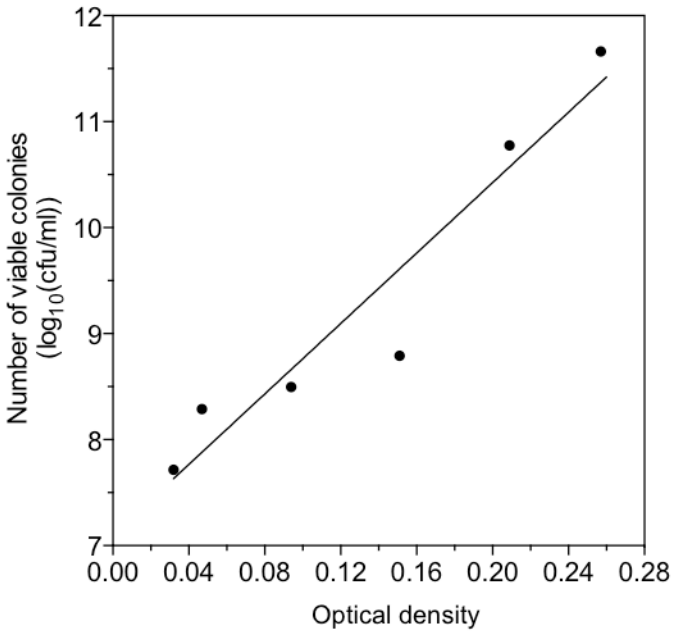


Figure A2.1. Calibration line showing the relationship between optical density (OD) and inoculum concentration ($\log_{10}$(cfu ml^{-1}) = 7.101 + 16.61 OD. R = 0.96).

The calibration line can be used to adjust the inoculum concentration. A consistent inoculum load of 10^8–10^9 cfu ml^{-1} can be obtained by suspending bacteria in sterile RO-DI water to give an OD of 0.10 at 640 nm.

APPENDIX 3. Protocol for the isolation of high-quality plant DNA

The following protocol is adapted from that developed by Afanador *et al*. (1993) for the extraction of DNA from young trifoliolate leaves of *Phaseolus vulgaris* plants.

CTAB extraction buffer composition (100-ml stock solution):

- 75.82 ml sterile high-purity water.
- 100 mM Tris-HCl, pH 8.0: 10 ml of 1-M stock solution.
- 20 mM EDTA, pH 8.0: 4 ml of 0.5-M stock solution.
- 1.4 M NaCl: Add 8.182 g NaCl and stir gently until complete dissolution.
- 2% CTAB: Add 2 g CTAB and mix gently until complete dissolution (can take time to dissolve; mix on a hot plate/magnetic stirrer).

TE_{sl} (super low TE) buffer composition (50-ml stock solution):

- 49.5 ml sterile DI water.
- 10 mM Tris, pH 7.8–8.0, 0.01 mM EDTA, pH 7.6–8: 500 µl of Tris 1-M stock solution + 1 µl of EDTA 0.5-M stock solution.

Dispense solutions through a sterile 0.2-µm syringe filter into a 50-ml Falcon tube to hold the working stock.

Protocol

1. Collect young trifoliolate leaves or leaf discs (100 mg) in 2-ml microcentrifuge tubes containing a 4-mm stainless steel grinding ball and place on ice. Samples can be processed as fresh, frozen or lyophilised tissue.
2. Keep mixer mill adapter set at −80 °C for 2 h and heat water bath to 65 °C.
3. In a 2-ml microcentrifuge tube containing a 4-mm stainless steel grinding ball, disrupt leaf tissue with the mixer mill for 1 min at 30 Hz, at ambient temperature if using lyophilised starting material (20 mg), or at −80 °C and on liquid nitrogen if using fresh or frozen starting material (100 mg). Swap the positions of the outermost and innermost tubes and repeat disruption.
4. Centrifuge samples at 13,000 rpm for 30 s.
5. Add 100 µl hot (65 °C) CTAB extraction buffer to each tube.

6. Homogenise each sample by vortexing vigorously.
7. Add 300 µl hot (65 °C) CTAB buffer and 6 µl RNase A (100 mg ml^{-1}).
8. Vortex vigorously to remove any tissue clumps.
9. Once samples are completely homogenised, allow tubes to incubate for 1 h at 65 ºC. Invert every 15 min during incubation.
10. Centrifuge at 13,000 rpm for 15 min.
11. Being careful to avoid the interface (white disc) between the aqueous and organic phases, transfer the supernatant to a new 1.5-ml Eppendorf LoBind® tube containing 400 µl chloroform:isoamyl alcohol (24:1). Discharge tip into container of water and rinse with water before discarding tip in labelled Disposafe bin.
12. Mix samples well by inverting 20–30 times to produce an emulsion.
13. Centrifuge at 13,000 rpm for 15 min.
14. Repeat above 3 steps once.
15. To precipitate, transfer the supernatant to a new 1.5-ml Eppendorf DNA LoBind® tube containing 280 µl isopropanol (0.7× volume).
16. Mix samples well by inverting 20–30 times and ensure that the precipitated DNA is submerged.
17. Place on ice for 5 min and centrifuge at 13,000 rpm for 15 min.
18. Carefully discard supernatant by tipping tube with the pellet on the upper side.
19. Clean and dry the pellet:
    - Wash with 500 µl cold (−20 °C) 70% EtOH by inverting 20–30 times and centrifuging at 13,000 rpm for 2 min.
    - Carefully discard the supernatant.
    - Wash with 500 µl cold (−20 °C) 100% EtOH by inverting 20–30 times and centrifuging at 13,000 rpm for 2 min.
    - Discard supernatant, pipette excess, and open tubes for 5–10 min to dry pellet.
20. Re-suspend the pellet in 100 µl TE_{sl} buffer—re-dissolve by rinsing the walls to recover all DNA—and mix well by inverting and flicking.
21. Incubate at ambient temperature overnight, quantify, and freeze at −20 °C.

APPENDIX 4. Agarose gel electrophoresis of genomic DNA samples for genotyping-by-sequencing

The quality of genomic DNA extracted from plants of the SOA-BN × Edmund and JDT inbred mapping populations was confirmed by agarose gel electrophoresis (Figure A4.1) and trial digestion with EcoRI of a subset of samples (Figure A4.2).

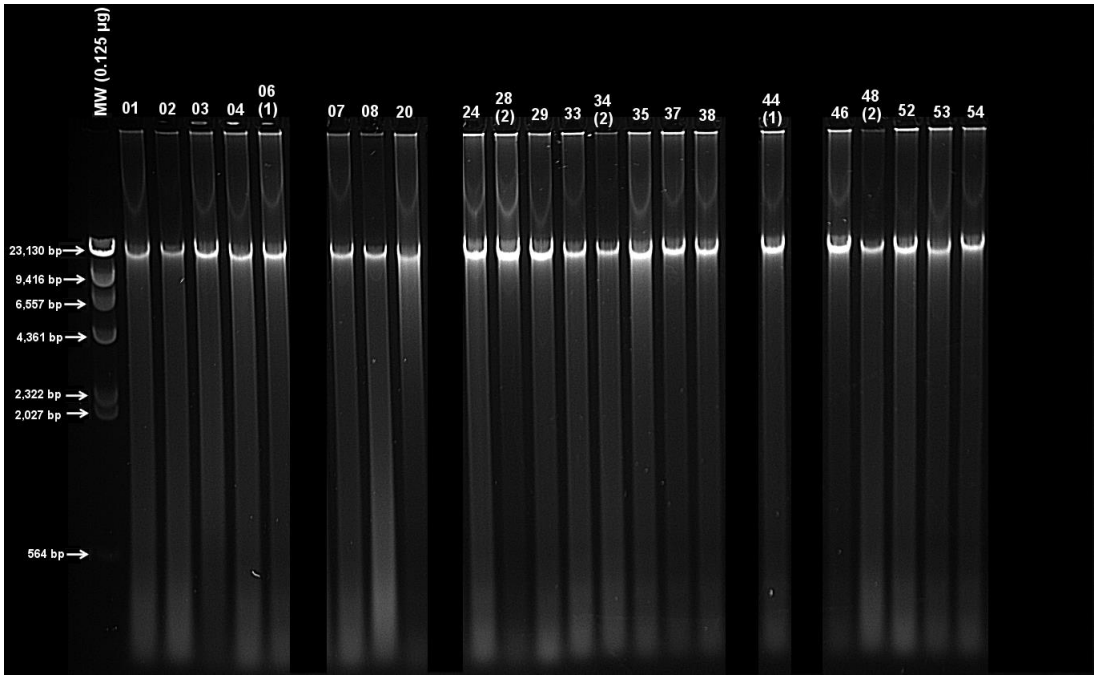


Figure A4.1. Genomic DNA of SOA-BN × Edmund (SE) and JDT inbred lines electrophoresed and visualised on 0.8% agarose gels. MW: HindIII digest of Lamda DNA supplied at 0.5 µg μl^{-1} (Gibco-BRL). Ed1/Ed2 and Br1/Br2: replicate samples for parental lines Edmund and SOA-BN. Ca2: Capulet, an improved SE inbred combining halo blight resistance with reported cold tolerance and earliness of maturity (Dodd and Taylor, 1991, unpublished data). Line numbers without decimal places denote SE inbreds that were at F_6 at the time of sampling, while those with decimal places were at F_7. JT-###: JDT inbred line samples pooled into halo blight-resistant and susceptible bulks for bulked segregant analysis. A number of breeding lines of interest (which did not contribute to the pedigree of the SE population) were included in the plate for genotyping-by-sequencing; corresponding DNA samples are denoted by two-letter abbreviations.

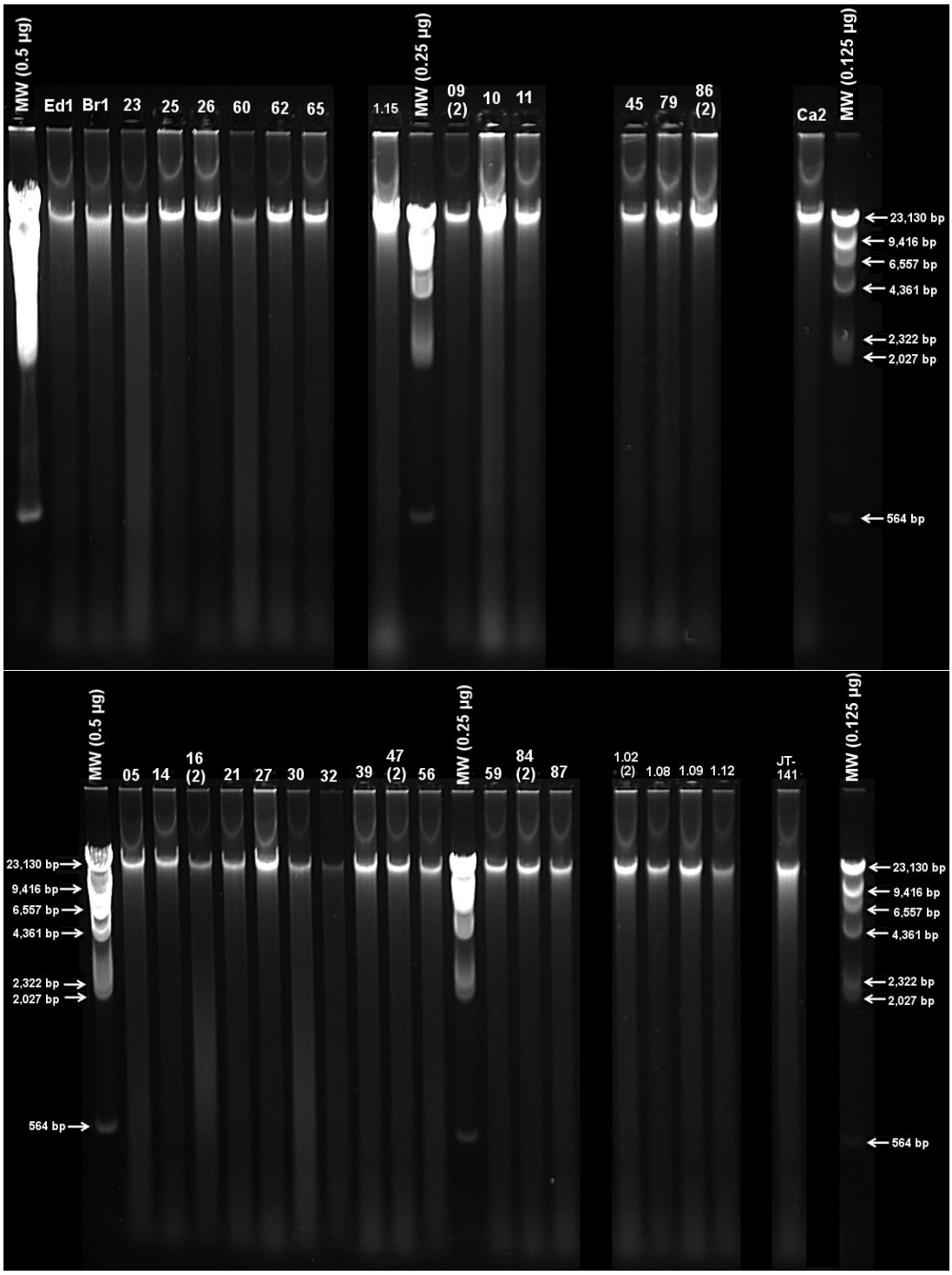


Figure A4.1. Continued.

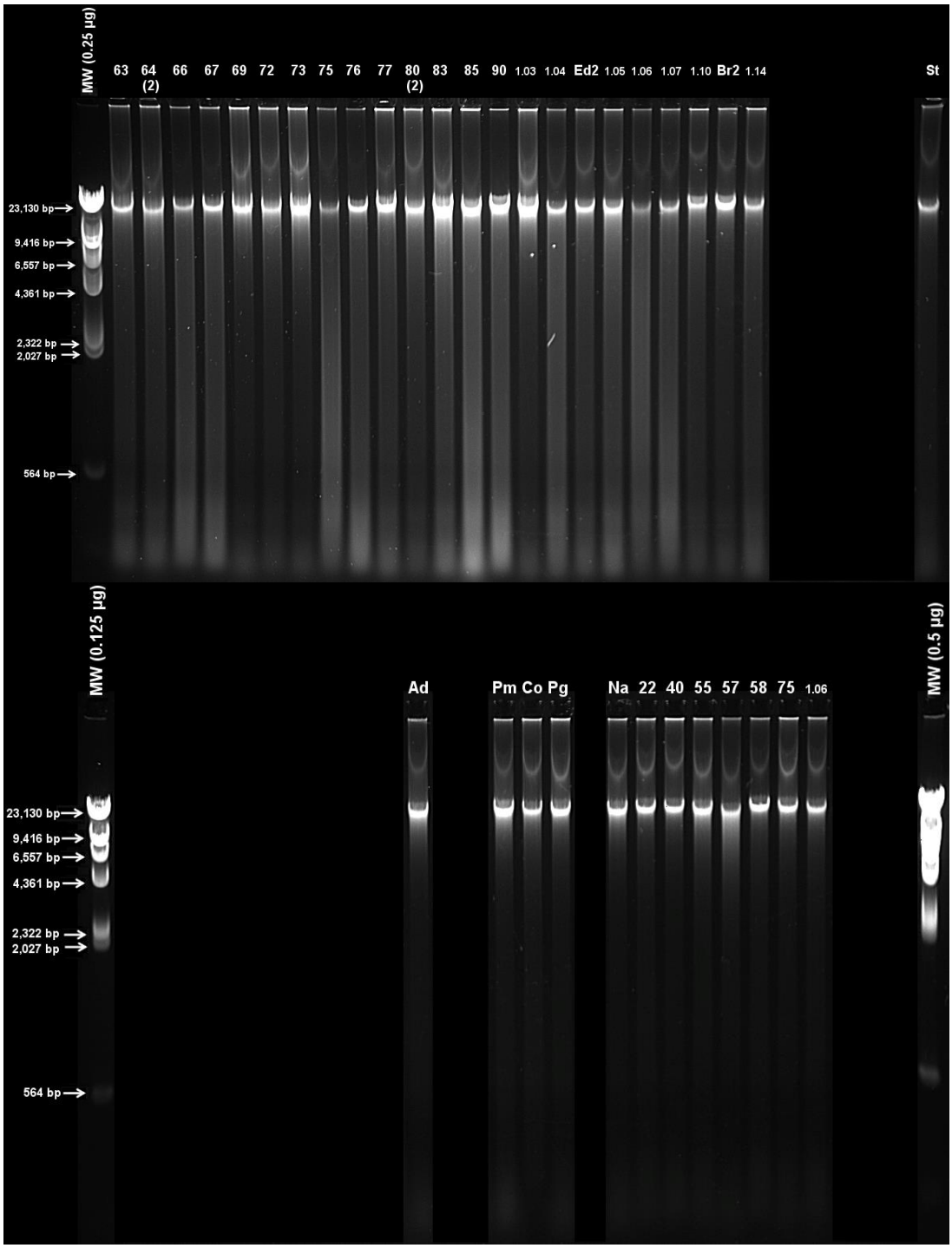


Figure A4.1. Continued.

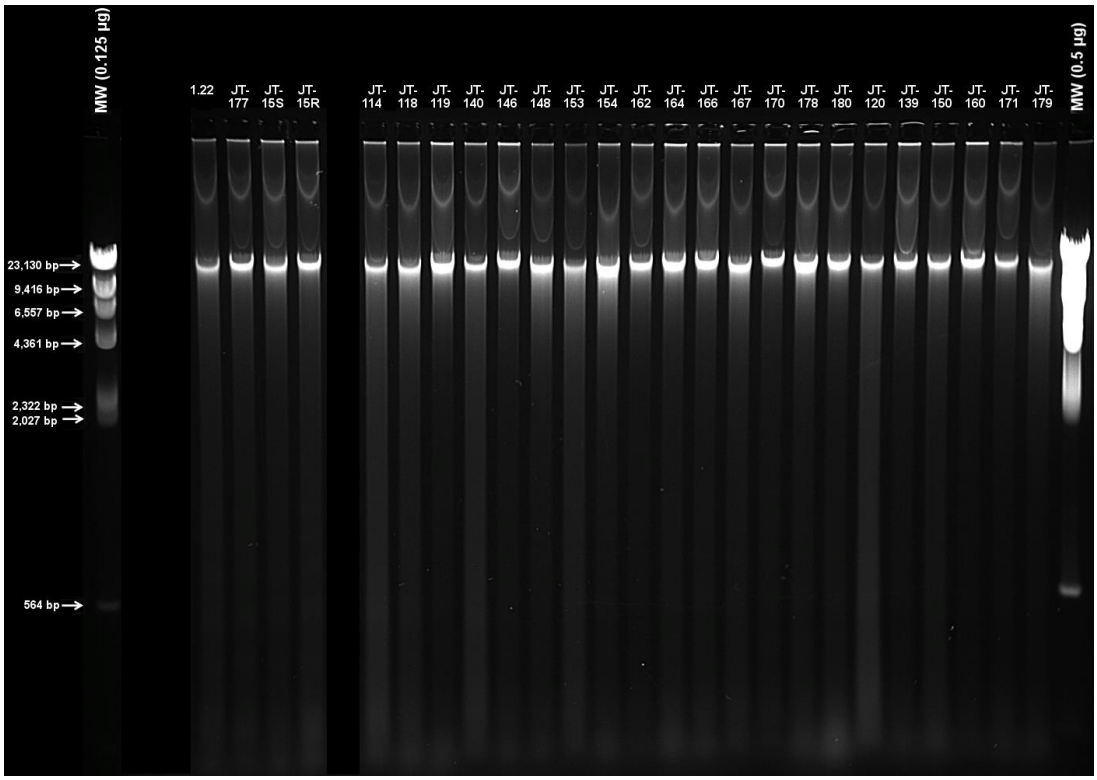


Figure A4.1. Continued.

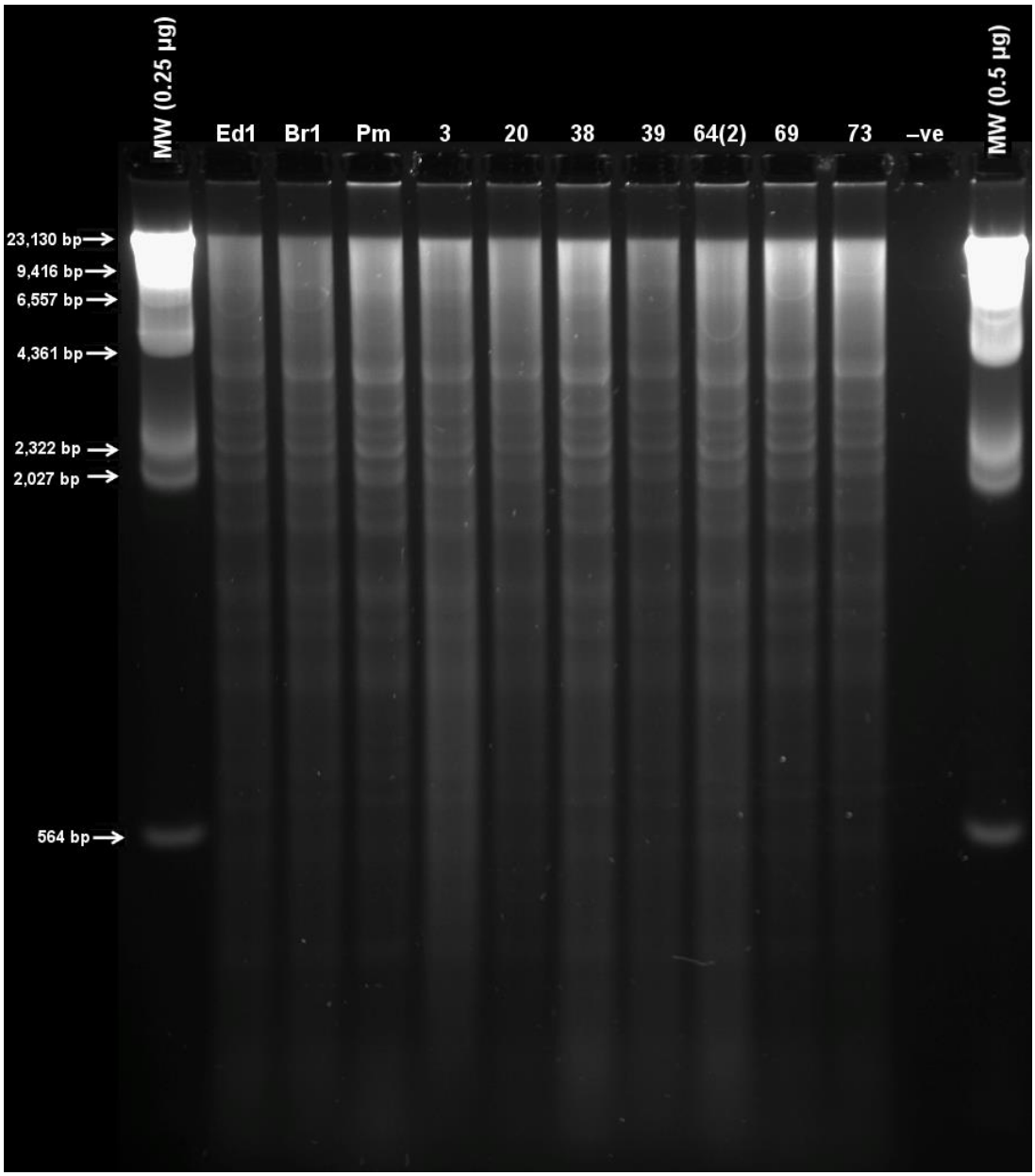


Figure A4.2. Trial digestion of genomic DNA of Edmund (Ed1) and SOA-BN (Br1) and of eight randomly selected lines with 5 U per sample of the restriction enzyme EcoRI (Roche, Basel, Switzerland) for 2 h at 37 °C followed by 20 min at 65 °C. Digested samples were electrophoresed and visualised on a 1% agarose gel to confirm suitability for genotyping-by-sequencing.

APPENDIX 5. Restricted maximum likelihood analysis to estimate fixed- and random-effects variance components of plant phenotypic data obtained for reactions to *Pseudomonas syringae* pv. *phaseolicola*

Table A5.1. Summary of fixed- and random-effects variance components of phenotypic data obtained for reactions to *Pseudomonas syringae* pv. *phaseolicola* (*Psph*) races 6 and 1 in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population estimated by restricted maximum likelihood analysis

***Psph* race 6**						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	75	106.32	115.787	−53.746	13.033	< 2.2E−16***
Random-effects term	χ^2	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	0	1	1			
Box	0.295	1	0.587			
Tray	0.302	1	0.583			
Position	0	1	1			
***Psph* race 1**						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	72	111.9	127.06	−54.022	10.454	< 2.2E−16***
Random-effects term	χ^2	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	0.514	1	0.473			
Box	0.092	1	0.761			
Tray	0	1	1			
Position	2.526	1	0.113			

d.f.: degrees of freedom. AIC: Akaike information criterion.
***: significant at the 0.001 probability level.

APPENDIX 6. Type III secretion system (T3SS) protein database compilation

The following keyword search criteria were used to compile a database of 18,080 T3SS protein sequences from plant and animal pathogens stored in the NCBI RefSeq non-redundant protein database on 3rd April 2016:

```
srcdb_refseq[prop] AND (
 (bacteria[ORGN] AND (type III effector[Protein Name] OR (type[All Fields] AND III[All Fields] AND effector[All Fields])))
 OR (bacteria[ORGN] AND (T3SS[All Fields] AND effector[All Fields]))
 OR (bacteria[ORGN] AND (virulence factor family protein[Protein Name] AND (type[All Fields] AND III[All Fields])))
 OR (bacteria[ORGN] AND (virulence protein[Protein Name] AND (type[All Fields] AND III[All Fields])))
 OR (bacteria[ORGN] AND (hop*[Protein Name]) NOT (hopanoid*[Protein Name]) NOT (hopene*[Protein Name]))
 OR (bacteria[ORGN] AND (hrp*[Protein Name]))
 OR (bacteria[ORGN] AND (avr*[Protein Name]))
 OR (bacteria[ORGN] AND (T3SS[All Fields] AND chaperone[All Fields]))
 OR (bacteria[ORGN] AND (T3SS[All Fields] AND helper[All Fields]))
 OR (bacteria[ORGN] AND (type[All Fields] AND III[All Fields] AND chaperone [All Fields]))
 OR (bacteria[ORGN] AND (type[All Fields] AND III[All Fields] AND helper [All Fields]))
 OR (Pseudomonas syringae[ORGN] AND (T3SS[All Fields]))
 OR (Pseudomonas savastanoi[ORGN] AND (T3SS[All Fields]))
 OR (Pseudomonas syringae[ORGN] AND (type[All Fields] AND III[All Fields] secret*[All Fields]))
 OR (Pseudomonas savastanoi[ORGN] AND (type[All Fields] AND III[All Fields] secret*[All Fields]))
)
```

APPENDIX 7. Type III secretion system (T3SS) protein hits in the genomes of 34 *Pseudomonas syringae* pv. *phaseolicola* isolates and five control pathovars of *P. syringae*

Table A7.1. Predicted presence and absence of type III secretion system (T3SS) proteins in 34 pathogenically and geographically diverse isolates of *Pseudomonas syringae* pv. *phaseolicola* (*Psph*) and control isolates of five pathovars of *P. syringae*

	Psph race																																							
	1				2				3			4			5		6												7			8		9	Other pathovars					
T3SS protein hit	1281A_1	1644R_1	1816A_1	725A_1	1502A_2	1678_2	2698A_2	882_2	1301A_3	1567A_3	2475A_3	1302A_4	1334A_4	2240A_4	11_5	1516A_5	1010_6	1294_6	1299A_6	1308_6	1448A_6	1599A_6	1715A_6	1769A_6	2109_6	2654A_6	2693A_6	716B_6	1314_7	1354A_7	1549A_7	1645_8	2654C_8	2732E_9	*glycinea*	*pisi*	*maculicola*	*tomato*	*syringae*
DNA_polymerase	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
DnaK	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EscC_YscC_HrcC	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EscD_YscD_HrpQ	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EscU_YscU_HrcU	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
EscV_YscV_HrcV	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
fimbrial_protein_	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
fimbrial_protein_	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
GroEL	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HopAJ2	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HopAN1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HopI1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
hopM1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
hopX1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HrpK	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
Insulin-cleaving_r	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
membrane_prote	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
secretin_WP_057	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
SepL_TyeA_HrpJ	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
type_III_secretior	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
ATP_synthase_or	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
HopAK1	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
hopR1	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	–
hrpH	+	–	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+

Table A7.1. Continued

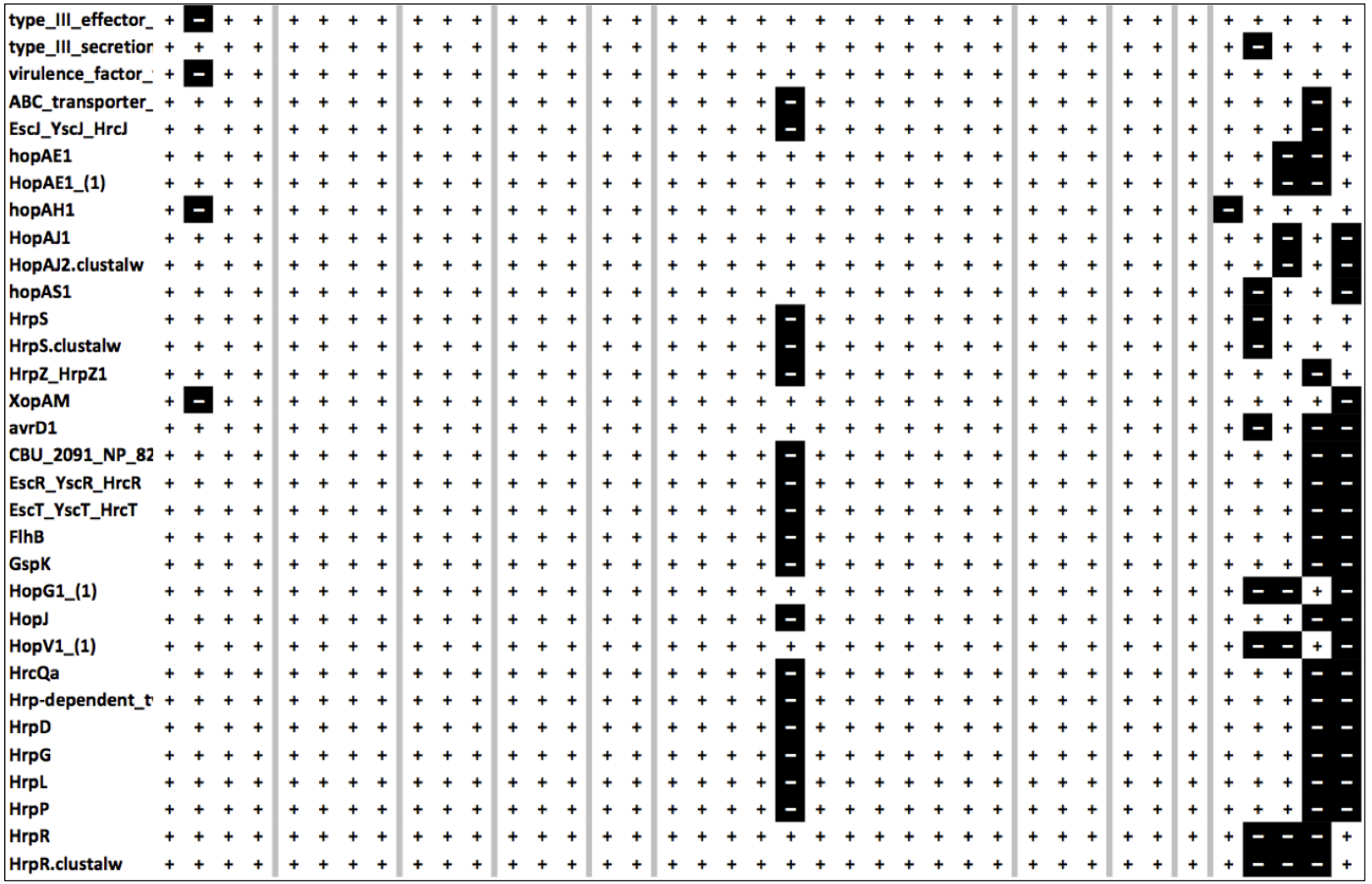


Table A7.1. Continued

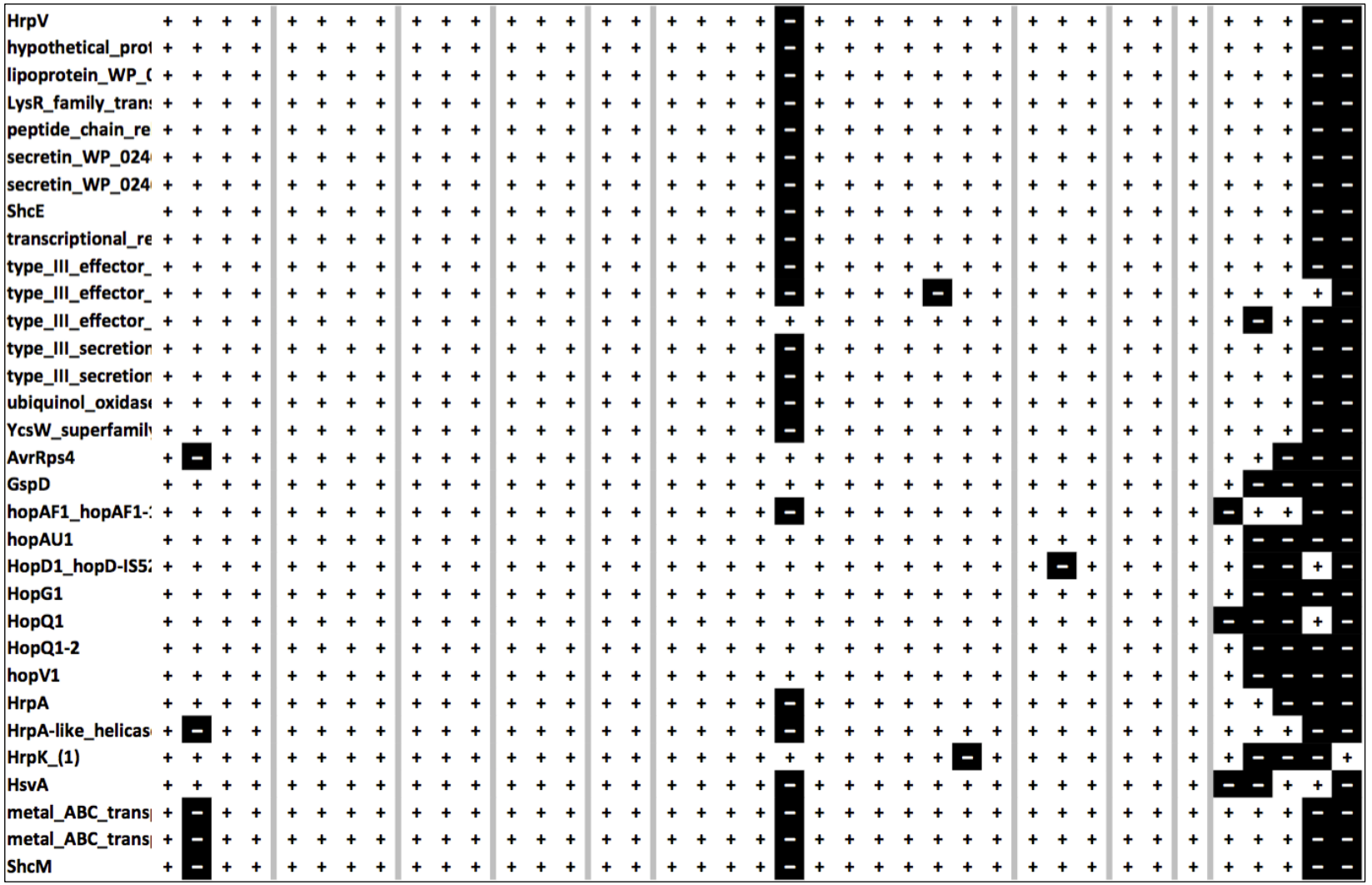


Table A7.1. Continued

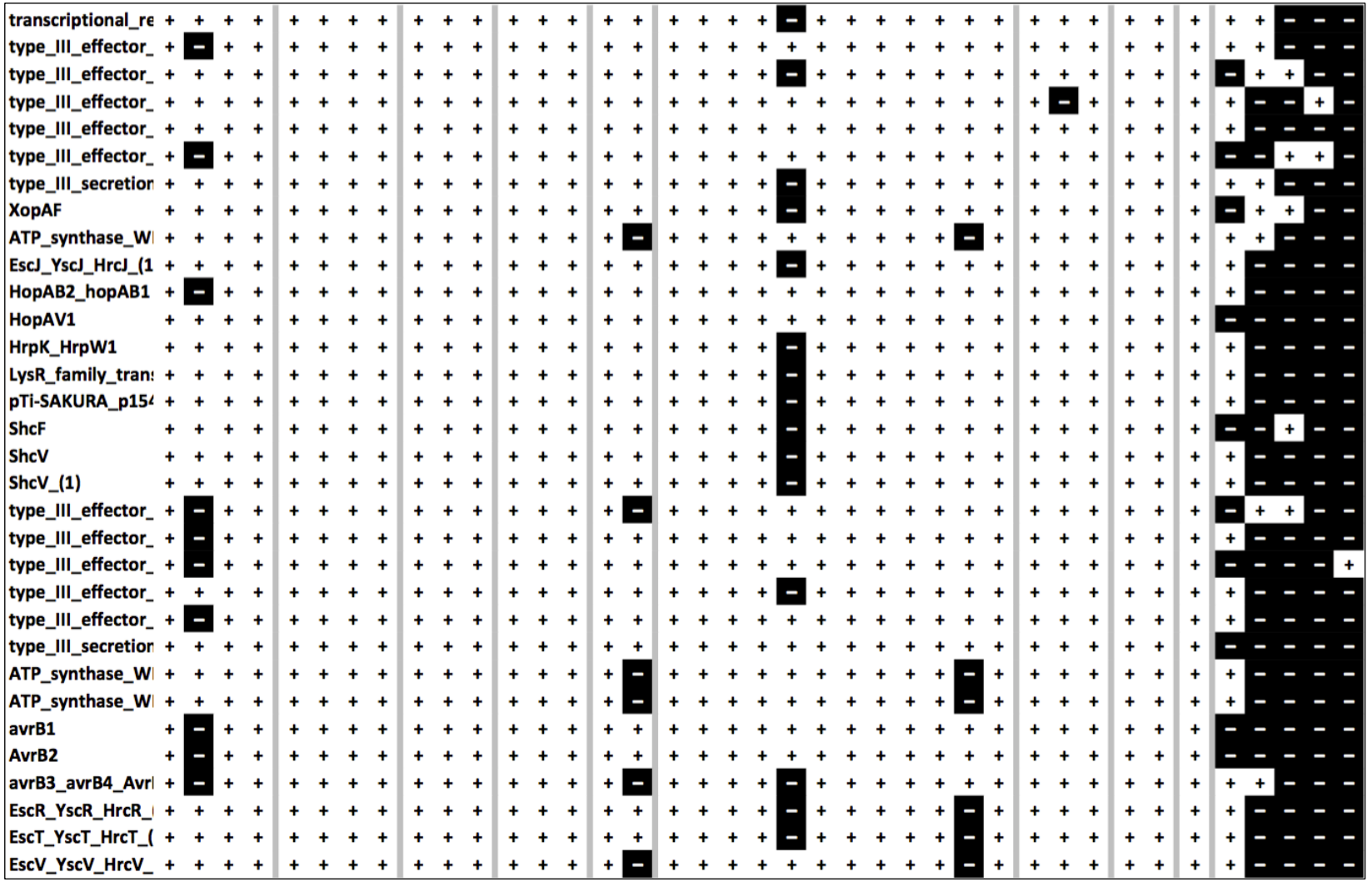


Table A7.1. Continued

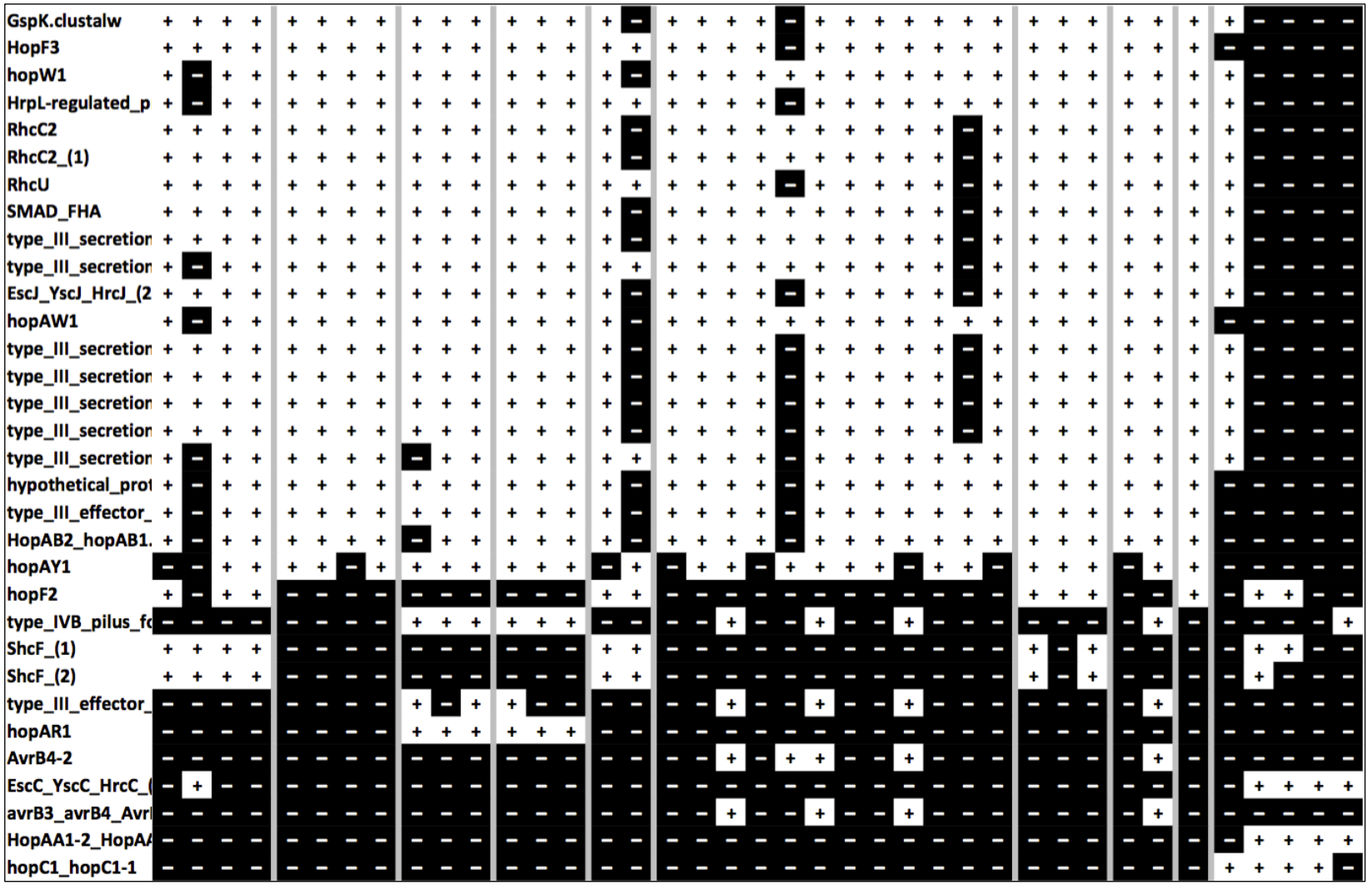


Table A7.1. Continued

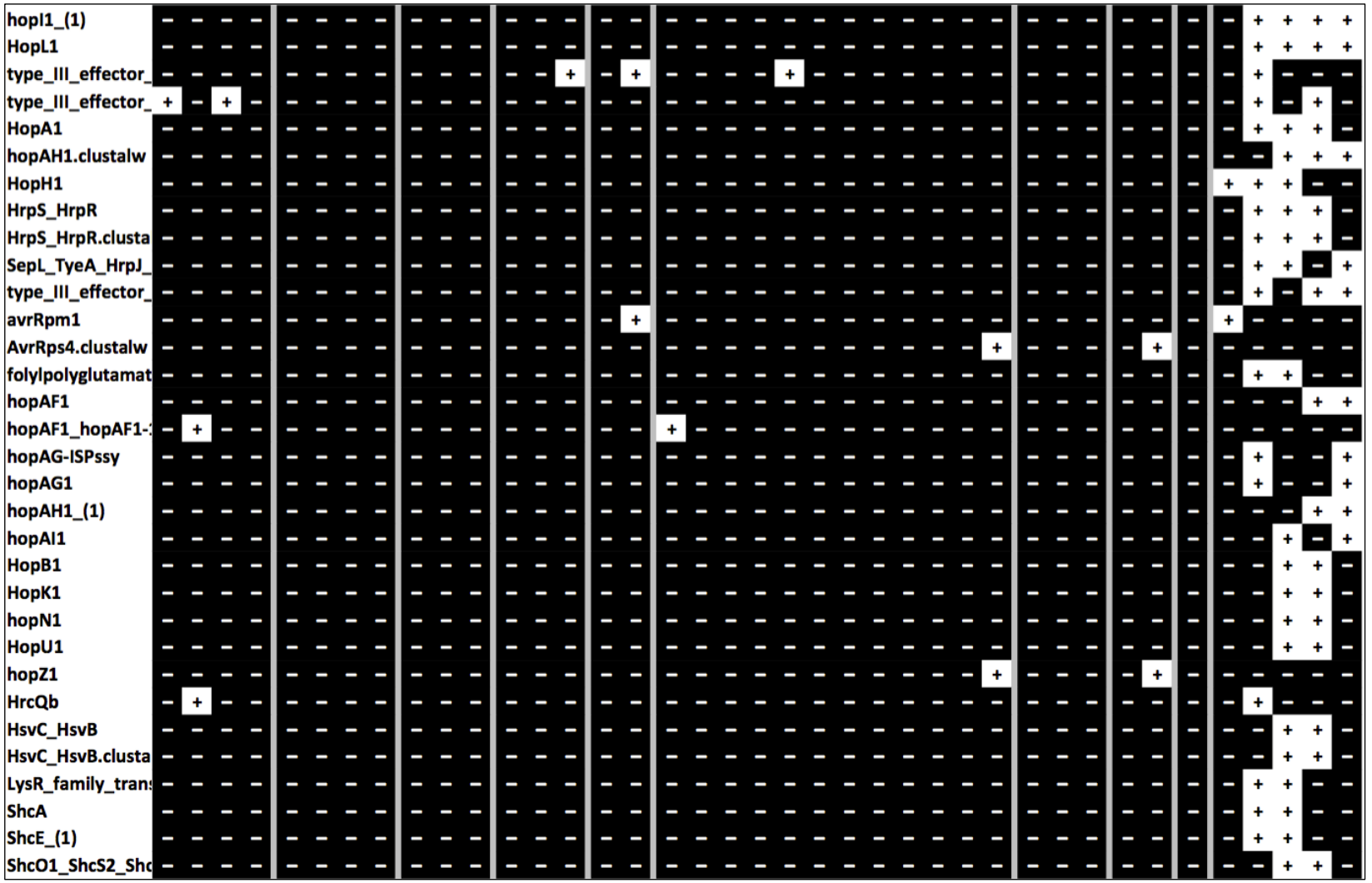


Table A7.1. Continued

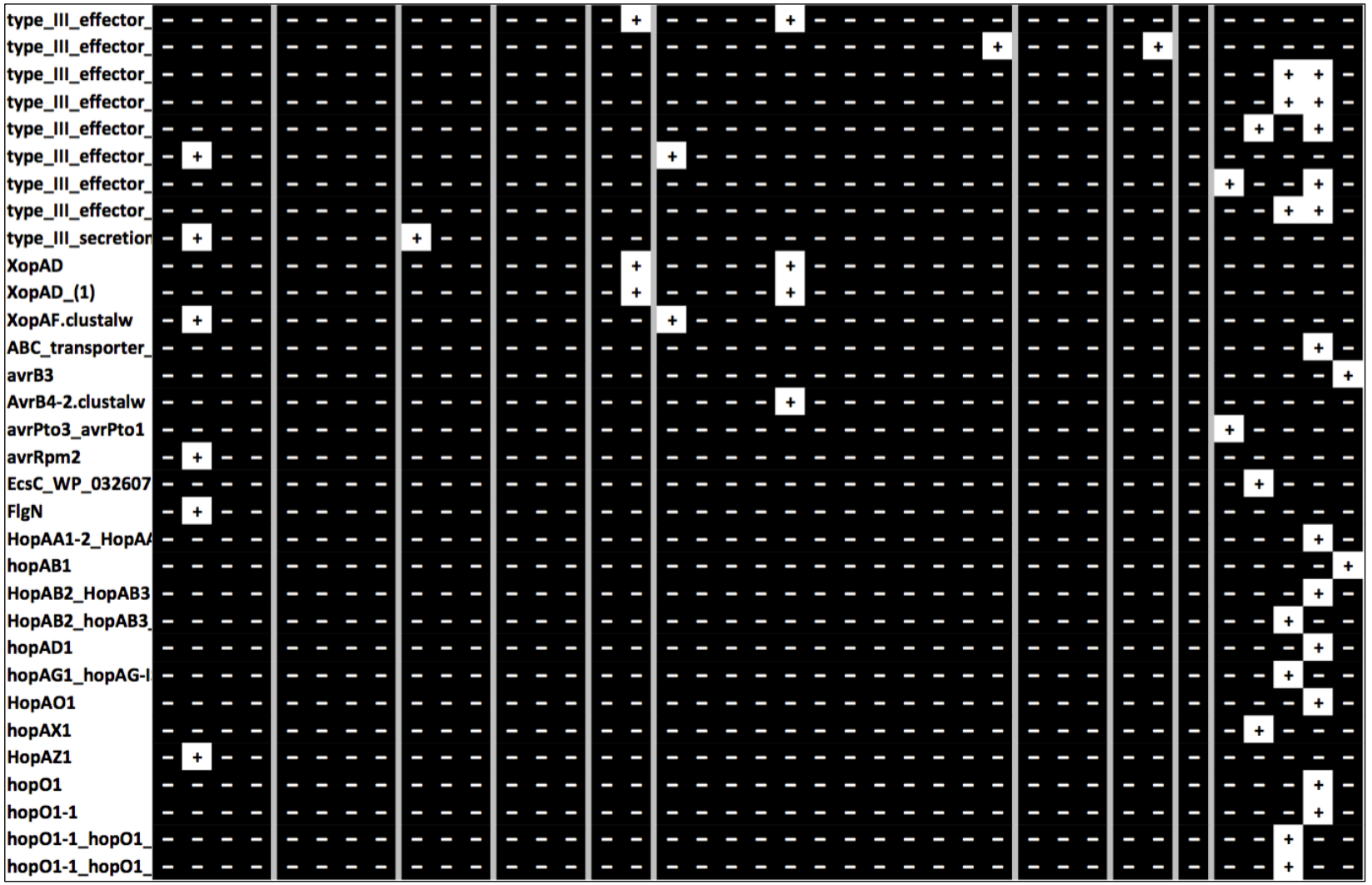


Table A7.1. Continued

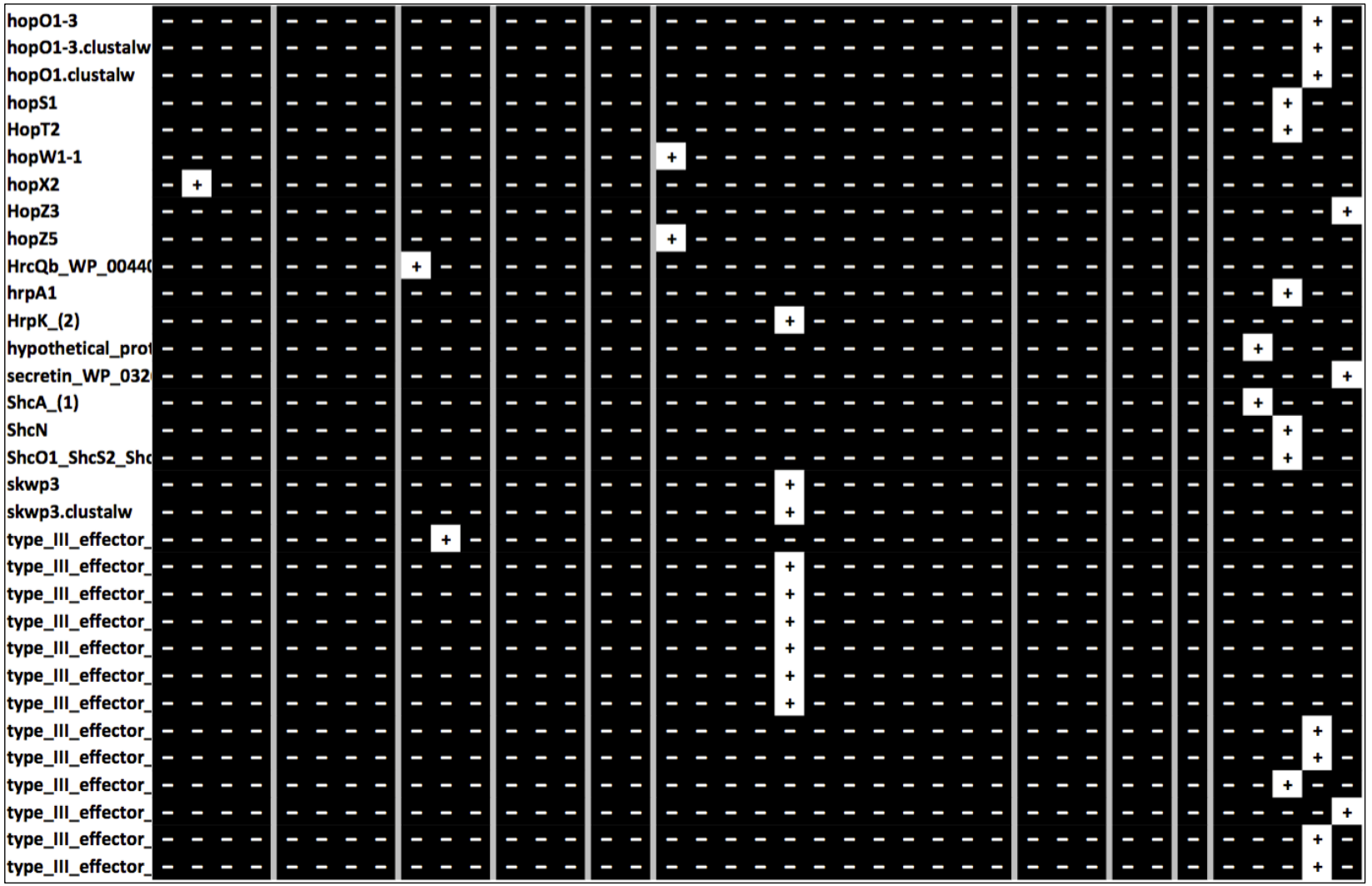
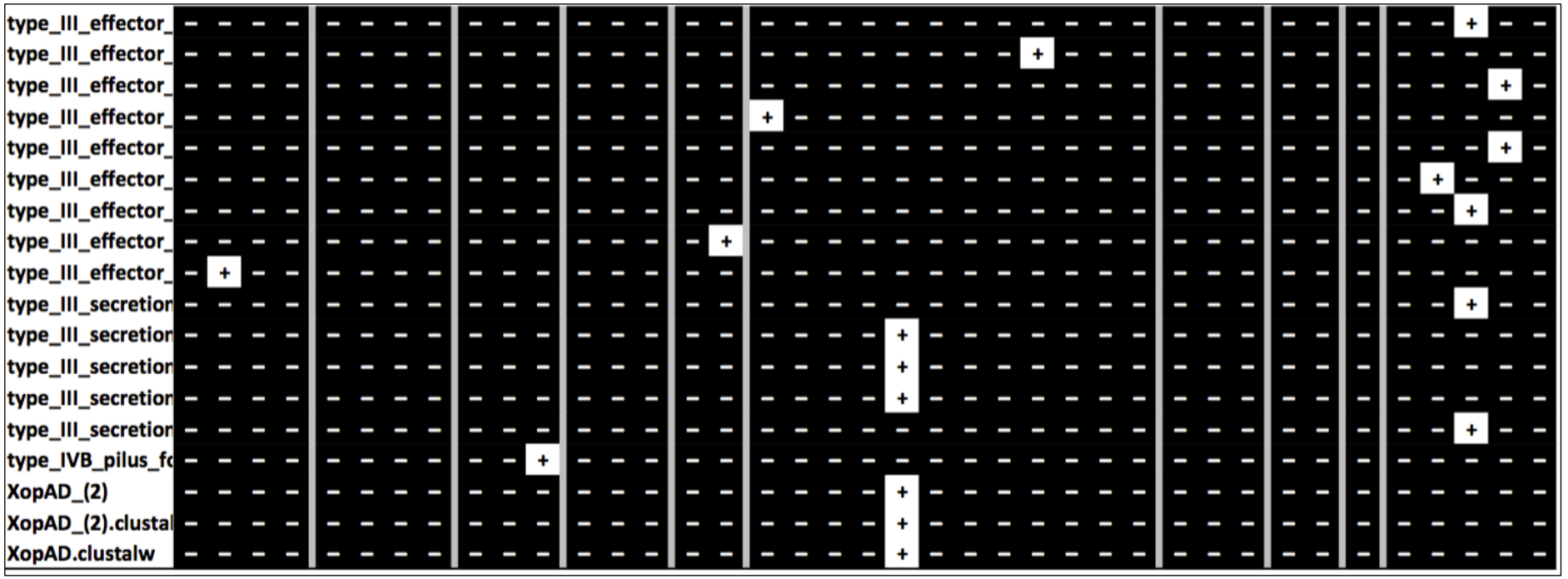


Table A7.1. Continued



+ (white boxes): homologue present. – (black boxes): homologue not found.
Previously assembled genomes for *Psph* isolates 1448A and 1644R were analysed (Joardar *et al*., 2005; Baltrus *et al*., 2012). Previously assembled genomes for *P. syringae* pathovars *glycinea* strain race 4 (NCBI BioSample accession: SAMN02471825), *maculicola* strain CFBP 1657 (SAMN03328955), *pisi* strain PP1 (SAMN02471574), *syringae* strain B728A (SAMN02604347) and *tomato* strain DC3000 (SAMN02604017) were included for identification and comparison of sequences with homology to T3SS proteins.

APPENDIX 8. Restricted maximum likelihood analysis to estimate fixed- and random-effects variance components of plant phenotypic data obtained for seedling root and shoot traits

Table A8.1. Summary of fixed- and random-effects variance components of phenotypic data obtained for seedling root and shoot traits in the *Phaseolus vulgaris* SOA-BN × Edmund recombinant inbred population estimated by restricted maximum likelihood analysis

Taproot diameter						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	72	15.746	23.635	−450.4	3.742	1.93E−11***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	1.753	1	0.185			
Row	0	1	1			
Column	0.298	1	0.585			
Range of basal root growth angles						
Fixed-effects term	**Sum of squares**	**Mean square**	**Num. d.f.**	**Den. d.f.**	***F***	**Probability (*F*)**
Genotype	10209.4	141.797	72	116.09	2.952	9.97E−08***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	0	1	1			
Row	3.237	1	0.072 (kept)			
Column	0	1	1			
Maximum basal root growth angle						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	72	7002.8	12404.9	846.18	2.395	6.77E−06***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	0.914	1	0.339			
Row	2.477	1	0.116			
Column	0	1	1			
Minimum basal root growth angle						
Fixed-effects term	**Sum of squares**	**Mean square**	**Num. d.f.**	**Den. d.f.**	***F***	**Probability (*F*)**
Genotype	1356.71	18.843	72	111.84	2.047	3.24E−04***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	0	1	1			
Row	1.1E−13	1	1.4E−10*** (kept)			
Column	18.874	1	1			

Table A8.1. Continued

Basal root number						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	72	501.36	756.44	270.55	3.685	3.22E−11***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	1.1E−13	1	1			
Row	0.231	1	0.631			
Column	1.097	1	0.295			
Mean basal root whorl number per file						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	72	31.436	47.335	−305.9	3.707	2.64E−11***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	2.8E−14	1	1			
Row	0.028	1	0.868			
Column	0.627	1	0.429			
Root dry weight						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	72	2680297	4517032	2025	2.615	9.95E−07***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	4.5E−13	1	1			
Row	1.3E−04	1	0.991			
Column	0	1	1			
Shoot dry weight						
Fixed-effects term	**Sum of squares**	**Mean square**	**Num. d.f.**	**Den. d.f.**	***F***	**Probability (*F*)**
Genotype	1.4E+06	19388.43	72	113.84	3.677	2.78E−10***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	0	1	1			
Row	7.117	1	0.008** (kept)			
Column	0.199	1	0.655			
Shoot diameter						
Fixed-effects term	**d.f.**	**Sum of squares**	**Residual sum of squares**	**AIC**	***F***	**Probability (*F*)**
Genotype	72	27.078	32.695	−382.9	9.038	< 2.2E−16***
Random-effects term	**χ^2**	**d.f. (χ^2)**	**Probability (χ^2)**			
Block	1.680	1	0.195			
Row	1.264	1	0.261			
Column	0	1	1			

Num./Den. d.f.: numerator/denominator degrees of freedom. AIC: Akaike information criterion.
*,**,***: significant at the 0.05, 0.01 and 0.001 probability levels.

APPENDIX 9. List of candidate genes with previously described functions in physiological resilience, morphological and developmental characteristics in common bean and other legumes

Table A9.1. Candidate sequences associated with drought response, low-temperature response, earliness of maturity, nutrient uptake and use, plant architecture, and seed morphological and nutritional traits

Species	Candidate	Accession no.*	Homologue	Protein family	Location	Description (taken or adapted from that provided at https://www.ncbi.nlm.nih.gov)	Function	Source
Drought response								
Phaseolus vulgaris	*Dreb2A* (Phvul.002G254500.1)	CV535836.1	*Glycine max*	AP2/EREBP transcription factor	Pv02: 42081187..42081870	Dehydration-responsive element binding (DREB) protein-encoding gene; stress-regulated transcription factor. *P. vulgaris* nodule EST library, cDNA	Abscisic acid (ABA)-independent responses to drought stress	Cortes *et al*., 2012a: 1072; Ramirez *et al*., 2005
P. vulgaris	*Dreb2B* (Phvul.001G251200.1)	BQ481823.1	*G. max; Populus trichocarpa*	AP2/EREBP transcription factor	Pv01: 50886928..50887671	Dehydration-responsive element binding (DREB) protein-encoding gene; stress-regulated transcription factor. *P. vulgaris* leaf EST library, cDNA	ABA-independent responses to drought stress	Cortes *et al*., 2012a: 1072
P. vulgaris	*Asr1* (Phvul.007G080400.1)	PHAVU_007G080400g	*G. max; Prunus armeniaca*	ABA-, stress-, ripening-induced (ASR) transcription factor	Pv07: 7715980..7717710 (– strand)	*P. vulgaris* hypothetical protein similar to ASR protein 1	ABA-dependent stress responses	Cortes *et al*., 2012b: 11; Schmutz *et al*., 2014
P. vulgaris	*Asr1* (region 1)	TC42828 (TIGR Gene Indices, TGI)	*G. max; P. armeniaca*	ASR transcription factor	Pv07: 7716075..7717561	*P. vulgaris* EST similar to ASR protein 1 (region 1)	ABA-dependent stress responses	Cortes *et al*., 2012b: 11
P. vulgaris	*Asr1* (region 2)	TC42828 (TIGR Gene Indices, TGI)	*G. max; P. armeniaca*	ASR transcription factor	Pv07: 7716075..7717039	*P. vulgaris* EST similar to ASR protein 1 (region 2)	ABA-dependent stress responses	Cortes *et al*., 2012b: 11
Phaseolus coccineus	*Asr2* (Phvul.004G094600.1)	CA910244.1	*P. vulgaris*	ASR transcription factor	Pv04: 25700400..25700740	Suspensor region TriplEx2 *P. coccineus* cDNA 5´ similar to ASR protein 2, mRNA sequence	ABA-dependent stress responses	Cortes *et al*., 2012b: 11
P. vulgaris	*PvLEA3* (*PvR16*) (Phvul.007G225200.1)	DQ196430.1	*G. max*	Group 3 late embryogenesis abundant (LEA3)	Pv07: 46501427..46504143	Group 3 late embryogenesis abundant protein (LEA3) gene, complete CDS. Corresponding cDNA (DT661620.1) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence (below)	Stress protection	Barrera-Figueroa *et al*., 2007: 371–374

Table A9.1. Continued

P. vulgaris	*PvR16 (PvLEA3)*	DT661620.1	*G. max*	Group 3 late embryogenesis abundant (LEA3; high identity)	Pv07: 46501497. .46501850	cDNA clone similar to Group 3 LEA protein (high identity; above) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Stress protection	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvR25*	DT661621.1	N/A	Unknown	Pv03: 30380695. .30380814	cDNA up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Unknown	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvR33* (Phvul.007G1 96900.1)	DT661622.1	N/A	Unknown	Pv07: 43675705. .43675839	cDNA up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Unknown	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvR123* (Phvul.009G2 45900.1)	DT661623.1	*Arabidopsis thaliana*	Periplasmic cytochrome C-related (moderate identity)	Pv09: 35946250. .35946408	cDNA clone similar to periplasmic cytochrome C-related (moderate identity) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Unknown	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvR160* (Phvul.008G1 08700.1)	DT661624.1	*G. max*	Acetyl-CoA carboxylase (high identity)	Pv08: 12634824. .12634910	cDNA clone similar to acetyl-CoA carboxylase (high identity) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Cellular metabolism	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvR168*	DT661625.1	N/A	Unknown	Pv02: 2091465.. 2091543	cDNA up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Unknown	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvR169* (Phvul.011G1 29100.1)	DT661626.1	*A. thaliana*	Unknown	Pv11: 28553695. .28553777	cDNA up-regulated in the roots of P. vulgaris cv. Pinto Villa under early drought stress, mRNA sequence	Unknown	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*Pv188*	DT661627.1	*Phaseolus acutifolius*	Heat shock protein Hsp70 (moderate identity)	Pv03: 35968527. .35968618	cDNA clone similar to Hsp70 (moderate identity) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Stress protection	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*Pv222* (Phvul.002G0 18700.1)	DT661628.1	*P. armeniaca*	Flavoprotein; zeaxanthin epoxidase precursor (low identity)	Pv02: 2051735.. 2051832	cDNA clone similar to zeaxanthin epoxidase precursor (low identity) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	ABA biosynthesis	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvR247* (Phvul.003G1 64200.1)	DT661629.1	*Oryza sativa*	Glycosylation enzyme (low identity)	Pv03: 37279416. .37279562	cDNA clone similar to glycosylation enzyme (low identity) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Protein glycosylation	Barrera-Figueroa *et al*., 2007: 371–374

Table A9.1. Continued

P. vulgaris	*PvR250* (Phvul.002G118400.1 / Phvul.002G118400.2)	DT661630.1	*P. vulgaris*	Leucine-rich repeat (LRR) (moderate identity)	Pv02: 23963868. .23963920	cDNA clone similar to *P. vulgaris* leucine-rich repeat (LRR) protein (CB539147; moderate identity) up-regulated in the roots of *P. vulgaris* cv. Pinto Villa under early drought stress, mRNA sequence	Unknown	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvL2* (Phvul.010G139300.1)	EE253606.1	*Cicer arietinum*	Beta-galactosidase (glycoside hydrolase; high identity)	Pv10: 41140049. .41140218	cDNA clone similar to β-galactosidase (high identity) up-regulated in the leaves of *P. vulgaris* cv. Pinto Villa under intermediate drought stress, mRNA sequence	Carbohydrate metabolism	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	*PvL185* (Phvul.002G090900.1)	EE253622.1	*Solanum tuberosum*	Translation initiation factor eIF-5A (high identity)	Pv02: 15370556. .15370674	cDNA clone similar to translation initiation factor eIF-5A (high identity) up-regulated in the leaves of *P. vulgaris* cv. Pinto Villa under intermediate drought stress, mRNA sequence	Protein synthesis	Barrera-Figueroa *et al*., 2007: 371–374
P. vulgaris	Phvul.003G044200.1 (BSNP28) (Restriction enzymes: BstNI or BstUI)	PHAVU_003G044200g	*Phaseolus lunatus*	*S*-adenosylmethionine decarboxylase (provisional)	Pv03: 5011354.. 5014384 (+ strand)	Enzyme contributing to transcriptional and post-transcriptional regulation of gene expression, with function in plant sensing of changes in the abiotic environment, including increased water deficit stress		Galeano *et al*., 2009a; Hu *et al*., 2005; Micheletto *et al*., 2007
Low-temperature response								
P. vulgaris	HSP83 (Phvul.008G281300.1)	FD791776.1	*Medicago sativa*	Heat shock protein 83 (putative)	Pv08: 58751801. .58752358	cDNA similar to heat shock protein 83 differentially expressed in total aerial vegetative tissue of *P. vulgaris* cv. ICA Pijao under cold stress, mRNA sequence	Stress protection	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
P. vulgaris	GABAr	FD795692.1	*M. sativa*	Gamma-aminobutyric acid A receptor (GABAr; putative)	Pv01: 31973886. .31974083	cDNA similar to a GABAr, subunit epsilon differentially expressed in total aerial vegetative tissue of *P. vulgaris* cv. ICA Pijao under cold stress, mRNA sequence	Signalling	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
P. vulgaris	ABAp (Phvul.010G016600.1)	FD797215.1	*M. sativa*	ABA-activated protein kinase (putative)	Pv10: 2655521.. 2655791	cDNA similar to an ABA-activated protein kinase differentially expressed in total aerial vegetative tissue of *P. vulgaris* cv. ICA Pijao under cold stress, mRNA sequence	ABA-dependent stress responses	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
P. vulgaris	CYP72A14 (Phvul.005G063200.1)	FD788569.1	*M. sativa*	Cytochrome P450 (CYP450; putative)	Pv05: 9098676.. 9099159	cDNA similar to a cytochrome P450 differentially expressed in total aerial vegetative tissue of *P. vulgaris* cv. ICA Pijao under cold stress, mRNA sequence	Abiotic stress responses	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011

Table A9.1. Continued

Phaseolus angustissimus	Py D (Phvul.008G095400.1)	FD800156.1	*M. sativa*	Pyruvate decarboxylase (PDC; putative)	Pv08: 9866707..9867000	cDNA similar to a pyruvate decarboxylase isozyme differentially expressed in total aerial vegetative tissue of *P. angustissimus* PI 535272 under cold stress, mRNA sequence	Abiotic stress responses	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
P. angustissimus	CYP72A14 (Phvul.005G063500.1 / Phvul.005G063500.2)	FD783151.1	*M. sativa*	CYP450 (putative)	Pv05: 9236613..9237085	cDNA similar to a cytochrome P450 differentially expressed in total aerial vegetative tissue of *P. angustissimus* PI 535272 under cold stress, mRNA sequence	Abiotic stress responses	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
P. angustissimus	bHLH (Phvul.006G093900.1)	FD782541.1	*M. sativa*	Basic helix-loop-helix (bHLH) transcription factor (putative)	Pv06: 21189471..21189721	cDNA similar to a bHLH transcription factor differentially expressed in total aerial vegetative tissue of *P. angustissimus* PI 535272 under cold stress, mRNA sequence	Low temperature response	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
P. angustissimus	Nfu2 (Phvul.006G218300.1)	FD797482.1	*M. sativa*	NFU domain protein 2 (putative)	Pv06: 31762369..31762654	cDNA similar to NFU domain protein 2 differentially expressed in total aerial vegetative tissue of *P. angustissimus* PI 535272 under cold stress, mRNA sequence	Chloroplast biogenesis	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
P. angustissimus	CIP1 (Phvul.002G296300.1)	FD782079.1	*M. sativa*	Photomorphogenic interactive protein 1 (putative)	Pv02: 45972050..45972270	cDNA similar to constitutive photomorphogenic interactive protein 1 differentially expressed in total aerial vegetative tissue of *P. angustissimus* PI 535272 under cold stress, mRNA sequence	Chloroplast biogenesis	Woronuk *et al*., 2010: 96; Vijayan *et al*., 2011
Low-phosphorous response								
P. vulgaris	PIP2;1 (putative; Q506K1; Phvul.011G079300.1)	PHAVU_011G079300g	*P. vulgaris* cv. "Pinto" (AY995195.1)	Aquaporin PIP; major intrinsic protein (MIP) (putative)	Pv11: 7352592..7354459 (–strand)	*P. vulgaris* putative aquaporin (PIP2;1; Q506K1) mRNA, complete cds	Aquaporin / P transport	Graham *et al*., 2006
P. vulgaris	FBP1 (Q9XFL3; Phvul.006G129600.1)	PHAVU_006G129600g	*P. vulgaris* cv. "Mon[t]calm" (AF149277.1)	Peroxidase	Pv06: 24415433..24417689 (– strand)	*P. vulgaris* peroxidase 1 precursor (FBP1; Q9XFL3) mRNA, partial cds	Response to oxidative stress; phenylpropanoid biosynthesis; metabolism	Graham *et al*., 2006
P. vulgaris	TGA2.2 (Q93XA0; Phvul.001G123300.1)	PHAVU_001G123300g	*P. vulgaris* cv. "Red Kidney" (AF402609.1)	bZIP transcription factor	Pv01: 34642377..34648835 (– strand)	*P. vulgaris* TGA-type basic leucine zipper protein TGA2.2 (Q93XA0) mRNA, complete cds	Sequence-specific DNA binding	Graham *et al*., 2006

Table A9.1. Continued

P. vulgaris	PvPR1 (P25985; Phvul.003G109100.1)	PHAVU_003G109100g	*P. vulgaris* cv. "Canadian Wonder" (X61365.1)	Pathogenesis-related protein Bet v I family	Pv03: 27013976..27014992 (– strand)	*P. vulgaris* mRNA for pathogenesis-related protein 1 (PvPR1; P25985) of unknown function	Pathogenesis-related protein; defence response; response to biotic stimulus	Graham *et al*., 2006
P. vulgaris	PvPR2 (P25986; Phvul.002G209500.1)	PHAVU_002G209500g	*P. vulgaris* cv. "Canadian Wonder" (X61364.1)	Pathogenesis-related protein Bet v I family	Pv02: 36948404..36949286 (+ strand)	*P. vulgaris* mRNA for pathogenesis-related protein 2 (PvPR2; P25986) of unknown function	Pathogenesis-related protein; defence response; response to biotic stimulus	Graham *et al*., 2006
P. vulgaris	Q41125 (Phvul.003G219200.1 / Phvul.003G219100.1)	PHAVU_003G219200g / PHAVU_003G219100g	*P. vulgaris* cv. unspecified (U34333.1)	Protease inhibitor/seed storage/LTP family	Pv03: 43634374..43635634 (– strand) / Pv03: 43630346..43631294 (+ strand)	*P. vulgaris* proline-rich 14 kDa protein (Q41125) mRNA, complete cds	Plant lipid transfer	Graham *et al*., 2006
P. vulgaris	CNGC-A (Q8H6U3; Phvul.006G112200.1)	PHAVU_006G112200g	*P. vulgaris* cv. "Red Kidney" (AF492816.1)	Voltage and ligand gated potassium channel	Pv06: 22804166..22808740 (+ strand)	*P. vulgaris* cyclic nucleotide-gated channel A (CNGC-A; Q8H6U3) mRNA, partial cds	Cyclic nucleotide-binding; ion transport	Graham *et al*., 2006
Roots								
P. vulgaris	Phvul.010G129400.1	PHAVU_010G129400g	*Medicago truncatula* (MTR_8g020840; NSP1 protein)	GRAS domain family (provisional)	Pv10: 39948943..39951082 (– strand)	Transcript for hypothetical protein (GRAS domain family [provisional]) in *P. vulgaris* landrace G19833 similar to nodulation-signalling pathway 1 (NSP1) transcription factor in *M. truncatula* (79% identity; 83% query [MTR_8g020840] coverage)	Nodule organogenesis; gibberellin (GA) signalling	Ariel *et al*., 2012
P. vulgaris	Phvul.009G122700.1	PHAVU_009G122700g	*M. truncatula* (MTR_3g072710; NSP2 protein)	GRAS domain family (provisional)	Pv09: 18210950..18212455 (– strand)	Transcript for hypothetical protein (GRAS domain family [provisional]) in *P. vulgaris* landrace G19833 similar to nodulation-signalling pathway 2 (NSP2) transcription factor in *M. truncatula* (72% identity, 98% query [MTR_3g072710] coverage)	Nodule organogenesis / gibberellin (GA) signalling	Ariel *et al*., 2012

Table A9.1. Continued

P. vulgaris	Phvul.008G165200.1	PHAVU_008G165200g	*M. truncatula* (MTR_5g058860; NSP protein)	GRAS domain family (provisional)	Pv08: 42672531..42673898 (– strand)	Transcript for hypothetical protein (GRAS domain family [provisional]) in *P. vulgaris* landrace G19833 similar to nodulation-signalling pathway (NSP) transcription factor in *M. truncatula* (73% identity, 76% query [MTR_5g058860] coverage)	Nodule organogenesis / gibberellin (GA) signalling	Ariel *et al*., 2012
P. vulgaris	Phvul.011G200900.1 (Pv-RHS24-13)	EY457919.1	*G. max* (*Rpg1-b*)	Nucleotide-binding site–leucine-rich repeat-type disease resistance protein	Pv11: 47907355..47911290 (– strand)	Pv-RHS24-13 *Phaseolus vulgaris* cDNA similar to NBS-LRR type disease resistance protein, mRNA sequence	Nodulation / defence response	Meschini *et al*., 2008
P. vulgaris	Phvul.011G061100.1 (Pv-RHS24-03; *Rab11c*)	EY457909.1	*Lotus japonicus*	Pfam: Ras family; KOG: GTPase Rab11/YPT3, small G protein superfamily	Pv11: 5292708..5294877 (+ strand)	Pv-RHS24-03 *Phaseolus vulgaris* cDNA similar to RAB11C, mRNA sequence. Small GTP binding protein Rab11c	Nodulation	Meschini *et al*., 2008
P. vulgaris	Phvul.001G236600.1 (Pv-RHS24-01)	EY457907.1	*O. sativa*	PANTHER: leucine-rich repeat receptor-like protein kinase; Pfam: Protein tyrosine kinase; KOG: Serine/threonine protein kinase	Pv01: 49698036..49701035 (+ strand)	Pv-RHS24-01 *Phaseolus vulgaris* cDNA similar to Serine/Threonine protein kinase, mRNA sequence. Putative receptor-like kinase	Nodulation	Meschini *et al*., 2008
Growth habit and maturity								
P. vulgaris	*PvTFL1y*	"G12873" (indeterminate Mesoamerican): EF643248.2. "G19833" (ref. genome): PHAVU_001G189200.1	*A. thaliana* (831683; Terminal Flower 1 protein, TFL1)		Pv01: 45561512..45563326 (– strand)	RCN4 Centroradialis-like1 homologous to *TFL1* gene; contains Pfam profile PF01161: Phosphatidylethanolamine-binding protein	Determinacy / flowering	Kwak *et al*., 2008; Repinski *et al*., 2012
P. vulgaris	*ELF3* orthologue	PHAVU_010G142900g	*G. max* (LOC102663657 & LOC100820110 proteins EARLY FLOWERING 3-like)		Pv10: 41472218..41478694 (– strand)		Early flowering	Kwak *et al*., 2008

Table A9.1. Continued

P. vulgaris	*ELF3* orthologue	PHAVU_001G032900g	*G. max* (LOC100788038 & LOC100818279 protein EARLY FLOWERING 3-like)		Pv01: 3178404..3183178 (+ strand)		Early flowering	Kwak *et al*., 2008
Seed coat colour and nutritional traits								
P. vulgaris	Phvul.007G108500.1	CW652102.1 (Myb15-CW652102)		Myb-like DNA-binding domain; Myb superfamily	Pv07: 13461806..13464239 (– strand)	Myb15-3 *Phaseolus vulgaris* (variety OAC Rex) / DNA *Phaseolus vulgaris* genomic clone Myb15-3 5- similar to *P. coccineus* EST for R2R3-MYB transcription factor (AtMYB15) CA902486, genomic survey sequence. [One of two nearest phenylpropanoid pathway genes to "ground factor" colour gene *P* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013
P. vulgaris	Phvul.007G150500.1	CV670734.1 (PAL2-P19142-1)		Phenylalanine and histidine ammonia-lyase	Pv07: 37020027..37024239 (– strand)	Pv PAL2-4 *Phaseolus vulgaris* (variety OAC Rex) / cDNA Phaseolus vulgaris cDNA clone Pv PAL2-4 5- similar to *Phaseolus vulgaris* phenylalanine ammonia lyase 2 (PAL2) P19142., mRNA sequence. [One of two nearest phenylpropanoid pathway genes to "ground factor" colour gene *P* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013
P. vulgaris	Phvul.008G072700.1	CV670753.1	*P. coccineus*		Pv08: 6750350..6750668	Pv HD-1 *P. vulgaris* (variety OAC Rex) / cDNA *P. vulgaris* cDNA clone Pv HD-1 5- similar to *P. coccineus* homeodomain protein GL2 like 1/ANL2 (HD-GL2) EST CA902455., mRNA sequence [One of three nearest phenylpropanoid pathway genes to "complete colour" gene *C* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002
P. vulgaris	Phvul.008G076600.1	PHAVU_008G076600g	*G. max* (vestitone reductase, EU925557.1)	Pfam: NAD dependent epimerase/dehydratase family; KOG: Flavonol reductase/cinnamoyl-CoA reductase	Pv08: 7325513..7328038 (– strand)	[One of three nearest phenylpropanoid pathway genes to "complete colour" gene *C* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002

Table A9.1. Continued

P. vulgaris	Phvul.008G113600.1	PHAVU_008G113600g	*P. acutifolius* (glutathione S-transferase, AY220095.1)	Glutathione S-transferase (GST), superfamily, GST-domain-containing	Pv08: 13646227. .13647651 (– strand)	Glutathione S-transferase [One of three nearest phenylpropanoid pathway genes to "complete colour" gene *C* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002
P. vulgaris	Phvul.004G104100.1	PHAVU_004G104100g	*G. max* (Flavonoid glycosyltransferase, BE190894.1)	UDP-glucoronosyl and UDP-glucosyl transferase	Pv04: 33042774. .33044196 (+ strand)	Glycosyl transferases group 1 gene [One of two nearest phenylpropanoid pathway genes to "yellow-brown" gene *G* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002
P. vulgaris	Phvul.004G145500.1	CW652099.1	*G. max* (chalcone reductase, CHR, AI965636)	Aldo/keto reductase family	Pv04: 42558699. .42560395 (– strand)	CHR-2 *Phaseolus vulgaris* (variety OAC Rex) / DNA *Phaseolus vulgaris* genomic clone CHR-2 5- similar to *G. max* EST for chalcone reductase (CHR) AI965636, genomic survey sequence. [One of two nearest phenylpropanoid pathway genes to "yellow-brown" gene *G* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002
P. vulgaris	Phvul.006G054200.1	PHAVU_006G054200g	*G. max* (cytochrome P450 71D9-like, XM_003551443.2)	Cytochrome P450	Pv06: 16967791. .16970297 (- strand)	Cytochrome P450. [One of two nearest phenylpropanoid pathway genes to "violet" gene *V* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002
P. vulgaris	Phvul.002G268500.1	CV670740.1	*M. sativa* (cinnamyl alcohol dehydrogenase 1, CAD 1, L46856)	Alcohol dehydrogenase GroES-like domain; zinc-binding dehydrogenase	Pv02: 43368131. .43371549 (– strand)	Pv CADn-2 *Phaseolus vulgaris* (variety OAC Rex) / cDNA *Phaseolus vulgaris* cDNA clone Pv CADn-2 5- similar to *Medicago sativa* cinnamyl alcohol dehydrogenase 1 (CAD 1) L46856, mRNA sequence. Probable cinnamyl alcohol dehydrogenase. [One of three nearest phenylpropanoid pathway genes to "no colour expression" gene *Rk* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002
P. vulgaris	Phvul.002G290000.1	PHAVU_002G290000g	*G. max* (EW678711.1)	Pfam: Helix-loop-helix DNA-binding domain; PANTHER: MYC; GL3 (GLABRA 3), TF	Pv02: 45309544. .45313699 (+ strand)	bHLH-MYC and R2R3-MYB transcription factors N-terminal. [One of three nearest phenylpropanoid pathway genes to "no colour expression" gene *Rk* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002
P. vulgaris	Phvul.002G294800.1	PHAVU_002G294800g	*G. max* (dirigent protein, BE806864.1)	Pfam: Dirigent-like protein; PANTHER: nucleoporin-related; disease resistance response protein-related	Pv02: 45829124. .45829837 (– strand)	Dirigent-like protein. [One of three nearest phenylpropanoid pathway genes to "no colour expression" gene *Rk* in *P. vulgaris*]	Phenylpropanoid pathway	Reinprecht *et al*., 2013; McClean *et al*., 2002

Table A9.1. Continued

P. vulgaris	Phvul.002G040100 (*4CL-1*)	KF303286.1	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv02: 3830262..3833723 (– strand)	*Phaseolus vulgaris* 4-coumarate:coenzyme A ligase 1 (*4CL-1*) gene, complete cds	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013
P. vulgaris	Phvul.003G146900 (*4CL-2*)	KF303287	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv03: 34628465..34633500 (+ strand)	*Phaseolus vulgaris* polyprotein gene, complete cds; retrotransposon copia-type, complete sequence; and disrupted 4-coumarate:coenzyme A ligase 2 (*4CL-2*) gene, partial sequence	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013
P. vulgaris	Phvul.003G147000 (*4CL-3*)	KF303288	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv03: 34642251..34646976 (+ strand)	*Phaseolus vulgaris* 4-coumarate:coenzyme A ligase 3 (*4CL-3*) gene, complete cds	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013
P. vulgaris	Phvul.005G131400 (*4CL-4*)	KF303289	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv05: 35684228..35688356 (+ strand)	*Phaseolus vulgaris* 4-coumarate:coenzyme A ligase 4 (*4CL-4*) gene, complete cds	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013
P. vulgaris	Phvul.005G184500 (*4CL-5*)	KF303290	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv05: 40478925..40481188 (+ strand)	*Phaseolus vulgaris* 4-coumarate:coenzyme A ligase 5 (*4CL-5*) gene, complete cds	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013
P. vulgaris	Phvul.006G057300 (*4CL-6*)	KF303291	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv06: 17403293..17406728 (+ strand)	*Phaseolus vulgaris* 4-coumarate:coenzyme A ligase 6 (*4CL-6*) gene, complete cds	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013
P. vulgaris	Phvul.011G020200 (*4CL-7*)	KF303292	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv11: 1586744..1589607 (+ strand)	*Phaseolus vulgaris* 4-coumarate:coenzyme A ligase 7 (*4CL-7*) gene, complete cds	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013
P. vulgaris	Phvul.011G084300 (*4CL-8*)	KF303293	Pfam: AMP-binding enzyme; EC: 4-coumarate--CoA ligase; KOG: Acyl-CoA synthetase	Pv11: 8078453..8081905 (– strand)	*Phaseolus vulgaris* 4-coumarate:coenzyme A ligase 8 (*4CL-8*) gene, complete cds	Phenylpropanoid pathway; 4CL	Reinprecht *et al*., 2013

Table A9.1. Continued

P. vulgaris	*Alpha-phaseolin* (*Phs*) gene (Phvul.007G059700.1)	KJ544107.1	*P. vulgaris* cultivar Red Mexican	Cupin	Pv07: 4959769.. 4961920 (+ strand)	*P. vulgaris* cultivar Red Mexican *alpha-phaseolin* (*Phs*) gene, complete cds (seed protein and weight locus)	Nutrient reservoir activity (formerly storage protein)	Kelly *et al*., 2003, and references therein

*National Center for Biotechnology Information (NCBI) accession number unless otherwise stated.