Genetics and Genomics of Key Agronomic Traits in Field pea (*Pisum sativum* L.)

Submitted by

Shimna Sudheesh

Master of Biotechnology

A thesis submitted in total fulfilment

of the requirements for the degree of

Doctor of Philosophy by published work

School of Applied Systems Biology

Faculty of Science, Technology and Engineering

La Trobe University

Bundoora, Victoria 3086

Australia

September 2015

TABLE OF CONTENTS

TABLE OF CONTENTS i

LIST OF TABLES v

LIST OF FIGURES vi

SUMMARY vii

STATEMENT OF AUTHORSHIP ix

ACKNOWLEDGEMENTS x

THESIS PREFACE xii

LIST OF ABBREVIATIONS xiii

CHAPTER 1 1

Introduction 1

1.1 Taxonomy, origin, genetics and characteristics of the genus *Pisum* 1

1.2 Agronomy and breeding 6

1.3 Abiotic stresses of field pea 8

1.3.1 Soil salinity stress 8

1.3.2 Soil boron toxicity stress 10

1.4 Biotic stresses of field pea 11

1.4.1 Powdery mildew of field pea 12

1.4.2 Bacterial blight of field pea 15

1.4.2.1 Disease cycle of bacterial blight 17

1.4.2.2 Biology of bacterial blight pathogens 18

1.4.2.3 Bacterial virulence factors contributing to disease and pathogenesis 19

1.4.2.4 Factors influencing disease development 19

1.4.2.5 Effects of bacterial blight infection on production 20

1.4.2.6 Dissemination of bacterial blight 22

1.4.2.7 Control measures 22

1.4.2.8 Genetics of bacterial blight 24

1.5 Plant innate immunity: insight into defence mechanisms 25

1.6 Molecular genetic markers 29

1.7 Construction of genetic linkage maps and trait dissection based on QTL analysis 33

1.7.1 The mapping population 33

1.7.2 Computer software for genetic mapping and trait dissection based on QTL analysis 35

1.8 Conserved synteny analysis in legumes 38

1.9 Gene expression analysis 40

1.10 Objectives of this study 43

CHAPTER 2 45

SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.) 45

2.1 Chapter preface 45

2.2 Publication details 45

2.3 Statement of contributions of joint authorship 46

2.4 Statement from the co-author confirming the authorship contribution of the PhD candidate 47

2.5 Manuscript 48

CHAPTER 3 62

Consensus genetic map construction for field pea (*Pisum sativum* L.), trait dissection of biotic and abiotic stress tolerance and development of a diagnostic marker for the *er1* powdery mildew resistance gene 62

3.1 Chapter preface 62

3.2 Publication details 63

3.3 Statement of contributions of joint authorship 63

3.4 Statement from the co-author confirming the authorship contribution of the PhD candidate 64

3.5 Manuscript 65

CHAPTER 4 78

De novo assembly and characterisation of the field pea transcriptome using RNA-Seq 78

4.1 Chapter preface 78

4.2 Publication details 78

4.3 Statement of contributions of joint authorship 79

4.4 Statement from the co-author confirming the authorship contribution of the PhD candidate 79

4.5 Manuscript 81

CHAPTER 5 96

Construction of an integrated linkage map and trait dissection for bacterial blight resistance in field pea (*Pisum sativum* L.) 96

5.1 Chapter preface 96

5.2 Publication details 96

5.3 Statement of contributions of joint authorship 97

5.4 Statement from the co-author confirming the authorship contribution of the PhD candidate 97

5.5 Manuscript 99

CHAPTER 6 112

General discussion 112

6.1 Background and context of the research 112

6.2 Overview of the research 114

6.3 Future directions 118

REFERENCES 127

LIST OF TABLES

Table 1.1: Attributes of species..... 5

Table 1.2: Effector molecules, their sources, targets, corresponding R-proteins and R-protein structure 28

LIST OF FIGURES

Figure 1.1: Phylogenetic relationships of Papilionoideae legume featuring the two clades 2

Figure 1.2: Statistics of Australian pea production 6

Figure 1.3: Field pea plants infected with (A) *P. syringae* pv. *pisi* race 6 (B) *P. syringae* pv. *syringae* in glasshouse inoculation experiment 16

Figure 1.4: Pea cultivation in the field in Horsham, Victoria during 2012, showing the susceptible cultivar Kaspa infected by bacterial blight 21

Figure 1.5: Representation of the location and structure of the five main classes of plant disease resistance proteins 26

SUMMARY

Field pea (*Pisum sativum* L.) is a self-pollinating, diploid, cool-season food legume. Crop production is constrained by multiple abiotic (including soil salinity and boron [B] toxicity) and biotic (powdery mildew and bacterial blight infection) stress factors, that cause reduced growth and yield. Development of varieties which are tolerant to these stresses is essential, and application of molecular genetic markers could greatly accelerate this process. Advances in genomic technology have permitted the development of 768 single nucleotide polymorphism (SNP) markers associated with expressed sequence tags (ESTs). Comprehensive genetic linkage maps were generated based on existing simple sequence repeat (SSR) markers and new SNP markers for four RIL-based mapping populations of field pea. By combining these four bi-parental maps with previously published consensus maps, a comprehensive integrated structure was obtained. Comparative genomic analysis of field pea with other legume species revealed high levels of conserved synteny between the genomes. Trait dissection of resistance to bacterial blight caused by *Pseudomonas syringae* pv. *syringae* and salinity tolerance identified multiple contributory genomic regions leading to quantitative inheritance. In contrast, B toxicity tolerance, powdery mildew resistance and bacterial blight (caused by *P. syringae* pv. *pisi*) resistance were predominantly controlled by a single genomic region. Resequencing of the *PsMLO1* candidate gene from powdery mildew resistant and susceptible genotypes allowed the design and validation of a putative diagnostic marker. This study also generated a comprehensive transcriptome data sets from two genotypes of field pea through the use of RNA sequencing technology (RNA-Seq), and performed comparison to gene complements in related species, sequence annotation and assessment of tissue-specific expression. The resources generated in this study will support further development of genetic markers, map construction and enhancement, identification of marker-trait associations, genomics-assisted breeding, map-based gene

cloning, comparative genetics as well to identify target genes for genetic modification approaches on the basis of annotation and expression analysis.

STATEMENT OF AUTHORSHIP

This thesis consists primarily of work by the author that has been published or accepted for publication as described in the text. Except where reference is made in the text of the thesis, this thesis contains no other material published elsewhere or extracted in whole or in part from a thesis submitted for the award of any other degree or diploma. No other person's work has been used without due acknowledgment in the main text of the thesis. This thesis has not been submitted for the award of any degree or diploma in any other tertiary institution.

Shimna Sudheesh

21 September 2015

ACKNOWLEDGEMENTS

First and foremost I express my gratitude to my supervisor Prof John Forster, for his excellent guidance, support and encouragement. It has been my honour and privilege to have such a caring, well-organised and efficient supervisor during my PhD research. I am also deeply grateful to Dr Tim Sawbridge for his scientific advice and knowledge and many insightful discussions and suggestions.

I would like to express my special appreciation and thanks to my co-supervisor Dr Sukhjiwan Kaur, you have been a tremendous mentor for me. I would like to thank you for encouraging my research and guiding me whilst allowing me the room to work in my own way, and permitting me to grow as a research scientist. Your advice has always been priceless. My sincere thanks also goes to Dr Noel Cogan, the funniest and one of the smartest people I know, for his support, motivation, and immense knowledge. His guidance helped me in all the time of research and writing of manuscripts and this thesis.

My gratitude is also extended to Dr Sareena Sahab and Dr Daniel Isenegger, for providing me with direction and technical support in tissue culture and microscopy. Appreciation also goes out to all molecular genetics team members in Bundoora and also the pulse improvement team members in Horsham for their support throughout this journey.

I am grateful for the financial contribution from the Victorian Department of Economic Development, Jobs, Transport and Resources (DEDJTR) and the Grains Research and Development Corporation (GRDC), which allowed me to undertake this research.

A special thanks to my family. Words cannot express how grateful I am to my daughters, Ananya and Vyshali and to my mother Shobha, for all of the sacrifices that you have made on my behalf. I would also like to thank all of my friends who supported me and incentivised me to strive towards my goal. Last, I must acknowledge my gratitude to my husband Sudheesh, through his love, patience, support and unwavering belief in me, I have been

able to complete this long dissertation journey. I could not have succeeded without your invaluable support and encouragement.

THESIS PREFACE

This thesis is composed of six chapters, and the original experimental content is presented in the form of a series of peer-reviewed journal articles. The first chapter provides a general overview of the literature in this area of research. The manuscripts that have been published are presented in chapters 2 to 5. Each of these chapters contains its own detailed introduction, methodology, results, and discussion sections. Each chapter is prefaced by summary of the work, the publication details of that manuscript, the contribution of co-authors and a statement from a co-author confirming the authorship contribution of the PhD candidate. The sixth chapter provides a general discussion that integrates the major themes from each of these manuscripts, as well as providing suggestions for future research directions. Because each experimental chapter corresponds to an independently published manuscript, some redundancy of content has arisen between the introduction and materials and methods sections of the respective journal articles. In addition, the individual experimental chapters employ the respective distinct referencing and citation styles of the corresponding journals. In contrast, a single referencing and citation style has been employed for chapters 1 and 6, and the bibliography is provided at the end of the thesis. Supplementary materials for the manuscripts featured in chapters 2 to 5 are available at the relevant journal web sites.

LIST OF ABBREVIATIONS

AFLP	Amplified fragment length polymorphism
BC	Backcross
B.C.	Before Christ
BLRV	Bean leaf roll virus
bp	Base pair
cDNA	Complementary deoxyribonucleic acid
c.	Circa
CIM	Composite interval mapping
cM	Centimorgan
CRISPR	Clustered regulatory interspaced short palindromic repeat
DEDJTR	Department of Economic Development, Jobs, Transport and Resources
DH	Double haploids
DNA	Deoxyribonucleic acid
EPS	Exopolysaccharides
EST	Expressed sequence tags
ERIC	Enterobacterial repetitive intergenic consensus
GBS	Genotyping-by-sequencing
GEBV	Genomic estimated breeding value
GMO	Genetically modified organism

GS	Genomic selection
GWAS	Genome-wide association study
HR	Hyper-sensitive response
ILDIS	International legume database and information service
kt	Kilotonnes
LG	Linkage group
LOD	Base 10 logarithm of odds
LRR	Leucine-rich repeat
LTR	Long terminal repeat
MAGIC	Multi-parent advanced generation inter-cross
MAS	Marker-assisted selection
Mb	Mega base pair
MIP	Major intrinsic protein
MLST	Multilocus sequence typing
mRNA	Messenger ribonucleic acid
Myr	Million years
NAM	Nested association mapping
NBS	Nucleotide-binding site
NGS	Next-generation sequencing
NIL	Near isogenic line
OPA	Oligonucleotide pool assay

ORF	Open reading frame
PCR	Polymerase chain reaction
PSbMV	Pea seed-borne mosaic virus
pv.	Pathovar
QTL	Quantitative trait locus/loci
R	Resistance
RAD-seq	Restriction site-associated DNA sequencing
RAPD	Randomly amplified polymorphic deoxyribonucleic acid
RE	Restriction endonuclease
REP	Repetitive extragenic palindromic
RF	Recombination frequency
RFLP	Restriction fragment length polymorphism
RIL	Recombinant inbred line
RNA	Ribonucleic acid
RNA-Seq	Ribonucleic acid sequencing technology
SCAR	Sequence characterised region
SIM	Simple interval mapping
SMR	Single marker regression
SNP	Single nucleotide polymorphism
SnRNA	Small nuclear ribonucleic acid
SSN	Sequence-specific nuclease

SSR	Simple sequence repeat
STS	Sequence tagged site
TALEN	Transcription activator-like effector nuclease
T3SS	Type III secretion system
tRNA	Transfer ribonucleic acid
USCE	Unequal sister-strand chromatid exchange
VNTR	Variable-number tandem repeat
Vp	Phenotypic variance
WGD	Whole-genome duplication
ZFN	Zinc finger nuclease

CHAPTER 1

Introduction

1.1 Taxonomy, origin, genetics and characteristics of the genus *Pisum*

The legumes (Fabaceae or Leguminosae) are the third largest family of angiosperms after the Orchidaceae and Asteraceae, and is second in size only to the Poaceae (grasses and cereals) in terms of agricultural and economic importance. This family contains 700 genera and over 19,000 species distributed globally in different ecoclimatic zones, from deserts of high latitude to dry or wet tropical forests. Many legume species are harvested as crops for human and animal consumption as well as for oils, fibre, fuel, fertilizer, timber, medicinal, chemical, and horticultural varieties (Lewis *et al.* 2005). The family is classified into three major sub-families: Mimosoideae, Caesalpiniodeae and Papilionoideae. Most agriculturally important species are located within the Phaseoloid and Galegoid clades of the Papilionoideae (Fig. 1.1). Phaseoloid species include most of the tropical or warm-season legumes, such as common bean (*Phaseolus vulgaris* L.), mung bean (*Vigna radiata* [L.] Wilczek.), soybean (*Glycine max* [L.] Merr.) and pigeon pea (*Cajanus cajan* [L.] Millsp.). In contrast, Galegoid species are cool-season legumes such as clovers (*Trifolium*), pea (*Pisum sativum* L.), lentil (*Lens culinaris* Medik.), faba bean (*Vicia faba* L.), chickpea (*Cicer arietinum* L.), as well as the model species *Medicago truncatula* Gaertn. and *Lotus japonicus* L. (Doyle and Luckow 2003). Genomic comparisons and analyses of duplication patterns within papilionoid legume genomes strongly suggest that an ancestral whole-genome duplication (WGD) event occurred c. 58 million years (Myr) ago (Pfeil *et al.* 2005). Subsequent to this WGD event, papilionoid taxa radiated into several clades between 45-54 Myr ago (Lavin *et al.* 2005). Even though papilionoid species share relatively close phylogenetic relationships, the genetic systems represented within the group are diverse, ranging from simple autogamous diploids (such as *L. japonicus*) to

complex outbreeding allotetraploid (white clover) (Choi *et al.* 2004b). Among the broader group, pea is an important grain legume (pulse) that is widely cultivated throughout the world. Pea is a cool-season crop which is extensively grown in temperate zones, but also cultivated at cooler altitudes in the tropics and winter seasons in the sub-tropics.

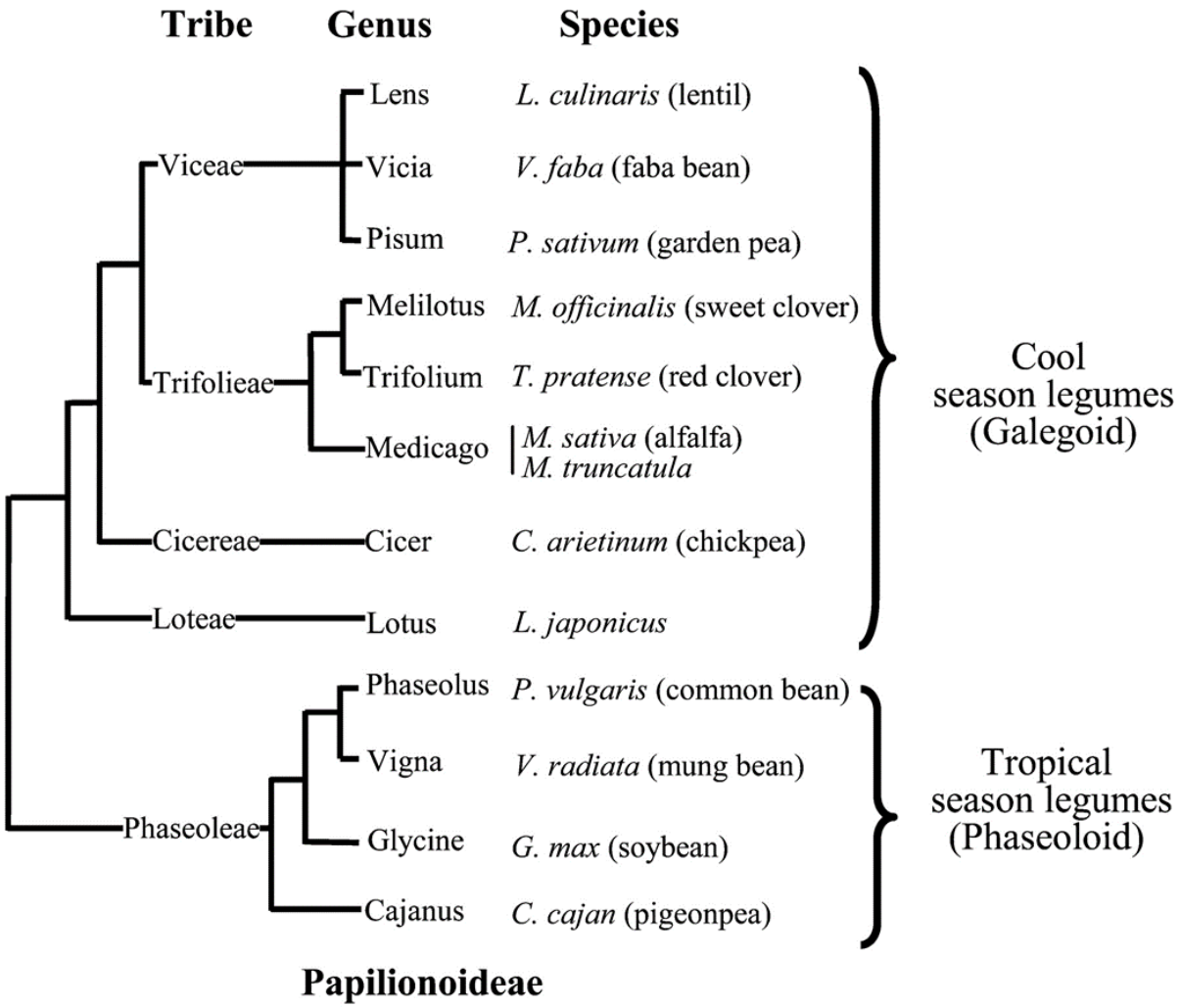

Figure 1.1: Phylogenetic relationships of Papilionoideae legume featuring the two clades (Zhu *et al*. 2005)

Taxonomic classification of *Pisum* species on the basis of morphological and karyotypic variation has been the subject of much dispute. The International Legume Database and

Information Service (ILDIS) accepts three species, *Pisum abyssinicum*, *Pisum fulvum* and *Pisum sativum*, the latter being further divided into two sub-species: *elatius* and *sativum* (http://www.ildis.org). Based on the analysis of morphology, ecology, cytogenetics and molecular diversity, it was concluded that *P. elatius*, *P. humile*, and *P. sativum* form a single-species complex comprised of two main morphotypes: weedy forms such as *P. elatius* and *P. humile*, and cultivated derivatives such as *P. sativum* (Ben-Ze'ev and Zohary 1973; Davis 1970). *P. sativum* is now generally viewed as a complex species that includes a wide variety of cultivated and wild forms.

Previous studies have shed light on the origin and progenitors of *P. sativum* (Vavilov 1926; Schaefer *et al.* 2012; Vershinin *et al.* 2003). Four centres of origin based on genetic diversity were proposed by Vavilov (1926): Near East, Central Asia, Ethiopia and the Mediterranean. Archaeological evidence indicates that peas were cultivated around 9000-8000 B.C. in early Neolithic farming villages in a number of centres including Iran, Israel and Turkey. Other centres in Greece and Eastern Europe, dated between 7500-6500 B.C., are associated with spread of the crop to the west and north (Zohary and Hopf 1973). A phylogeographical study based on DNA-level polymorphisms associated with the plastid genome and selected nuclear genome sequences suggested that wild pea was spread from its centre of origin in the Middle East eastwards to the Caucasus, Iran and Afghanistan, and westwards to the Mediterranean (Smykal *et al.* 2011). However, research to date has been unable to verify the primary centre of domestication, although this is likely to be within the range of distribution of the wild progenitor. The modern gene pool of cultivated *Pisum* is diverse, reflecting the effects of early domestication and subsequent widespread cultivation. The genetic diversity of *Pisum* as a whole, and high levels of polymorphism in all species (apart from *P. abyssinicum*) indicate a high contribution of recombination between multiple ancestral lineages (Vershinin *et al.* 2003). The two independently domesticated pea species, *P. abyssinicum* and *P. sativum*, arose in contrasting ways from the common processes of hybridisation, introgression and selection

without associated transpositional activity (Vershinin *et al.* 2003). Though *Pisum* is well known to be an inbreeding species, a substantial level of heterogeneity is nonetheless conserved within pea populations (Jing *et al*. 2007).

The genome organisation of higher plant species may differ considerably, despite the presence of similar biological features and developmental mechanisms (Murray *et al*. 1981). Most of this variation is due to differences in ancestral ploidy and differences in repetitive DNA content (primarily attributable to retroelements, especially the long terminal repeat [LTR]-retrotransposons) (Bennetzen *et al*. 2005). Such effects have been observed in a number of families, such as the Brassicaceae and Poaceae (Devos and Gale 2000). Crop legumes also differ greatly in terms of genome size and basic chromosome number (Table 1) (Choi *et al.* 2004b). Pea has a large genome size (c. 5000 Mb/1C or 4.6 pg of DNA) (Murray *et al.* 1978), which is c. 10-fold larger than that of the model legume species *M. truncatula* (c. 500 Mb) and c. 4-fold larger than the genome of soybean (c. 1100 Mb) (Schmutz *et al.* 2010). The *Pisum* genome is rich in repetitive DNA, which is estimated to comprise c. 50-70% of the nuclear genome complement (Macas *et al*. 2007; Murray *et al*. 1978). The repeats belong to DNA sequence classes such as satellite DNAs, simple sequence repeats and multiple families of transposable elements. Ty3/gypsy LTR-retrotransposons were found to contribute the major component of the repetitive DNA, followed by Ty1/copia elements, while non-LTR retrotransposons account for only a minor fraction of pea nuclear DNA. Most of the single copy sequences (accounting for 30% of the pea genome) are actually relics of ancient families of repeated sequences, which have diverged by mutational processes sufficiently to behave essentially as unique sequences (Murray *et al.* 1978, 1981).

Table 1.1: Attributes of species (Choi *et al.* 2004b)

Species	Genome size (Mbp)	Gametic chromosome number (n)	Tribe	Clade
M. truncatula	500	8	Trifoleae	Galegoid
L. japonicus	500	6	Loteae	Galegoid
V. radiata	520	11	Phaseoleae	Phaseoloid
P. vulgaris	620	11	Phaseoleae	Phaseoloid
G. max	1,100	20	Phaseoleae	Phaseoloid
M. sativa	1,600	16	Trifoleae	Galegoid
P. sativum	5,000	7	Viceae	Galegoid

The pea genome is diploid in nature, as confirmed by karyotypic analysis that has revealed a 2n = 2x = 14 chromosomal constitution (Cannon 1903; Cooper 1938; Murray *et al.* 1978). Based on both morphology and patterns of fluorescent *in situ* hybridisation, pea chromosomes are fully distinguishable and may be related to linkage groups as defined by genetic mapping studies (Fuchs *et al*. 1998; Neumann *et al*. 2002). Two of the seven chromosome pairs are nearly metacentric and five are submetacentric. Two of the latter carry satellites of different sizes on their long arms. The currently accepted scheme for chromosome nomenclature in pea arises from an early attempt to coordinate the designations for linkage groups and chromosomes. The chromosome numbers and linkage group (LG) numbers are referred to using Arabic and Roman numerals, respectively (1 = VI, 2 = I, 3 = V, 4 = IV, 5 = III, 6 = II and 7 = VII) (Fuchs *et al*. 1998).

1.2 Agronomy and breeding

Pea is the fourth most important legume crop after soybean, common bean and chickpea. In 2013, global pea cultivation was practiced on 6.4 million ha, and production reached 11 million tonnes (FAOSTAT 2014). The most important countries for pea cultivation are Canada, the Russian Federation, China, India, France and Australia. Australian field pea production in 2013 was 319.7 kilotonnes (kt) from 280,500 ha (Fig. 1.2), 60% of the product being exported to Asia, the Middle East and Europe (ABARES 2014). South Australia is the leading field pea growing state, contributing 40% of national production, followed by New South Wales and Victoria (each 20%). Varieties used in Australian production include the Dun - Kaspa type, Dun - others, White and Blue field pea. The Dun - Kaspa type (including Kaspa, PBA Gunyah and PBA Twilight) is now the most prevalent class of cultivars in Australian agriculture, due to favourable properties of environmental adaptability, good seed quality and high yield potential (Pulse Australia 2009).

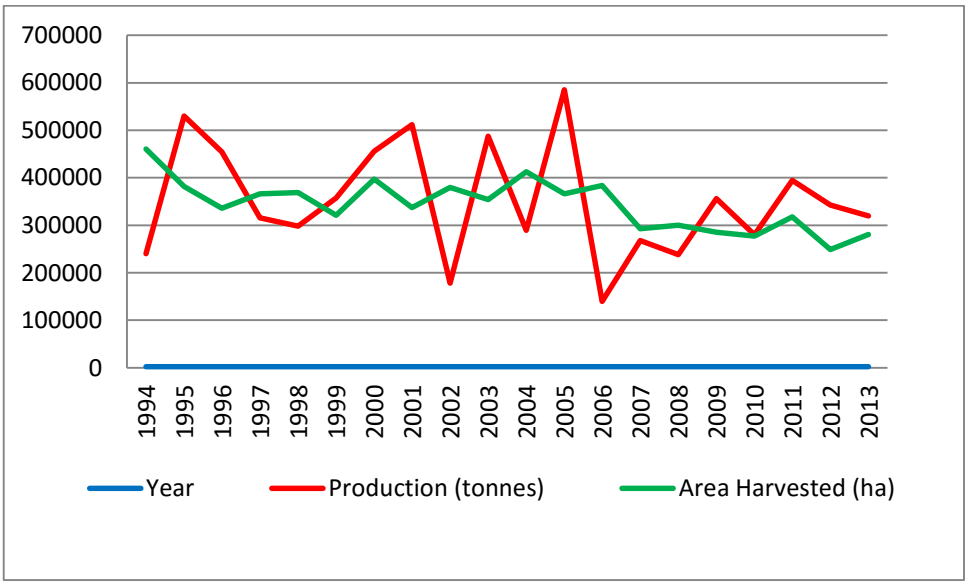

Figure 1.2: Statistics of Australian pea production (FAOSTAT 2014)

Field pea is primarily used for human consumption and livestock feed. Pea seeds are rich in protein (23-25%), starch (50%), soluble sugars (5%), fibre, minerals and vitamins (A, B and C). Peas are cultivated in order to obtain fresh green seeds, tender green pods, dried seeds and foliage. The green pods are eaten either raw or cooked, while dried peas are consumed split as dhal, roasted, parched or boiled. Pea plays a major role in crop rotations, breaking the disease cycles of cereals and enhancing weed control, as well as enhancing soil health through provision of nitrogen fixation (Nikolopoulou *et al.* 2007; Wang *et al.* 1998). Pea is also used for forage production (Kocer and Albayrak 2012).

The genetics of peas has been well-studied (Samatadze *et al.* 2008) ever since the pioneering work of Gregor Mendel in the nineteenth century (Mendel 1866). Certain features of pea biology, such as self-pollination, ease of cultivation, and provision of readily distinguishable phenotypic traits inspired Mendel to choose pea as the target organism for his experiments, such that the species has remained as a major subject for modern genetic studies. Many morphophysiological characteristics, such as the properties of upright growth habit associated with dwarf varieties that resist lodging and disease incidence, have played a role in increase of pea yield and production. Many of these characteristics display qualitative inheritance patterns (Blixt 1972), and are hence presumably under relatively simple genetic control (McPhee 2003).

To meet the present and future trends of world population growth, there is a need to develop sustainable high-yielding varieties which persist under conditions of abiotic and biotic stress. Molecular breeding programs have been used for crop improvement in several crops, including legumes such as soybean and common bean (Chamarthi *et al.* 2011), and are suitable for application to field pea. In Australia, gains in grain yield for field pea have been achieved by breeding for optimisation of crop architecture (that is, reduced internode length), harvest index and phenology traits with growing season length and rainfall (Redden *et al.* 2005). Breeding programs targeting resistance to fungal diseases such as ascochyta blight, powdery and downy mildew, and viruses such as pea seed-

borne mosaic virus (PSbMV) and bean leaf roll virus (BLRV) were also previously conducted. In contrast, little effort has been made towards the improvement of abiotic tolerance traits, as well as complex and putatively multigenic traits such as yield and quality (Kumar *et al.* 2011).

Introgression of genes for stably inherited traits including pest and disease resistance, herbicide resistance and improved protein quality may not be feasible from natural populations of close relatives of cultivated peas. The adoption of new genomic tools will enable the identification and selection of superior gene variants for various complex traits. Advances in plant genetic modification through transformation and regeneration system permits the introduction of foreign genes provide a potentially powerful tool for achieving these goals by other means (Schroeder *et al*. 1993).

1.3 Abiotic stresses of field pea

Pea production is constrained by a number of abiotic stresses that produce highly detrimental effects, including frost, waterlogging, drought, heat, soil pH, salinity, sodicity and boron toxicity (Dita *et al*. 2006).

1.3.1 Soil salinity stress

Soil salinity is a natural property of many environments, arising from the geomorphic and geochemical processes that lead to presence of excessive salts on the top layer of the soil, and resulting in the deterioration of its chemical and physical properties. It is a form of land degradation that has become a major cause of low agricultural productivity. Soil salinity affects more than 20% of cultivated land world-wide (Gupta and Huang 2014), but is relatively more prevalent in arid and semi-arid regions as compared to humid regions (Hussain *et al*. 2009). In Australia, different types of salinization, with a prevalence of sodium salts, affect c. 30% of the land area (Rengasamy 2006).

Cool season legumes in general are sensitive to salinity stress and the corresponding effects on seed germination, seedling growth as well as some physiological activities (Okcu *et al*. 2005). Assessments of biomass reduction due to salinity indicate that field pea is significantly more sensitive to this form of stress than crop species such as barley (*Hordeum vulgare* L.) (Maas 1986; Saxena *et al*. 1994), wheat (*Triticum aestivum* L.) (Francois and Maas 1994) and canola (*Brassica napus* L.) (Steppuhn *et al.* 2001), but appears to be much more tolerant than other pulses such as chickpea (Sadiki and Rabhi 2001) and lentil (Maher *et al*. 2003). Excess soil salinity causes reductions in plant growth rate, height, shoot and root dry matter accumulation, as well as marginal necrosis of older leaves that progresses up the plant (Leonforte 2013). Excess salt in the soil limits availability of water and nutrients to plants, and the adverse effects of salinity cause ionic and osmotic stress conditions, leading to oxidative stress (Zhu 2001). The symptoms of salinity stress in plants are necrosis and yellowing on the outer margins of the older leaves, which gradually progress to younger leaves, while the older leaves eventually dies and undergoes abscission (Lee *et al*. 2004). Salinity stress symptoms in pulses are readily confused with those of other nutrient disorder, disease, herbicide damage and drought stress.

Plant growth responses to salinity are complex in nature, as they cause multiple and varying reactions in terms of whole plant physiology and cellular function (Läuchli and Grattan 2007). Initially, salinity stress causes sudden reductions of plant growth due to osmotic adjustment (Chen and Jiang 2010), but the secondary growth response is slower, resulting from increasing sodium ion concentration in plant tissues (Jacoby 1999). In general, tolerance to salinity is achieved through: (1) ion homeostasis and compartmentalisation; (2) ion transport and uptake; (3) biosynthesis of osmoprotectants and compatible solutes; (4) activation of antioxidant enzymes and synthesis of antioxidant compounds; (5) synthesis of polyamines; (6) generation of nitric oxide (NO); and (7) modulation of phytohormone levels (Gupta and Huang 2014). Previously, different

physiological traits associated with salt tolerance in *M. truncatula* were studied, allowing the identification of a total of 19 putative quantitative trait loci (QTLs) (Arraouadi *et al.* 2011). For field pea, limited trait dissection studies of salinity stress tolerance mechanisms have so far been conducted to dissect different components, so only limited information is available at the physiological, chemical and molecular genetic levels (Kumar *et al*. 2011).

1.3.2 Soil boron toxicity stress

Boron is an essential micronutrient that is required for normal growth of plants. Requirements for boron vary between different crop species, but an excessive B concentration can result in toxicity effects, critical levels again varying between species (Schnurbusch *et al*. 2010). B toxicity is a significant agronomic problem worldwide, hindering food production in arid regions, including southern Australia, northern Africa and the Mediterranean basin (Bogacki *et al*. 2013). In Australia, 15% of agricultural soils in Western Australia (Lacey and Davies 2009) and 30% of soils in South Australia (Government of South Australia, Department of Environment and Natural Resources, http://www.environment.sa.gov.au/Home) are at moderate to high risk of B toxicity effects. The combined effects of salinity and high B concentration adversely affect both crop growth and grain yield, but the nature of the interaction is not clearly understood (Grieve and Poss 2000).

B toxicity can produce significant negative impacts on both seed yield and quality in legume species (Paull *et al.* 1991). Typical symptoms of B toxicity include leaf burn (that is, chlorotic and/or necrotic patches, often at the margins and tips of older leaves), excessive curling of tendrils, and reduced growth (Miwa *et al*. 2010). Formation of root nodules is also affected by B toxicity. Plants which are tolerant to B toxicity exhibit lower levels of root-specific uptake, which is followed by similar effects in the shoot. Such plants accumulate less B in their tissues, and are also associated with increased root growth, dry matter production and reduced leaf symptoms (Schnurbusch *et al*. 2010). Several

physiological mechanisms of tolerance to high boron concentration have been proposed, including B exclusion, altered distribution of accumulated B and internal B tolerance. Based on studies of a highly B tolerant barley landrace, it was concluded that tolerant cultivars must possess a membrane-active transporter that acts to export B from the root (Hayes and Reid 2004). It was reported that efflux-type B transporters and/or members of the major intrinsic protein (MIP) group or aquaporins are involved in the process of B uptake and translocation (Bogacki *et al*. 2013). The mechanism of boron tolerance in wheat was hypothesised to involve reduced uptake and differential translocation of boron (Wimmer *et al.* 2005). The internal B tolerance mechanisms include inactivation of excess B with solutes such as sorbitol (Rozema *et al*. 1992), and phloem immobility which prevents access to key sites of metabolism (Brown and Hu 1996).

B toxicity is difficult to manage by manipulation of agricultural systems, so genetic improvement of the target plant species is the ideal solution. Genetic analysis of B toxicity tolerance in cereal crops such as wheat and barley allowed identification of QTLs, and also the genes underlying these QTLs, so permitting development of closely linked molecular genetic markers (Jefferies *et al*. 1999, 2000; Sutton *et al*. 2007; Reid 2007). In legumes such as *M. truncatula* (Bogacki *et al*. 2013) and lentil (Kaur *et al*. 2014), QTL identification studies have indicated that the plant response to B is subject to control by a single major gene.

1.4 Biotic stresses of field pea

Biotic stresses of field pea include infestation by insects, arachnids, nematodes, parasitic weeds (Solh *et al.* 1994), and infection by pathogens such as bacteria, fungi and viruses (Allen and Lenne 1998). In Australia, the pea crop suffers chiefly from diseases such as blackspot, bacterial blight, powdery mildew, downy mildew, septoria blotch and PSbMV. Blackspot is the most common foliar disease of field pea in Australia, and has traditionally been the main disease-based constraint to production. This disease complex is caused by

up to 4 fungi: mycosphaerella (*Mycosphaerella pinodes* [Berk. & A. Bloxam] Vestergr.), two phomas (*Phoma medicaginis* var *pinodella* [L. K. Jones] Boerema and *Phoma koolunga*), ascochyta (*Ascochyta pisi* Lib.) and macrophomina (*Macrophomina phaseolina* [Tassi] Goid.) (Wang *et al*. 1999, 2007). The mycosphaerella component is by far the most predominant and damaging disease in majority of areas. Bacterial blight is prevalent in areas of the eastern states, but is not considered a major problem in Western Australia. In the southern region of Australia, bacterial blight has the potential to cause yield losses as high as 45% of total production (Murray and Brennan 2012). Moreover, bacterial blight and powdery mildew are the two diseases identified as contributing to the majority of seed quality losses in field pea (Murray and Brennan 2012). Powdery mildew is caused mainly by *Erysiphe pisi* DC, and the disease occurs sporadically when warm humid conditions favour spread of incidence late in the season. Downy mildew (caused by *Peronospora viciae* f. sp. *pisi*) contributes to significant yield losses in field pea crops located in areas of several eastern states, and the disease is sometimes apparent on the older, lower leaves of pea crops in late winter, being favoured by cool, wet growing conditions and occurring most often on early-seeded crops. This disease rarely requires any specific intervention, as crop growth evades infection during longer and warmer spring days (Davidson *et al.* 2004). Sclerotinia (*Sclerotinia sclerotiorum* [Lib.] de Bary and *S. minor* Jagger) may become a problem in wet seasons in which pea crops are sown into paddocks that have been previously cropped with canola, sunflower or chickpea. PSbMV is the major virus that affects field pea, and may cause significant losses in seed yield and quality, especially when infections occur before or during bloom (Provvidenti and Hampton 1991).

1.4.1 Powdery mildew of field pea

Powdery mildew is an important and serious fungal disease that affects a wide range of plant species, including legumes (Panstruga 2005). *Erysiphe pisi* DC is the most common powdery mildew pathogen of pea, but two other distinct *Erysiphe* species (*E. baeumleri*

and *E. trifolii*) have also been reported to cause similar symptoms on pea (Ondřej *et al*. 2005; Attanayake *et al*. 2010). Powdery mildew of pea has a worldwide distribution, being particularly important in climates with warm, dry days and cool nights (Smith *et al*. 1996). The disease adversely affects total biomass, number of pods per plant, number of seeds per pod, plant height, number of nodes and seed quality (Gritton and Ebert 1975), severe infection leading to 25-50% yield losses (Munjal *et al*. 1963; Warkentin *et al*. 1996; Janila and Sharma 2004; Katoch *et al.* 2010).

Powdery mildew affects all of the green parts of pea plants. The characteristic symptoms include a white powdery film on plants and a blue-white colour on severely infected foliage; tissues located below these infected areas may turn purple (Falloon and Viljanen-Rollinson 2001). Severe pod infection can cause a grey-brown discolouration of the seeds. Under favourable conditions, the disease rapidly spreads to adjacent areas. However, heavy rainfall is not conducive to dissemination of the disease, as spores may be washed off the plant tissue.

Control measures for powdery mildew include crop rotation, selection of resistant varieties and strategic use of foliar fungicides. The use of natural products (such as soluble silicon, oils, salts and plant extracts) and biological agents (mycolytic bacteria, mycophagous arthropods, fungi, yeasts and other possible non-fungal biological control agents) to control powdery mildew have received study, but are not yet fully ready for commercial application (Fondevilla and Rubiales 2012).

The genetics of resistance to powdery mildew in pea has been well studied, leading to identification of three major loci, designated *er1*, *er2* and *Er3* (Smykal *et al*. 2012). Large majority of naturally occurring resistant cultivars possess the recessive *er1* gene (Harland 1948; Timmerman *et al*. 1994; Vaid and Tyagi 1997; Janila and Sharma 2004). The recessive *er2* locus was identified in relatively few lines (Heringa *et al*. 1969; Sokhi *et al.* 1979; Kumar and Singh 1981). The *Er3* resistance locus exhibits a dominant inheritance pattern and been recently characterised from a related species, tawny pea (*Pisum fulvum*

Sibth. & Sm.). The two recessive genes, *er1* and *er2,* act in different manners, such that the *er1* gene confers systemic resistance under field and controlled conditions by preventing epidermal cell penetration, so resulting in very few haustoria or colony formation. In contrast, the effects of *er2* are largely confined to leaves, in which expression is influenced by temperature and leaf age, and complete resistance is only expressed at 25°C or in mature leaves. The resistance phenotype conferred by *er2* is mainly based on reduced penetration rate, along with post-penetration cell death in mature leaves (Fondevilla *et al*. 2006; Marx 1986; Tiwari *et al.*1997). The *Er3* resistance locus is known to segregate independently of *er1* and *er2* (Fondevilla *et al*. 2011) and resistance mechanism is due to a high frequency of cell death that occurs both as a rapid response to attempted infection and a delayed response that follows colony establishment (Fondevilla *et al*. 2007, 2008). Expression of *Er3* is not temperature-dependent, and this gene confers complete resistance under both field and controlled environmental conditions (Fondevilla *et al.* 2007).

Trait dissection studies for powdery mildew resistance have permitted detection of various molecular genetic markers (such as members of the restriction fragment length polymorphism [RFLP], randomly amplified polymorphic DNA [RAPD], sequence characterised region [SCAR], SSR, sequence tagged site [STS] and SNP classes) in close linkage with *er1* on LG VI (Dirlewanger *et al.* 1994; Ek *et al*. 2005; Janila and Sharma 2004; Timmerman *et al*. 1994; Tiwari *et al*. 1998; Katoch *et al*. 2010; Tonguc and Weeden 2010; Pavan *et al*. 2013). The *er2* gene was shown to be linked to several amplified fragment length polymorphism (AFLP) and SCAR markers, and was localised on pea LG III (Tiwari *et al.* 1999; Katoch *et al*. 2010). SCAR markers in linkage with the *Er3* gene have also been reported (Fondevilla *et al.* 2008).

Previous literature has reported that in order for successful penetration of the host cell wall to be achieved by the powdery mildew pathogen, the host-encoded MLO protein is absolutely required by many plant species including barley, wheat, rice (*Oryza sativa* L.),

maize (*Zea mays* L.*)*, arabidopsis and grape (*Vitis vinifera* L.) (Jørgensen 1994; Panstruga 2005; Yahiaoui *et al*. 2004; Feechan *et al*. 2009). In the absence of a functional *Mlo* gene, plants are resistant to the fungal pathogen. In pea, a sequence analysis of four field pea accessions revealed that the *PsMLO1* gene provides the functional basis for allelic variation of the field pea *er1* gene (Humphry *et al*. 2011) and that loss-of-function of *PsMLO1* is the direct cause of resistance to the powdery mildew pathogen.

1.4.2 Bacterial blight of field pea

Bacterial blight is an important disease of field pea that is caused by the pathogens *P. syringae* pv. *pisi* and *P. syringae* pv*. syringae*. Bacterial blight due to infection by *P. syringae* pv. *pisi* is a serious disease throughout pea-growing regions of the world. All aerial parts of the plant are susceptible to attack across the life-history of the plant (Hollaway *et al.* 2007). Bacterial blight of field pea causes 'water-soaked' lesions, generally around the leaf base, when bacteria accumulate in water that is trapped for long periods. Initially, the pathogen causes shiny, water-soaked lesions near the nodes and stipules. As the disease progresses, lesions then spread to the stems, peduncles and tendrils and become darker in colour. Disease symptoms usually occur in small patches where bacteria are splashed across to adjoining plants from the site of initial infection. The undersides of leaves and stipules develop water-soaked spots, on the upper surface that appear dark-green to brown in colour. With age, the spots develop an angular shape, with dark margins and a light-coloured, papery appearance in the centre. Lesions appear translucent when held up to the light (Lawyer 1984). The lesions enlarge under humid conditions, and spots of considerable size are sometimes formed. A white- to cream-coloured ooze may also collect on the surface of the lesions (Fig. 1.3) (Rennie and Cockerell 2006). In contrast, if the weather turns dry, the infection may in turn be reversed. The infected tissue of the leaves and stipules eventually turns brown and becomes papery in texture (Harter *et al.* 1934).

Photo from Mathew Rodda, DEDJTR, Horsham

Figure 1.3: Field pea plants infected with (A) *P. syringae* pv. *pisi* race 6 (B) *P. syringae* pv. *syringae* in glasshouse inoculation experiment

Lesions on pods are also water-soaked, and darken as they mature, becoming sunken. The suture area is often a site of infection, and infected seeds may or may not show injury symptoms. Watery, dark spots sometimes appear on the seed, but it is more common for no symptoms to be apparent (Skoric 1927). Seedlings attacked by *P. syringae* pv*. pisi* may not survive. If the infection starts from seed, or if the plants are small, the vine may be killed prior to production of a crop. The earlier the infection, the more serious is the damage to the pea crop (Skoric 1927). Infections that are established later may considerably reduce yield, extent of injury depending largely on weather conditions. Extensively infected pods can harbour the disease pathogen, and act as a source of further infection. The

symptoms of bacterial blight caused by *P. syringae* pv. s*yringae* are difficult to distinguish in the field from those of *P. syringae* pv*. pisi* (Hollaway *et al.* 2007). Under laboratory conditions, however, it is easier to distinguish between the two diseases, especially at the seedling stage.

1.4.2.1 Disease cycle of bacterial blight

P. *syringae* pv. *pisi* overwinters predominantly as a seed-borne pathogen, in which the pathogen can be carried on the seed surface, as well as internally, as a dry, white film of bacteria that is invisible to the naked eye (Skoric 1927). Seed can remain infected for up to three years. During germination, infection proceeds as the plumule contacts the infected seed coat. The lowest three stipules are the most common sites of infection. Soil moisture plays a major role in the infection process, higher levels leading to elevated rates of infection (Skoric 1927; Wark 1954; Roberts 1992; Hollaway *et al.* 1996). Growth occurs between wide temperature ranges (7-37.5^{o}C) with optimal temperatures from 26 to 28^{o}C, and a minimum of 3^{o}C. Rain splash, wind, machinery, animals and irrigation assist the secondary spread of the disease (Lawyer 1984).

P. syringae pv. *syringae* may be a seed- or soil-borne pathogen, which is generally encountered on the surface of the seed coat. As a consequence, the length of the survival of this pathogen on the seed is not as long as for *P. syringae* pv. *pisi.* However, *P. syringae* pv. *syringae* displays a higher capacity to survive in soil than *P*. *syringae* pv. *pisi.* The optimum temperature for pathogen growth is 24°C. Otherwise, the structure of the disease cycle is largely similar to that of bacterial blight caused by *P. syringae* pv. *pisi* (Lawyer 1984). Bacterial blight pathogens can survive on crops without causing any disease symptoms or crop loss. Once conditions are suitable, disease can develop sporadically and hence cause significant crop damage (Armstrong *et al.* 2001). Both bacterial species can survive on pea stubble, while *P. syringae* pv*. syringae* can survive on a variety of host plants. Infection may occur at any stage of plant growth, and is most prevalent following

frosts. Under wet weather conditions, bacteria are spread from infected to healthy plants by rain splash, wind-borne water droplets and plant-to-plant contact. Severe disease can be associated with prolonged periods of rain combined with heavy dews, strong winds and cold temperatures (Roberts 1997; Wade 1951; Knight 1944)

1.4.2.2 Biology of bacterial blight pathogens

As previously established, bacterial blight is a sporadic and destructive disease caused by *P. syringae* pv. *syringae* and *P. syringae* pv. *pisi*. Isolates of the latter are categorised into 7 races based on their differential interactions with a range of pea cultivars (Taylor *et al.* 1989). In Australia, *P. syringae* pv. *pisi* is the most prevalent pathogen, but the incidence of bacterial blight caused by *P. syringae* pv. *syringae* is increasing (Richardson and Hollaway 2011). Race 3 of *P. syringae* pv. *pisi* is most common (64%) followed by race 6 which is the most virulent (Hollaway and Bretag 1995), largely due to the susceptibility of commonly cultivated pea cultivars to races 3 and 6.

P. syringae is a gram-negative, aerobic, non-spore forming, motile organism with a single flagellum or 1-5 polar flagellae (Suzuki *et al.* 2003). On agar media, colonies are round, slightly-convex, greyish-white, translucent, and smooth with undulate margins (Karimi-Kurdistani and Harighi 2008). The LOPAT tests (levan production - positive, oxidase reaction - negative, potato soft rot - negative, arginine dihydrolase - negative and tobacco hypersensitivity - positive) are most commonly used to identify *P. syringae* (Karimi-Kurdistani and Harighi 2008). Different pathovars of *P. syringae* may also be identified by nutritional, biochemical, physiological and nucleic acids-based tests (Hirano and Upper 2000). Both enterobacterial repetitive intergenic consensus (ERIC) sequences and repetitive extragenic palindromic (REP) sequences have been used to design universal PCR primers that generate highly reproducible, strain-specific fingerprints capable of genetic differentiation between pathovars of *P. syringae*. Other molecular methods such as AFLP, multilocus sequence typing (MLST) and RAPD-PCR (based on use of markers

designated as AN3 and AN7) have been used to characterise pathovars of *P. syringae* (Arnold *et al.* 1996; Martin-Sanz *et al.* 2012b).

1.4.2.3 Bacterial virulence factors contributing to disease and pathogenesis

P. syringae is a heterogeneous plant pathogen that causes diseases in >200 different plant species, with symptoms such as blights, leaf spot, and galls (Bender *et al.* 1999; Martin-Sanz *et al.* 2013). The species is sub-divided into 60 pathovars, which differ both in host- and tissue-specificity (Young 2010; Cunnac *et al.* 2009). These bacteria rely upon the direct delivery of pathogen-derived virulence factors in order to cause successful infection *in planta*. The genome sequence of *P. syringae* reveals the presence of genes for multiple protein secretion pathways, some of which have been shown to contribute to quantitative variation of virulence (such as effectors, pectic enzymes, insecticidal toxins, regulatory proteins, and lipoproteins), that are likely to be of particular importance to plant pathogens. The most important pathway is the type III secretion system (T3SS), which is encoded by *hrp*/*hrc* genes and is required for elicitation of the hyper-sensitive response (HR) in non-host or resistant host plants and for pathogenesis in susceptible plants (Losada and Hutcheson 2005; Ramos *et al.* 2007). Plant-pathogenic *Pseudomonas* species deliver about 30 effectors into the plant (Hou *et al.* 2009). The other important virulence factors for *P. syringae* infection include phytotoxins, adhesins, exopolysaccharides (EPS), cell wall-degrading enzymes and regulators of phytohormone production (O'Brien *et al.* 2011).

1.4.2.4 Factors influencing disease development

The severity and prevalence of disease symptoms depend largely on weather, crop rotations, and presence of disease in the field. Any type of mechanical injury to crops such as by frost, wind-abrasion, hail, damage by wheel tracks, herbicide application, water-logging etc. will lead to higher susceptibility to bacterial blight infection. A high level of atmospheric humidity is also necessary for disease development in the absence of frost

(Wark 1954). Increased humidity within intercellular spaces following frosting of the plant creates a micro-environment that is favourable for the growth and multiplication of the bacteria (Scarth 1944). Rupture of cells, which occurs due to frosting, provides a nutrient source for the bacteria. Early infection, combined with prolonged wet, cold and windy conditions that favour disease spread, can greatly increase the severity of pathogen damage. Soil moisture content and soil type also influence disease development (Wark 1954). Disease progresses more rapidly in soils with high moisture content than at lower moisture content following exposure to high atmospheric humidity (Roberts 1992). Infection rates as low as one infected seed in 10,000 (0.01%) are sufficient to give rise to diseased seedlings, even in a relatively dry seed-bed, which would be sufficient to cause an epidemic (Taylor and Dye 1976). Disease transmission from seed to seedling is also influenced by the quantity of inoculum. Roberts (1997) investigated the effect of weather conditions on simultaneous local spread of bacterial blight, and the results demonstrated the importance of heavy rainfall and high wind in dispersal of the disease. Wade (1951) reported frost to be the favoured condition for disease development, while Knight (1944) observed spread of the disease following hail-storms.

1.4.2.5 Effects of bacterial blight infection on production

Annual losses to pea production due to bacterial blight infection vary from year to year, depending upon local weather conditions. Severity of bacterial blight disease varies greatly from crop to crop (Fig. 1.4), and between seasons. Severe epidemics occur once in every 10 years, and can cause some crops to fail completely. Bacterial blight caused by *P. syringae* pv. *syringae* and *P. syringae* pv. *pisi* is responsible for a present loss of 10% or more in Australia, but has the potential to cause yield losses of 45% or more in the southern region of Australia (Murray and Brennan 2012).

Photo from Peter Kennedy, DEDJTR, Horsham

Figure 1.4: Pea cultivation in the field in Horsham, Victoria during 2012, showing the susceptible cultivar Kaspa infected by bacterial blight

As a seed-borne disease, bacterial blight often strikes early in plant growth, causing poor crop establishment and reduced plant vigour, so resulting in lower yields due to plant death, crop damage and production of small seeds. Extensive infection in pods considerably reduces market value due to blemishes, unsightly appearance and reduced seed size. Low quality of seeds, as well as the demands of phytosanitary regulation in some countries, excludes Australian growers from certain export markets due to a requirement for 'zero-*Pseudomonas syringae* in seed' (Hollaway *et al.* 2007).

Glasshouse-based studies by Roberts (1993) and Roberts *et al*. (1995) using plants artificially inoculated with *P. syringae* pv. *pisi* revealed reductions in seed yield of 24, 47, or 71% due to bacterial blight infection during reproductive, vegetative or both reproductive

and vegetative growth stages, respectively. The yield reductions were observed as reduced number of seeds per pod, combined with changes in the number of pods per plant. Infection before flowering reduced the number of pods formed, as compared to after flowering, which led to reduced pod-filling ability (Roberts 1993).

1.4.2.6 Dissemination of bacterial blight

Disease can be carried from one place or plant to other by various means, such as invertebrate and bird vectors, infected seeds, machinery etc. It is virtually impossible to determine the extent of disease dissemination through the action of insects such as aphids, plant lice, thrips and leaf hoppers. Wounds formed by insects that feed on plants provide favourable sites for infection to initiate. An estimated 90% of the world's food crops are grown from seeds (Schwim 1994), and the uncontrolled movement of infected seed between regions can result in rapid expansion of areas affected by such diseases. However, for many pathogens, this is an efficient means of transmission into new areas, as well as means of survival between seasons (Rennie and Cockrell 2006). Many serious diseases of pulses, including bacterial blight, can be seed-borne pathogens, and may be carried by the seed either internally or externally. Extent of pathogen transmission from seed to crop can vary considerably depending on growing conditions.

1.4.2.7 Control measures

Although bacterial blight cannot be cured once a crop has become infected, the risks of contracting disease and subsequent damage can be greatly reduced through judicious choice of variety, minimisation of the inoculum source (infected seeds) and good management practices. There are a number of ways to reduce the incidence of the infection and, hopefully, impact on crop growth, such as early sowing, paddock selection, crop rotation and the use of pathogen-free seeds (Hollaway *et al.* 2007).

Planting of resistant cultivars can also contribute to reduction of disease impact. The incidence of bacterial blight can be reduced by sowing of pathogen-free seeds, either from pre-testing, or material obtained from fields that have been free of bacterial blight during the past few seasons (Armstrong *et al.* 2001). Seed production should be limited to more arid areas, and irrigation should be avoided. In addition, cultivation should not be performed in paddocks that contain sulfonylurea residues or are more prone to frost. Selection of paddocks with low weed incidence, in order to minimise herbicide usage, may therefore reduce the damage to crop plants. Machinery should be disinfected between field applications, and spraying of pesticides should be performed by air instead of through use of ground-based applicators, in order to avoid mechanical injury. In-crop hygiene should be observed, such as disinfection of equipment after entry into a bacterial blight-infected field. For the *P. syringae* pv. *syringae* pathogen, however, seed storage for one year prior to planting should be sufficient to virtually eliminate the pathogen. Infected stubble can be a potential source of inoculum, and so must be buried, burned or destroyed to eliminate further spread (Armstrong *et al.* 2001). Prevalence of bacterial blight can be reduced by avoidance of sowing on land that has sown peas in the previous year, or is adjacent to pea stubble, and also by avoiding growth on the same land more than once in three years. Early-sown crops are more vulnerable to bacterial blight than late-sown crops, and so in order to reduce the likelihood of infection, sowing at the later end of the recommended period for a particular area is desirable. Once bacterial blight is detected in the field, appropriate measures should be taken to prevent the spread of disease. If possible, infected crops should be harvested last, in order to avoid contamination of healthy crops, and machinery that has been used on an infected crop should be thoroughly cleaned. In addition, machine operators and farm workers should wear appropriate clothing, and only migrate from crop to crop after taking precautions against the spread of bacteria. The incidence of bacterial blight may also be reduced by application of seed treatment chemicals or by sowing of healthy seed, in order to decrease the seed-borne inoculum. Legislation in many countries is aimed at prevention or limitation of the

introduction of non-indigenous seed-borne pathogens (Ebbels 2003), and national seed certification schemes may provide standards for the most damaging seed-borne diseases. Crop monitoring, especially in areas which have been exposed to environmental stresses such as frost, waterlogging and chemical damage, will facilitate more rapid identification of disease incidence. If detected, seed from the crop should not be used for subsequent plantings (Rennie and Cockrell 2006).

1.4.2.8 Genetics of bacterial blight

Taylor *et al.* (1989) explained the resistance of 7 commercial pea cultivars to *P. syringae* pv. *pisi* races 1, 2, 3, 4 and 5 in terms of gene-for-gene relationships involving different matching gene pairs. Genetic analysis indicated that up to six gene pairs are involved (Bevan *et al.* 1995; Hunter *et al.* 2001). Resistance to race 6 has been identified only in *P. abyssinicum* accessions and a *P. sativum* landrace from Spain (Schmit *et al.* 1993; Elvira-Recuenco and Taylor 2001). Recently Martin-Sanz *et al.* (2011) proposed the presence of an eighth race. The genetic transmission studies showed that *R1* [*Ppi1*], *R2* [*Ppi2*] and *R3* [*Ppi3*] are the dominant resistance alleles at different single loci, while *R4* [*Ppi4*] was found to be a dominant allele at a single locus with expression that may be altered in certain genetic backgrounds. Inheritance of *R5* and *R6* could be due either to recessive alleles at two further loci, or dominant alleles at loci linked to *R2* and *R4* respectively (Bevan *et al.* 1995; Hunter *et al.* 2001). For *P. syringae* pv. *syringae* resistance, QTL analysis study identified quantitative variation for resistance, through detection of two QTLs accounting for 22% and 8% of the phenotypic variation (Vp), respectively (Fondevilla *et al*. 2012). A number of pea lines exhibited resistance to *P. syringae* pv. *syringae* both in growth chamber-based studies and under field conditions (Martin-Sanz *et al.* 2012a; Richardson and Hollaway 2011) but more tests need to be conducted under different experimental conditions in order to determine the value of resistance in those lines to breeding programs.

1.5 Plant innate immunity: insight into defence mechanisms

Interactions between plants and pathogens are governed by a series of well-understood mechanisms that involve signal activation, frequently resulting in rapid defence responses. These responses help the host plant to avoid further disease infection.

Plants exhibit two major types of disease resistance, termed basal defence and resistance (R)-gene mediated defence. Basal defence, which can be a constituent of both non-host and host-specific resistance, provides the first line of defence protection against infection by a wide range of pathogens (Mysore and Ryu 2004). In contrast, plant disease resistance which is attributable to a particular cultivar or accession is known as host-specific resistance. Most of the various mechanisms of non-host resistance are regulated by multiple genes, unlike host-specific resistance which is most often controlled by segregation of a single gene (Cunha *et al.* 2006).

As many as 70 different R genes showing resistance to major plant pathogens, including bacteria, fungi, nematodes, insects and oomycetes, have been isolated, cloned, and characterised from different plant species in the last 15 years (Dangl and Jones 2001; Meyers *et al.* 2003; Liu *et al.* 2007). Sequence comparisons between such R genes have revealed structural similarities, even though overall sequence homology is low. The conserved domains participate in protein-protein interactions and signal transduction, and provide opportunities for isolation in other plant species based on PCR amplification using degenerate oligonucleotide primers (Meyers *et al.* 1999; Kobe and Deisenhofer 1994; Meyers *et al.* 2003). Based on these conserved motifs, R genes have been grouped into several classes (Fig. 1.5).

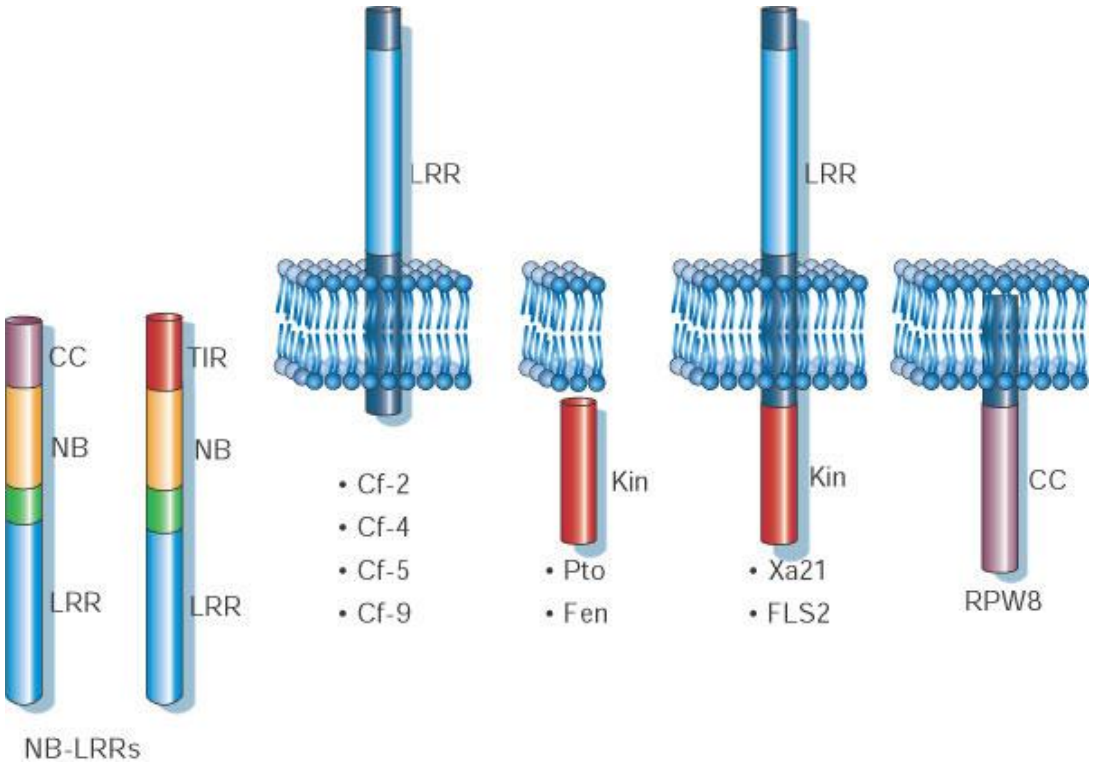

Figure 1.5: Representation of the location and structure of the five main classes of plant disease resistance proteins (Dangl and Jones 2001)

The largest class of R genes encodes proteins with a nucleotide-binding site (NBS) and a Leucine-rich repeat (LRR) region. The second class, represented by the *Cf* genes of tomato (*Cladosporium fulvum* avirulence determinants - AVR4 and AVR9), encode transmembrane receptors with an extracellular LRR domain and an intracellular serine-threonine kinase activity. The third class, which encodes a serine-threonine protein kinase domain, includes the *Pto* gene from tomato (*P. syringae* strains carrying *avrPto)*. The class four includes genes such as rice *Xa21* (*Xanthomonas oryzae* pv. *oryzae* upon recognition of a small pathogen-encoded protein), which encode an extra-cytoplasmic LRR domain and an intracellular serine/threonine kinase domain. Class five includes genes such as *RPW8* (which confers broad-spectrum resistance to powdery mildew pathogens) that encode a small, probable membrane-located protein with a possible coiled-coil domain, and essentially no other similarity to other known proteins (Bent 1996; Dangl and Jones

2001; Hammond-Kosack and Jones 1997). Table 1.2 provides a comprehensive summary of current data on effector molecules, their sources, targets, corresponding R-proteins and R-protein structure as relevant to *P. syringae*, the major pathogen of interest to this study (Muthamilarasan and Prasad 2013).

There are significant benefits from the use of disease resistance genes in plant breeding programs, such as efficient reduction of pathogen growth, minimal damage to the host plant, zero input of pesticides from the farmers, and most importantly, environmental sustainability. However, in the case of conventional breeding practices, introgression of resistance genes from one species into the gene pool of another by repeated back-crossing is a long-term process which usually requires many hybrid generations. The application of functional genomics tools such as cloning, characterisation and genetic transformation of plant resistance genes could assist researchers to overcome such problems (Chen *et al.* 2002; Nimchuk *et al.* 2001). Breeding of resistant varieties protected by single R genes of large effect is a simple strategy, which will eventually lead to presence in a larger number of varieties, providing a major source of genetic protection against infection. However, this can easily lead to catastrophic collapse, in which breakdown of R genes leads to susceptibility of many varieties. However, use of a combination of R genes could enhance the longevity of resistance, although implementation of such a strategy (by gene 'pyramiding') is logistically challenging (Singh 2012).

Table 1.2: Effector molecules, their sources, targets, corresponding R-proteins and R-protein structure (Muthamilarasan and Prasad 2013)

Pathogen	**Effector**	**Target**	**R-protein**	**R-protein structure**
Pseudomonas syringae	AvrPtoB	FLS2, BAK1, FEN, CERK1	Prf	CC-NB-LRR
Pseudomonas syringae	AvrPphB	PBS1	RPS5	CC-NB-LRR
Pseudomonas syringae	AvrRpt2	RIN4	RPS2	CC-NB-LRR
Pseudomonas syringae	AvrRpm1	RIN4	RPM1	CC-NB-LRR
Pseudomonas syringae	AvrB	RIN4	RPG1-B/TAO1	TIR-NB-LRR
Pseudomonas syringae	AvrRps4	unidentified	RPS4	TIR-NB-LRR
Pseudomonas syringae	HopAI-1	MPK3, MPK6	unidentified	unidentified
Pseudomonas syringae	HopU1	GRP7	unidentified	unidentified
Pseudomonas syringae	HopM1	AtMIN7	unidentified	unidentified
Pseudomonas syringae	HopAO1	Downstream of PTI signalling	unidentified	unidentified
Pseudomonas syringae	HopO1–1	Chloroplast protein	unidentified	unidentified
Pseudomonas syringae	HopO1–2	Chloroplast protein	unidentified	unidentified
Pseudomonas syringae	HopU1	AtGrp7, AtGrp8	unidentified	unidentified

1.6 Molecular genetic markers

A molecular genetic marker is associated with variation of a specific DNA sequence with a known location within a genomic region on a particular chromosome, which may be related to diversity of a particular gene or phenotypic trait. Such variation may arise due to mutation or structural alteration in the relevant genomic locus. A molecular genetic marker may arise from variation of a short DNA sequence, such as a single base-pair change (leading to a SNP), or a longer sequence, such a variable-number tandem repeats (VNTRs), commonly known as 'mini-satellites' (Vignal *et al.* 2002; Wright and Bentzen 1994).

The exploitation of molecular genetic markers has revolutionised the pace and precision of plant genetic analysis, leading to significant increases in knowledge. The last three decades have seen major advances in the evolution of genetic marker systems and the respective analysis and detection platforms. Molecular genetic markers have been extensively used to prepare saturated genetic linkage maps and to identify association with genes or QTLs that control traits of economic importance. Other uses of molecular markers include: assessment of gene introgression through backcrossing strategies; intensive germplasm characterisation; genetic diagnostics; characterisation of transgenic plants, including detection of site of genomic integration; studies of comparative genome organisation; phylogenetic analysis; and others (Gupta *et al.* 2002).

The evolutionary development of molecular markers has been primarily driven by considerations of throughput, appropriate cost of detection method, and requisite level of reproducibility. The majority of early molecular genetic marker-based studies for agronomic plant species were based on the use of RFLP, RAPD and AFLP loci (Botstein *et al*. 1980; Williams *et al*. 1990; Vos *et al.* 1995). SNPs and SSRs provide the most reliable and efficient current forms of molecular genetic marker technology. SSRs have been termed 'second-generation' molecular markers, while SNPs provide the present, 'third-

generation', of markers (Gupta *et al.* 2001). SSRs containing tandemly-repeated di-, tri-, or tetranucleotide units, which can be identified in either coding or non-coding regions of DNA, are ubiquitous in eukaryotes (Gupta *et al.* 1996), and abundant in plant genomes. The number of repeat units present at a particular genomic site is the variable that is assayed by PCR, which may vary substantially between individuals (Jeffreys 1990; Gupta and Varshney 2000). SSR analyses are generally performed by fluorescence detection following capillary sequencing, allowing automated data transfer and analysis using specialised programs (Gupta *et al.* 2002). SSRs originate from genomic processes such as unequal sister-strand chromatid exchange (USCE), or DNA polymerase slippage during replication (Pearson and Sinden 1998). If the resulting tandem repeats are located within the coding regions of specific genes, the structure of the resulting protein and its function may be altered. Due to constraints on protein structure, the majority of SSR expansions are, however, located either within non-transcribed regions of gene units, such as promoters, 5'- and 3' untranslated region and introns, or intergenic non-coding regions (Liu *et al*. 2013).

SSRs have the advantages of being highly reproducible, highly polymorphic, abundant and co-dominant in nature. They require only a small quantity of DNA template for PCR in order to detect polymorphisms, and are amenable to automation (Smith *et al.* 1997). Because of these properties, SSRs provide suitable markers for use in breeding programs to allow screening of a large number of breeding lines. In addition, SSRs are valuable for a broader range of applications, such as genome mapping, trait-dissection, and a range of molecular ecology and diversity studies (Robinson *et al.* 2004). With the increase in the availability of DNA sequence information, automated processes to identify and design PCR primers for amplification of SSR loci are valuable tools. Genomic DNA-derived SSRs and expressed sequence tag-derived SSRs (EST-SSRs) provide complementary desirable and undesirable properties for genetic analysis. EST-SSRs are readily identified from transcriptome datasets, and due to the conservation of coding sequences, are generally

less prone to null allele amplification arising from primer binding site mutations. Genomic DNA-derived SSRs are highly polymorphic in nature and tend to be widely distributed throughout the genome, resulting in better map coverage (Wu *et al.* 2011). However, discovery methods for genomic DNA-derived SSRs may require generation and interrogation of larger genomic DNA-based sets, and primer design to hypervariable regions may be problematic. SSR markers developed for one species generally exhibit less transferability across same or different taxa which require the development of species specific markers (Roa *et al.* 2000; Kindiger 2006), and also have much higher rates of mutation compared to nucleotide substitution and therefore suffers from homoplasy (Cieslarova *et al*. 2011).

SNPs and insertion/deletion (indel) events are the most commonly occurring type of DNA sequence variation, arising either due to identity changes at a single nucleotide position in the genome, or the presence and absence of a small number of nucleotides (typically 1-2). SNPs may fall within coding sequences of genes, non-coding regions of genes, or in the intergenic regions between genes. SNPs within a coding sequence will not necessarily change the amino acid sequence of the protein that is produced, due to the degeneracy of the triplet genetic code, as mutational changes may give rise to synonymous amino acid substitutions. Synonymous changes can influence gene expression, messenger RNA (mRNA) conformation and subcellular localisation of mRNAs and/or proteins (Shastry 2009; Sauna and Kimchi-Sarfaty 2011).

SNP markers are capable of development for the majority of target genes, and may be used as specific genetic markers for the detection of associations between genotype and phenotype, and as diagnostic markers to identify genotypes with superior allelic content (Rafalski 2002). Use of SNP markers has rapidly gained acceptance during recent years due to high genomic abundance and amenability to high-throughput detection formats and platforms. Moreover, computational methodologies dominate SNP discovery, due to the

enormous amount of sequence information that is available in public databases (Jones *et al.* 2007).

The emergence of next-generation sequencing (NGS) technologies such as HiSeq (Illumina, San Diego, CA), SOLiD (Life Technologies Corporation, Carlsbad, CA) and Ion Torrent (Life Technologies Corporation, Carlsbad, CA) has mitigated the problems associated with low rates of throughput and high costs associated with SNP discovery. In particular, transcriptome resequencing using NGS technologies allows rapid and inexpensive SNP discovery within genes, and avoids sampling of highly repetitive genomic regions, which are very prevalent in many higher plant species (Mardis 2008; Morozova and Marra 2008). As an alternative to transcriptome analysis, genome complexity may be reduced by approaches such as AFLP, high C_0t selection (in which genomic fractions that rapidly reassociate after denaturation, such as highly and moderately repetitive components, are selectively subtracted), methylation status-based filtration, restriction site-associated DNA sequencing (RAD-seq) and multiple matching oligonucleotide-based sequence capture (Vos *et al.* 1995; Yuan *et al*. 2003; Thompson *et al*. 2012; Baird *et al*. 2008; Clarke *et al*. 2013). Such techniques are effective for reducing repetitive sequence content, but may be unable to eliminate duplicated sequences, which can cause detection of false-positive SNPs (Mammadov *et al.* 2012).

A large number of specific genotyping technologies have been developed for detection of SNPs. Some of the widely used genotyping platforms include BeadXpress™, GoldenGate™ and Infinium from Illumina (http://www.illumina.com); GeneChip™ and GenFlex™ Tag array from Affimetrix (http://www.affymetrix.com); SNaPshot™ and TaqMan™ from the Applied Biosystems (http://www.appliedbiosystems.com); SNPstream™ from Beckman Coulter (http://www.beckmancoulter.com) and KASP™ from LGC (http://www.lgcgroup.com). Genotyping-by-sequencing (GBS) is a rapidly emerging method for SNP genotyping which is being increasingly adopted for several crop species (Liu *et al.* 2014; Sonah *et al*. 2013). However, GBS methods require computational

expertise and significant time for data analysis. Nonetheless, current developments in computer software packages and computational pipelines will enable wider use of GBS in multiple plant species in the future (Sonah *et al*. 2013).

1.7 Construction of genetic linkage maps and trait dissection based on QTL analysis

The main objective of plant breeding programs is the improvement of characteristics that will increase benefits for farmers and industries. Conventional breeding activities have contributed considerably to crop improvement, but have been slow to effectively target complex traits like abiotic stress, grain yield and grain quality (Kumar *et al.* 2011). A detailed understanding of mechanisms of gene inheritance, location on chromosomes and systems of regulation will provide important information for breeders.

One method for organisation of positional information for genes is generation of a linkage map, which determines the relative positions of, and distances between, genetic loci or DNA-based markers across multiple LGs based on recombination values obtained from mapping pedigrees (Jones *et al.* 1997; Semagn *et al.* 2006). Genetic linkage map construction requires both provision of an appropriate, generally biparental, mapping population and polymorphic molecular genetic markers that are used to genotype the mapping population (Semagn *et al.* 2006).

1.7.1 The mapping population

One of the most critical decisions in constructing a linkage map is the mapping population. The first essential step is selection of divergent parents that exhibit clear genetic differences for one or more traits of interest. However, the parental genotypes should ideally not be so divergent as to cause sterility in progeny sets, and hence segregation distortion during linkage analysis (Semagn *et al.* 2006). Such issues were particularly

observed in early linkage map construction studies, which frequently exploited interspecific crosses between closely related taxa, in order to maximise genetic polymorphism (Vallejos and Tanksley 1983; Paran *et al.* 1995). Sizes of mapping populations typically range from 50 to 250, a larger population size contributing to higher levels of map resolution (Collard *et al.* 2005). Different types of mapping populations such as F_2- or F_x- derived families, backcross (BC) families, recombinant inbred lines (RILs), double haploids (DH) and nested association mapping (NAM) populations can be utilised for linkage map construction in plants (Collard *et al.* 2005; Manikanda 2013). Each type of mapping population, when developed from inbred parents, provides distinct advantages and disadvantages. Such factors largely depend on project objective, trait complexity, available time for construction, and whether the molecular genetic markers to be used for genotyping are dominant or co-dominant in nature. Another type of mapping population, the near isogenic lines (NILs), cannot be used for linkage map construction but are most useful for tagging traits and fine mapping (Monforte and Tanksley 2000; Keurentjes *et al*. 2007). F_2 and BC populations are the most straightforward type of mapping families to generate, due to ease and speed. However, these population types suffer the disadvantage of residual heterozygosity, and inability to propagate for many generations through seeds (Manikanda 2013). RILs, NILs and DHs, in contrast, are permanent or 'immortalised' populations, as constituent genotypes are homozygous lines that can be multiplied and reproduced in the absence of genetic alteration. Moreover, seeds of each individual homozygous line may be transferred between different laboratories for mapping purposes. The major limitations of NIL and RIL populations include the long lead-time required for and costs associated with their development, as well as limited genetic diversity (Bergelson and Roux 2010). In comparative terms, use of a NIL population offers advantages over RILs for detection of minor QTLs, which are often missed in RILs (Keurentjes *et al.* 2007). DH populations are capable of more rapid generation than RILs and NILs, and have been constructed for agronomically important crops such as rice, maize, rapeseed, wheat, and barley. However, haploid induction in these and other plants usually requires tissue culture of microspores,

ovules, or rescued haploid embryos following distant hybridisation (Dunwell 2010). Such cell-culture based methods are highly influenced by genotype and/or environmental factors. Practical usefulness is consequently limited by the technical difficulty of producing a large doubled haploid population in a desired genetic background, even within a species (Schneider 2005). Information derived from biparental mapping populations based on homozygous parents is inherently limited, as logistical constraints on population size restrict resolution to QTL-containing intervals which may span several centimorgans (cM), and could correspond to the location of hundreds of genes.

Association mapping is an alternative approach for QTL mapping which exploits the properties of natural populations and collections of landraces, breeding materials and varieties (Risch and Merikangas 1996), which offer a broader range of genetic diversity. Association mapping provides advantages over linkage mapping due to higher resolution, capacity to survey a larger number of alleles, and time-saving in establishing a marker-trait association and its application in a breeding program (Flint-Garcia *et al*. 2003). In the recent past, there has been substantial interest in association genetics, due to increasing capacity for high-throughput gene discovery and polymorphism detection based on efficient genotyping. In association mapping, it is important to consider population structure and kinship among individuals, because false associations may be detected due to the confounding effects of population admixture (Soto-Cerda and Cloutier 2012).

1.7.2 Computer software for genetic mapping and trait dissection based on QTL analysis

Several programs are available for use in linkage mapping exercises, such as MAPMAKER (Lander *et al.* 1987), JoinMap (Stam 1993), CarthaGène (de Givry *et al*. 2005), MST_{MAP} (Wu *et al*. 2008), Map Manager QTX (Manly 2001) and R/qtl (Broman *et al*. 2003). Most of these programs support dominant, co-dominant and a mixture of marker types.

The linkage assignments in programs are established by considering all estimates of recombination frequencies. The markers are assigned to a particular LG on the basis of assessment of logarithm of odds (LOD). A significant LOD score (generally >3) indicates rejection of the null hypothesis that two loci are unlinked, and hence *prima facie* evidence for genetic linkage. During map construction, genotypes of parents and progeny for multiple loci are obtained, and an iterative process based on two-point and three-point genetic mapping is algorithmically implemented through use of a mapping function (Haldane, Kosambi or Morgan) to convert recombination frequencies (RFs) into map units, typically expressed as cM units, which correspond to an RF of 0.01 or 1%. The overall length of the resulting linkage map, and arrangement of individual loci within the map, are influenced by a multiplicity of factors including frequency of recombination, environmental effects, number and type of loci mapped, choice of mapping software, mapping population size, type and their generation used (Liu *et al*. 2008; Knox and Ellis 2002; Sim *et al*. 2012; Shirasawa *et al*. 2013).

For successful identification of genomic regions (and ultimately, of the underlying genes) that control agronomic traits of interest, sufficient genetic variation within the mapping population, a well-saturated genetic linkage map derived from the population, and a reliable phenotypic screening method are prerequisite. Using these tools, QTL mapping is performed based on the position of markers on the linkage map through integration of genotypic and phenotypic data. A QTL is defined as a chromosomal region that contains a gene or genes for which allelic variation contributes to a proportion of trait variation that is measured on a quantitative scale. Three analytical methods that are widely used for QTL detection are single marker regression (SMR) analysis, simple interval mapping (SIM) and composite interval mapping (CIM) (Liu 1998; Tanksley 1993). SMR identifies any statistically significant associations between allelic variation for a given marker locus and levels of trait expression to identify a QTL in the vicinity of that marker locus. In contrast, SIM performs a systematic scan of the whole genome for evidence of QTLs, locating a

position between two marker positions. SIM uses a map function (either Haldane or Kosambi) to translate from recombination frequency to genetic distance or *vice versa*, and calculates the LOD score at each increment in the interval. The LOD score profile for the whole genome is subsequently calculated, and when a threshold value is exceeded, QTLs are declared at that location, with appropriate confidence limits. CIM adds information on background loci to SIM and fits parameters for a target QTL in one interval while simultaneously fitting partial regression coefficients for background markers to account for variance caused by non-target QTLs. The major advantage of CIM over SMR and SIM is precision and resolution, especially when closely linked QTLs are involved. Many researchers have used the QTL Cartographer (Basten *et al.* 1994, 2002), and MapManager QTX (Manly *et al.* 2001) applications to perform CIM.

Mapping of major genes is relatively simple as they are inherited in a Mendelian manner and their allelic forms instruct distinct qualitatively variable phenotypes. In contrast, continuously variable traits, for which phenotypes typically vary along a continuous gradient depicted by a bell curve, are not readily explained by Mendelian inheritance patterns (Mendel 1866). Quantitative traits are complex as they are controlled by multiple QTLs, such that the same phenotype may result from different alleles at each of the many loci; plants with identical QTL genotypes may show different phenotypes when assessed in different environments; and the effect of a given single QTL can depend on the allelic constitution of the plant at other QTLs (Semagn *et al.* 2010). For these reasons, it is not possible to unambiguously infer QTL genotype from the observed phenotype. The number of QTLs detected in a given study depends on different factors, including type and size of mapping population used, trait under evaluation, number of environments employed for phenotyping, and genome coverage.

Apart from support for QTL analysis, genetic linkage maps also provide essential resources for comparative studies and genome sequences of other plant species, which may be valuable for prediction of candidate genes. Even for those plant species for which

genomes are yet to be sequenced, a linkage map can contribute to the process of sequence assembly and validation. Finally, information coordinated in the form of a linkage map provides vital tools for the use of marker-assisted selection in plant breeding programs (Semagn *et al*. 2006).

Numerous genes associated with disease resistance have been tagged with molecular genetic markers for selection in a broad range of crops (Brar 2002). For example in rice, the bacterial blight resistance genes, *Xa4*, *Xa5*, *Xa13* and *Xa21* were successfully used for gene pyramiding using molecular markers (Brar 2002). Disease resistance genes have been isolated from rice, wheat, tomato and barley through map-based cloning (Komori *et al*. 2004; Keller *et al*. 2005; Wing *et al*. 1994; Buschges *et al*. 1997). This process required high-density genetic maps and the capacity to perform chromosome walking on large genomic fragments.

1.8 Conserved synteny analysis in legumes

The term 'conserved synteny' refers to a set of conserved genomic features (genes or other genetic loci) that are in the same relative order when compared between chromosomal locations in a pair of species. Such analysis is important in order to reveal genomic conservation of related species, and for studies of evolutionary chromosomal dynamics (Devos and Gale 1997; Gale and Devos 1998). Conserved synteny is derived from the descent of contemporary taxa from a common ancestor, and the orthologous genes located within syntenic blocks that are predicted by the analysis are generally inferred to share similar functions.

M. truncatula is the closest model species in taxonomic terms to cool-season legume food crops such as pea, lentil, faba bean, and chickpea. Comparative synteny analysis between the genomes of *M. truncatula*, *L. japonicus* and soybean revealed evidence for an ancient WGD event within a common ancestor that occurred c. 58 Myr ago (Doyle and Luckow

2003; Young *et al.* 2005; Cannon *et al.* 2006). Synteny analysis between pea and the sequenced legume genomes (*M. truncatula*, *L. japonicus*, and soybean), as well as linkage maps of other legumes, allowed evaluation of macrosynteny and confirmed significant colinearity between species (Aubert *et al.* 2006; Kaló *et al.* 2004; Choi *et al.* 2004a). The availability of genomic resources for the model species (*M. truncatula* - *Medicago Truncatula* Genome Project [http://jcvi.org/medicago/]*; L. japonicus* - Sato *et al.* 2008) and also including, chickpea (Varshney *et al.* 2013) and pigeon pea (Varshney *et al.* 2011), has created more opportunities for comparative genetic analysis in crop legumes.

To understand complex biological process in plants, model species have been extensively utilised, and this process has been highly assisted by knowledge of the extent of genomic conservation. Comparative genetic analysis is consequently an effective approach for sharing genetic and genomic information between plant species, allowing sequence information and orthologous markers to be transferred from the model legumes to other closely related species. This approach was successfully demonstrated in a study by Mudge *et al.* (2005) in which two large soybean genomic regions (containing the soybean cyst nematode [SCN] resistance loci *rhg1* and *Rhg4*) exhibit conserved synteny with both the *M. truncatula* genome and a network of segmentally duplicated regions in *Arabidopsis thaliana*. Using comparative genetic approach, the orthologous loci of *EARLY FLOWERING GENE 3* (*ELF3)*, controlling the photoperiod responses in pea were identified in lentil (Weller *et al*. 2012). Pottorff *et al.* (2012) demonstrated the use of integrated genomic resources for the identification of candidate genes for leaf morphology in cowpea, through extrapolation of the sequences of a trait-linked genetic marker onto model legume genomes in order to confirm conserved synteny and permit candidate gene prediction. Recently, synteny-based candidate gene discovery has been practised in faba bean for the isolation of tannin gene (Webb *et al*. 2015), in cowpea for aphid resistance gene (Huynh *et al.* 2015) and in lentil for identification of candidate gene associated with boron tolerance (Kaur *et al.* 2014).

1.9 Gene expression analysis

Gene expression is a complex process in which information from a gene is used for the synthesis of functional gene products, which are usually proteins. However, in non-protein coding units such as transfer RNA (tRNA) or small nuclear RNA (snRNA) genes, the product is a functional RNA. Transcriptomics is the analysis of the full array of coding and non-coding RNAs and their expression profiles (Storz 2002). Transcriptome profiling was one of the first systems biology approaches to be developed, and is more tractable and sensitive than other '-omics' technologies (Lockhart and Winzeler 2000; Shi *et al.* 2013). The transcriptome provides a 'snap-shot', not only of genes that are active at a particular time or place during development of an organism, but also all isoform sequences which are produced through alternative splicing and variant alleles within the cells (Ward *et al.* 2012). Differences in gene expression during developmental stages and between different tissue types (Stelpflug *et al.* 2015; Moreau *et al*. 2011; Verdier *et al.* 2013a, 2013b), and also during environmental changes (in response to abiotic and biotic stresses) (Shu *et al*. 2015; Goyer *et al*. 2015), can be studied utilising appropriate approaches for producing transcript libraries and analysis tools.

Due to advances in NGS technologies, massively parallel sequencing of RNA has emerged as a powerful and cost-effective technique that can be used in the absence of genome information to obtain gene expression data (Zhao *et al.* 2011). One fundamental application of RNA-Seq is inclusive characterisation of the transcriptome of an organism (Wolf 2013), which generates genomic resources for a broad range of applications. In RNA-Seq, RNA (total or fractionated, such as the poly[A]+ fraction) is converted to a library of cDNA fragments with adaptors attached to one or both ends. After amplification, each molecule is sequenced in a high-throughput manner (using sequencing technologies such as Illumina, Applied Biosystems SOLiD and Roche 454 Life Science) to obtain short sequences from one end (single-end sequencing) or both ends (paired-end sequencing)

with read lengths from 30-400 bp, depending on the technology that has been used (Wang *et al*. 2009).

Two approaches may be used to define transcripts from the derived sequencing data: the mapping-first approach and the assembly-first (*de novo*) approach (Grabherr *et al*. 2011). The first approach requires a genome reference to which reads can be mapped, which in principle provides maximum sensitivity. However, this method may be complicated by the presence of splice variants, sequencing errors and incomplete status of the reference. The most commonly used programs for mapping are TopHat-Cufflinks (Trapnell *et al*. 2012), Bowtie (Langmead *et al*. 2009) and Scripture (Guttman *et al.* 2010). The *de novo* assembly approach is essential in the absence of a reference sequence. The common computational programs for *de novo* assembly are Velvet (Zerbino and Birney 2008), Oases (Schulz *et al*. 2012), ALLPATHS (Butler *et al*. 2008), SOAPdenovo-Trans (Xie *et al*. 2014), Trans-ABySS (Robertson *et al*. 2010) and Trinity (Grabherr *et al*. 2011; Haas *et al*. 2013). Both mapping and *de novo* assembly approaches display particular advantages. *De novo* assembly is independent of the reference data set, and so is capable of discovering transcripts that are missing from the reference assembly. However, the mapping-based methods are relatively faster, capable of filtering out sequencing errors, and may also be used for filling of gaps in comparison to the reference sequence, resulting in a higher confidence of obtaining full-length transcripts. This method, however, suffers from an inability to identify alternative splicing events (Zhao *et al*. 2011; Jain *et al*. 2013).

De novo assemblers create short contigs from overlapping reads, which may be extended through use of insert size estimates. This process involves either the construction of de Bruijn graphs using *k*-mers, or using an overlap-layout-consensus approach (Góngora-Castillo and Buell 2013). The de Bruijn graph-based assembly algorithm is implemented in successive steps. In the first step, short reads are broken into small pieces (*k*-mers) and a de Bruijn graph is constructed from these sections. In the next step, the sequence is derived from the de Bruijn graph. The Overlap graphs algorithm is more computationally

intensive, and mainly effective for assembly of fewer reads with a high degree of overlap (Góngora-Castillo and Buell 2013).

De novo construction of transcripts from short read sequences presents a number of operational challenges, arising from the following observations: some transcripts exhibit low abundance, while others are highly expressed; uneven read coverage may arise across the length of the transcript; transcripts encoded by adjacent gene loci may overlap, and so be fused to form a chimeric structure; sequence repeats from different genes could introduce ambiguity; sequencing errors may occur within reads; and complications may arise due to the presence of alternative splicing forms, allelic variants and paralogous sequences (Góngora-Castillo and Buell 2013; Grabherr *et al*. 2011).

Among other applications, RNA-Seq may be used to detect alternative splicing products, which include the rarest and most specific of transcripts. Moreover, read density can be used to measure the transcript abundance, and thereby the gene expression (Liu *et al.* 2011). RNA-Seq has been applied to both model and non-model organisms in order to understand the most important biological processes. Compared to microarrays, RNA-Seq has many advantages such as cost-efficiency, dynamic range, provision of full-genome coverage and high discriminatory power (Kawahara *et al.* 2012). Availability of genome sequence information, along with transcriptome analysis, will permit enhanced knowledge of the expression of genome at the transcript level, providing information on gene structure, regulation of gene expression, gene product function and genome dynamics (Dong and Chen 2013). Likewise, integration of transcriptomics and proteomics approaches will complement each other by reducing bias and increased data coverage, as well serving as a means of cross-validation to reveal novel biological insights that are not available through a single approach (Zhang *et al.* 2010).

Transcriptome sequencing datasets may be simultaneously analysed to provide insight into levels of gene expression, structures of genomic loci and sequence variation present at such loci (Morozova and Marra 2008). As a consequence, such datasets provide an

important resource for allele mining and marker development (including for new generations of molecular genetic markers, such as those obtained by GBS) (Thomson 2014), based on sequences with known or predictable functions. The use of such data can support assessment of genetic diversity, construction of linkage maps, trait-dissection and implementation of whole-genome selection strategies in varietal improvement programs. In addition, transcriptome datasets may support the identification and selection of candidate genes underlying different physiological and pathological processes, which may be tested for co-location with QTLs and for further fine-scale genetic analysis to verify function.

1.10 Objectives of this study

The general aims of this research were to develop comprehensive genetic linkage maps based on SSR and SNP markers for four RIL-based mapping populations of field pea and to perform trait dissection to identify major genes and QTLs for four key agronomic traits, two relating to abiotic stress (salinity tolerance, boron toxicity tolerance), two to biotic stress (resistances to powdery mildew and bacterial blight). The initial focus was on bacterial blight, as a highly damaging disease, and for this reason, the biology and genetics of the host-pathogen interaction has been described in detail in this chapter. However, development of comprehensive genetic tools allowed analysis of the other traits to be efficiently performed in parallel. Collectively, knowledge of the genomic regions controlling these stress-related characters will address a large proportion of breeding objectives for Australian field pea production. In addition to the trait-dissection studies, a new large-scale transcriptome resource was produced by RNA-Seq for future molecular breeding of field pea.

The specific aims were:

- Development and characterisation of novel SNP markers generated from an existing EST sequence database and characterisation of SSR marker polymorphism in four field pea mapping populations.
- Construction of genetic linkage maps for four trait-specific field pea mapping populations based on existing SSR and newly developed SNP markers.
- Development of a composite genetic map by combining the maps from these trait-specific populations, and then merging the composite genetic map with two previously described consensus maps to form an integrated reference map.
- Performance of comparative genetic analysis between the genome of field pea and those of other legumes of the sub-family Papilionoideae.
- Identification of genomic regions associated with salinity tolerance in field pea based on phenotypic assessment of the Kaspa x Parafield RIL mapping population.
- Identification of genomic regions associated with B toxicity tolerance, based on phenotypic data collected from glasshouse-based nursery screens of the Kaspa x ps1771 RIL mapping population.
- Identification of genomic regions associated with powdery mildew resistance based on phenotypic assessment of the Kaspa x Yarrum and Kaspa x ps1771 RIL populations, and development of a potential diagnostic marker based on resequencing of the *PsMLO1* gene.
- Identification of genomic regions associated with the genetics of bacterial blight resistance by phenotypic assessment in glasshouse conditions of the Kaspa x Parafield and Kaspa x PBA Oura RIL mapping populations.
- Development of comprehensive transcriptome sets from two genotypes of field pea that differ in terms of seed and plant morphological characteristics through use of RNA-Seq, followed by assembly, comparison to gene complements in related species, sequence annotation and assessment of tissue-specific expression.

CHAPTER 2

SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.)

2.1 Chapter preface

This chapter details the development of 768 SNP markers associated with field pea ESTs, and generation of a comprehensive genetic linkage map for the Kaspa x Parafield RIL-based mapping population based on existing SSR markers and newly developed SNP markers. Comparative genomic analysis with other legume species revealed higher levels of conserved synteny with the genomes of *M. truncatula* and chickpea than with soybean, *L. japonicus* and pigeon pea. Trait dissection analysis in the Kaspa x Parafield population revealed a quantitative basis for seedling-induced salinity tolerance in pea, identifying two QTLs.

This chapter is presented in published format.

2.2 Publication details

Title: SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.)

Journal Details: BMC Plant Biology. 2013, 13:161. doi:10.1186/1471-2229-13-161.

Stage of publication: Published

Authors: Antonio Leonforte, Shimna Sudheesh, Noel OI Cogan, Philip A Salisbury, Marc E Nicolas, Michael Materne, John W Forster, Sukhjiwan Kaur

This paper has also been presented at the following conference, and appears in the conference proceedings as:

Leonforte A, Sudheesh S, Cogan NOI, Salisbury PA, Nicolas ME, Materne M, Forster JW, Kaur S (2014) SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). In: *Proceedings of the 6th International Food Legume Research Conference and the 7th International Conference on Legume Genetics and Genomics, Saskatoon, Saskatchewan, Canada*. 7th-11th July 2014, P71.

2.3 Statement of contributions of joint authorship

Antonio Leonforte[+] performed population development, phenotypic assessment, data interpretation and contributed to drafting the manuscript.

Shimna Sudheesh[+] performed DNA extraction, SNP marker discovery, genotyping of the mapping population, map construction, QTL analysis, comparative genomics and contributed to drafting the manuscript.

Philip A Salisbury and Marc E Nicolas co-conceptualised the project and assisted in drafting the manuscript.

Michael Materne co-conceptualised the project.

John W Forster co-conceptualised and coordinated the project and assisted in drafting the manuscript.

Noel OI Cogan and Sukhjiwan Kaur co-conceptualised the project, contributed to data interpretation and assisted in drafting the manuscript.

[+]Equal contributors

2.4 Statement from the co-author confirming the authorship contribution of the PhD candidate

"As co-author of the manuscript 'Leonforte *et al.* (2013) SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). BMC Plant Biology. 13:161', I confirm that Shimna Sudheesh has made the following contributions,

- SNP marker discovery
- DNA extraction of the mapping population
- SSR and SNP genotyping of the mapping population
- Genetic linkage map construction
- QTL analysis
- Comparative genomics
- Data analysis and interpretation
- Writing the manuscript, critical appraisal of content and response to reviewers"

Dr Sukhjiwan Kaur

Date: 15/09/2015

Leonforte et al. BMC Plant Biology 2013, 13:161
http://www.biomedcentral.com/1471-2229/13/161

BMC Plant Biology

RESEARCH ARTICLE Open Access

SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.)

Antonio Leonforte[1,2†], Shimna Sudheesh[3,4†], Noel OI Cogan[3], Philip A Salisbury[2,3], Marc E Nicolas[3], Michael Materne[1], John W Forster[3,4*] and Sukhjiwan Kaur[3]

Abstract

Background: Field pea (*Pisum sativum* L.) is a self-pollinating, diploid, cool-season food legume. Crop production is constrained by multiple biotic and abiotic stress factors, including salinity, that cause reduced growth and yield. Recent advances in genomics have permitted the development of low-cost high-throughput genotyping systems, allowing the construction of saturated genetic linkage maps for identification of quantitative trait loci (QTLs) associated with traits of interest. Genetic markers in close linkage with the relevant genomic regions may then be implemented in varietal improvement programs.

Results: In this study, single nucleotide polymorphism (SNP) markers associated with expressed sequence tags (ESTs) were developed and used to generate comprehensive linkage maps for field pea. From a set of 36,188 variant nucleotide positions detected through *in silico* analysis, 768 were selected for genotyping of a recombinant inbred line (RIL) population. A total of 705 SNPs (91.7%) successfully detected segregating polymorphisms. In addition to SNPs, genomic and EST-derived simple sequence repeats (SSRs) were assigned to the genetic map in order to obtain an evenly distributed genome-wide coverage. Sequences associated with the mapped molecular markers were used for comparative genomic analysis with other legume species. Higher levels of conserved synteny were observed with the genomes of *Medicago truncatula* Gaertn. and chickpea (*Cicer arietinum* L.) than with soybean (*Glycine max* [L.] Merr.), *Lotus japonicus* L. and pigeon pea (*Cajanus cajan* [L.] Millsp.). Parents and RIL progeny were screened at the seedling growth stage for responses to salinity stress, imposed by addition of NaCl in the watering solution at a concentration of 18 dS $m^{-1.}$ Salinity-induced symptoms showed normal distribution, and the severity of the symptoms increased over time. QTLs for salinity tolerance were identified on linkage groups Ps III and VII, with flanking SNP markers suitable for selection of resistant cultivars. Comparison of sequences underpinning these SNP markers to the *M. truncatula* genome defined genomic regions containing candidate genes associated with saline stress tolerance.

Conclusion: The SNP assays and associated genetic linkage maps developed in this study permitted identification of salinity tolerance QTLs and candidate genes. This constitutes an important set of tools for marker-assisted selection (MAS) programs aimed at performance enhancement of field pea cultivars.

Keywords: Grain legume, Genetic marker, Trait dissection, Comparative genomics, Abiotic stress, Breeding

* Correspondence: john.forster@depi.vic.gov.au
†Equal contributors
[3]Department of Environment and Primary Industries, Biosciences Research Division, AgriBio, Centre for AgriBioscience, 5 Ring Road, La Trobe University Research and Development Park, Bundoora, VIC 3083, Australia
[4]La Trobe University, Bundoora, VIC 3086, Australia
Full list of author information is available at the end of the article

Background

Field pea (*Pisum sativum* L.) is widely cultivated on a global basis as an important legume crop for human dietary protein intake and livestock forage nutrition [1]. Field pea is especially beneficial in crop rotations with cereals in order to provide disease breaks and for provision of soil nitrogen [2].

Development of sustainable high-yielding varieties that persist under biotic and abiotic stresses is a prerequisite for meeting the food requirements of a growing world population. Molecular breeding strategies have been adopted for crop improvement programs in several crops, including legumes such as soybean and common bean [3], and are suitable for application in field pea. Most breeding gains for grain yield in field pea have been achieved by optimisation of crop architecture (i.e. reduced internode length), harvest index and phenology traits with growing season length and rainfall [4,5]. Breeding practices have also primarily focused on pyramiding genes for resistance to important fungal diseases such as ascochyta blight, powdery and downy mildew, and viruses such as pea seed-borne mosaic virus (PSbMV) and bean leaf roll virus (BLRV). However, comparatively little effort has been directed towards improvement of physiologically complex and putatively multigenic traits such as tolerance to salinity stress [6].

Genetic improvement for complex traits will be facilitated by new genomic tools through the identification and selection of preferred genes. For legume crops, only limited genomic resources were available until recently, so MAS adoption has been slow [6]. However, advances in DNA sequencing and genotyping technologies have recently delivered large-scale transcriptome sequence data sets for field pea [7,8]. These data can be exploited for the design of DNA-based genetic markers such as SSRs and SNPs, supporting linkage mapping, analysis of genetic diversity, trait-dissection [9,10], as well as gene-tagging for MAS [11].

For pea, a large number of genetic linkage maps have been developed previously [10,12-18]. SSR markers are generally co-dominant in nature and highly polymorphic, and have been extensively used for pedigree analysis in crop breeding and genetics research [6]. SNPs are highly prevalent, usually biallelic and co-dominant in nature, sequence-tagged, and amenable to development of low-cost multiplexed marker assays that can provide sufficiently dense genome coverage for the dissection of complex traits [19,20]. A number of methods have been developed for SNP detection. Medium to high-throughput array-based SNP genotyping systems are now available, depending on the number of samples and markers to be analysed, such as GoldenGate® and Infinium from Illumina Inc., SNPStream from Beckman Coulter, and GeneChip from Affymetrix [18].

In order to understand complex biological processes in plants, comparative genetic analysis with model species has been used extensively. In concert with extensive genomic resources that are available for a number of species of the legume sub-family Papilionoideae (e.g. *M. truncatula* [http://www.medicago.org], *L. japonicus* [21], chickpea [22], soybean [23] and pigeon pea [24]), such analysis provides opportunities for translational genomics to assist breeding of other, less well-studied crop legumes, such as field pea.

Soil salinity is a serious global problem due to limitation of plant growth and reduced crop yield [25]. Salinity tolerance in field pea has become increasingly important in Australia due to a geographical shift of crop production towards environments characterised by shorter seasons, greater water limitation and marginal soils with higher transient soil salinity [26]. Large effects of salinity and sodicity are predominantly due to levels of the Na^+ cation, and in Australia, are commonly associated with highly alkaline (pH > 8.5) soils [27,28]. In combination, these factors can cause nutrient (Fe, K) deficiencies and soil toxicities (such as to elevated levels of boron) that limit growth and grain yield potential. For field pea, relatively high and heritable genetic tolerances to Fe deficiency [29] and boron toxicity [30-32] have been identified. In terms of salinity tolerance, preliminary studies based on biomass reduction indicated that field pea is significantly more sensitive than other commonly cultivated Australian broad-acre crops such as barley [33,34], wheat [35] and canola [36], due to a low salinity threshold level [37] in pea. In comparison to other legumes, in contrast, pea [38-41], as well as faba bean [42], appear more tolerant than chickpea [43] and lentil [44].

Research on other major dry-land crops such as wheat [45] has demonstrated the difficulty of using yield-based response measurements from field studies as a measure of salinity tolerance, due to the complexity of interactions with other stress factors such as high pH and boron, Na^+ variability in the soil profile, and differential responses according to both growth stage and genotype. However, low-cost and reliable pot-based glasshouse screening methodologies have been developed for a range of crops, including pea [41], which can be used to identify useful variation at the seedling stage for breeding purposes. Considerable potential for genetic improvement appears to be available, on the basis of the outcome of screening experiments [41,46]. Identification and marker-tagging of genomic regions containing QTLs for aspects of salinity stress tolerance would hence highly facilitate the targeted introgression of this trait into otherwise unadapted germplasm.

The objectives of the present study were: development and characterisation of novel SNP markers and characterisation of existing SSR markers; construction of an

SSR- and SNP-based linkage map for a field pea population varying for salinity tolerance; comparative genetic analysis between field pea and other legumes of the subfamily Papilionoideae; and identification of genomic regions and molecular genetic markers associated with salinity tolerance in field pea.

Methods

Plant material and DNA extraction

Crosses were made between single genotypes of cultivar Kaspa (salinity sensitive), and Parafield (moderately tolerant). The crosses were performed at DEPI-Horsham in 2007 and F_2 generation progeny were produced. Single seed descent was undertaken from F_2 progeny-derived genotypes for 4 generations in the glasshouse from 2008 to 2011. The resulting F_6 mapping population consisted of 134 RILs.

Frozen leaf tissue from each progeny genotype was ground using a Mixer Mill 300 (Retsch®, Haan, Germany), and genomic DNA was extracted using the DNeasy® 96 Plant Kit (QIAGEN, Hilden, Germany). DNA was resuspended in 1 x TE buffer to a concentration of 50 ng/μl and stored at −20°C.

SNP discovery and validation

Putative SNPs were identified from transcriptome sequence data [8] using Next*GENe* software v1.96 (SoftGenetics, State College, PA, USA). Based on alignment of high-quality sequences from four genotypes (including Kaspa and Parafield [8]) with the consensus reference (obtained as a result of *de novo* assembly), all base variants were identified. All insertion and deletion (indel) variants were excluded from further analysis. Subsequently, high-confidence SNPs were filtered using the following criteria: (1) base variants in homozygous condition within each genotype; (2) read-coverage equal to or greater than 4; and (3) absence of any other base variants within 20 bp segments flanking each SNP.

A sub-set of 48 SNPs was selected for experimental validation by Sanger sequencing. Primer pairs were designed using Sequencher 4.7 (Gene Codes Corporation, USA) and OligoCalc: Oligonucleotide Properties Calculator (http://www.basic.northwestern.edu/biotools/oligocalc.html). PCR reactions contained 10 ng of genomic DNA in a 12 μl reaction with 5 μM of each primer pair. The amplification conditions were as follows: a hot start at 94°C for 15 min, followed by 35 cycles of 95°C for 30 s, 50°C for 30 s and 72°C for 1 min, and a final elongation step at 72°C for 7 min. PCR products were purified in a 15 μl reaction containing 0.5 U exonuclease I (New England Biolabs), 0.5 U shrimp alkaline phosphatase (USB-VWR International, Pennsylvania, USA) and 5 μl of PCR product. Sequencing reactions were performed in a total volume of 7.5 μl, each reaction contained 3.2 μM primer, BigDye® Terminator v3.1 (Life Technologies Australia Pty Ltd, Victoria, Australia), BigDye® sequencing buffer (Life Technologies Australia Pty Ltd, Victoria, Australia) and were subjected to cycling conditions as described in the BigDye® v.3.1 protocol. Extension products were purified by the ethanol/EDTA/sodium acetate precipitation method, resuspended in 12 μl Hi-Di™ formamide (Life Technologies Australia Pty Ltd, Victoria, Australia), and separated on the ABI3730xl automated capillary electrophoresis platform. DNA sequence analysis and alignment was performed using Sequencher 4.7, while contig assembly and the SNP validation was performed visually.

SSR genotyping

Genomic DNA- and EST-derived SSRs [8,17] were screened on the mapping parents for polymorphism detection. Primer synthesis and PCR amplifications were performed as described previously [8,47]. PCR products were combined with the ABI GeneScan LIZ500 size standard and analysed using an ABI3730xl (Life Technologies Australia Pty Ltd, Victoria, Australia) capillary electrophoresis platform according to the manufacturer's instructions. Allele sizes were scored using GeneMapper® 3.7 software package (Life Technologies Australia Pty Ltd).

Framework genetic map construction and selection of maximally recombinant individuals

A framework genetic map was constructed using Joinmap® 3.0 [48] with a threshold log-of-odds (LOD) score of 3 using SSR-derived genotyping data, providing the basis for selection of maximally recombinant individuals in the mapping population using MapPop version 1.0 [49].

SNP genotyping

A preliminary list of SNPs was selected for GoldenGate® primer design (Illumina Inc., San Diego, CA, USA). A designability rank score (0 to 1) was calculated for each SNP by Illumina. Finally, SNPs with designability scores between 0.7 and 1.0 were selected for development of an Illumina GoldenGate® oligonucleotide pool assay (OPA) for genotyping. Individuals were SNP genotyped according to the manufacturer's instructions using 250 ng of template genomic DNA. The genotyping assays were processed by the Illumina iScan reader. Automatic allele calling was achieved using the Illumina Genome Studio software v2011.1 with a GeneCall threshold of 0.20 and checking the output visually as well for the confirmation of cluster specificity.

Genetic linkage mapping

The genetic linkage map was generated using Map Manager Software version QTXb19 [50]. Markers with a χ^2 score >10 were not included in further analysis. Map

distances were calculated using the Kosambi mapping function [51] at a threshold LOD score of 3. LGs were assigned on the basis of marker loci [17] in common with publicly available linkage maps of pea, and by comparison with chromosomes of *M. truncatula* [52,53]. LGs were drawn using Mapchart software v 2.2 [54].

Comparative genome analysis

DNA sequences underlying map-assigned SSR and SNP markers were used to perform comparative analysis with genome assemblies of chickpea (NCBI, Project PRJNA175619), *M. truncatula,* v3.5 (http://www.medicago.org), *G. max* v189 (http://www.plantgdb.org), *L. japonicus,* v2.5 (ftp://ftp.kazusa.or.jp/pub/lotus/lotus_r2.5/pseudomolecule/) and *C. cajan* v5.0 (http://www.icrisat.org/gt-bt/iipg/Genome_Manuscript.html). BLASTN was used to conduct similarity searches against each genome sequence with a threshold E-value of 10^{-10}.

Phenotypic screening

The Kaspa x Parafield RIL population was screened for response to NaCl-induced stress applied at the seedling stage. Experiments were conducted during the autumn of 2012 in a semi-controlled (polyhouse) environment at DEPI-Horsham. Screening was undertaken by sowing six plants of each RIL at equidistant spacing in 13 cm diameter pots into a sand and gravel medium (to a depth of 2 cm in two pot replications). This provided 12 plants as replicates for each RIL. The medium was composed from a 1:1 ratio of coarse river sand and 5 mm bluestone chips. Each pot was treated daily with rainwater from sowing until emergence. From 6 days post-emergence, seedlings were watered with a complete nutrient solution (i.e. nitrosol, NPK ratio 12.2: 2.9: 8.5), in addition to supplementation with a calcium source (i.e. calcium nitrate). The required NaCl concentration was tested using an electrical conductance (EC) meter and was applied at an initial rate of 3 dS m^{-1} from day 9 post-emergence. The concentration of applied NaCl was increased by 3 dS m^{-1} at each watering time to avoid abrupt osmotic shock, up to a final rate of 18 dS m^{-1}, and maintained at this concentration until assessment. All watering with the nutrient and salt solution was undertaken over 3 day-intervals at a rate of 200 ml per pot applied directly to the growing medium surface. A null-salt application treatment (no added NaCl) was included for control lines (parental genotypes) and randomised in the experiment in order to eliminate effects due to other stress factors. Individual plants in each pot were assessed for symptom development (symptom score) as described previously [41] from 28 days post-emergence and thereafter on every 7th day until plant death. Final plant biomass cuts were also obtained and seed set was recorded per genotype pot. Averages for plant symptom score were calculated from individual plant assessments and used to estimate genotype-specific average values for symptom score using REML spatial row-column analysis. An index was used to quantify genotypic salinity tolerance values, and to describe tolerance levels according to sensitivity based on weighted symptom scores and final biomass.

Averages for plant symptom score (calculated from individual plant assessments) and salt index were used to generate frequency distribution histograms. Narrow sense heritabilities (h^2) were calculated for the trait by considering the spatial trends in the experiment using best linear unbiased prediction (BLUP) analysis.

QTL analysis and candidate gene selection

QTL detection was conducted using MapManager QTX software version QTXb19. Marker regression analysis was initially performed to identify markers significantly associated with trait variation (LOD threshold = 3). Simple interval mapping (SIM) and composite interval mapping (CIM) methods were used to identify and confirm QTLs associated with salt tolerance. The sequences underpinning SNP loci flanking the QTL-containing intervals were BLAST analysed against the *M. truncatula* genome to identify genomic regions containing putative candidate genes.

Results

SNP discovery and validation

A total of 36,188 putative SNPs were identified from comparison of transcriptome reads obtained from the mapping parents against the EST sequence database. An average frequency of 1.85 SNPs per kb between two haplotypes was observed. A preliminary set of 21,000 SNPs were selected following elimination of indels. After further filtration based on the criteria of homozygous status and absence of other known SNPs in the vicinity, a subset of 956 high quality SNPs was obtained. Of these, a total of 953 satisfied the required primer design criteria and a final sub-set of 768 SNP loci with a designability rank of 1 was selected for GoldenGate® assay.

Analysis of nucleotide variation revealed that transition substitutions were more predominant (2:1) than transversions. The two most common SNP variants were A/G and C/T, representing 36% and 32% of all changes, respectively. The other SNP variants (T/G, C/G, A/C and A/T) accounted for less than 10% of the total (Additional file 1). A subset of 48 SNP loci was verified through Sanger sequencing prior to 768-plex SNP OPA synthesis (Additional file 2), of which 45 were concordant with prediction (Additional file 3).

Framework genetic map construction and selection of maximally recombinant individuals

A total of 96 of 240 genomic DNA-derived SSRs and EST-SSRs (40%) revealed polymorphism between

the parental genotypes, of which 78 were selected for screening on the mapping population on the basis of consistency of amplification. A sub-set of 47 SSR markers generated data of sufficient quality to generate a framework genetic map, and 40 loci (85%) were assigned to 9 LGs. These data were then used to select 101 maximally recombinant individuals for use in bin mapping.

SNP genotyping

A total of 768 SNPs were used to genotype the 101 selected RILs. All SNPs were visually qualified, the majority producing two major clusters in Genome Studio representing the homozygous (AA and BB) genotypic classes, but occasionally a third small cluster of heterozygous (AB) genotypes was also observed (Additional file 4). The mapping population was descended to the F_6 level, so residual heterozygosity was expected to be low (c. 5 - 10%). A total of 705 SNPs (91.7%) produced coherent data, while those generating ambiguous cluster structures were removed from further analysis. A sub-set of 462 SNPs (65%) generated polymorphic clusters within the Kaspa x Parafield mapping population and were used for genetic linkage map construction.

Linkage mapping

A total of 73 markers (13.5%) were excluded from linkage analysis due to excessive heterozygosity, missing data, skewed segregation or ambiguity. A final set of 467 markers (53 SSRs and 414 SNPs) was used for linkage map construction. A small proportion of markers were ungrouped, such that 458 (98%), comprising 48 SSRs and 410 SNPs (Table 1) were assigned to 9 LGs (Additional file 5). The estimated cumulative total map length was 1916 cM with an average inter-locus interval of 4.2 cM (Figure 1; Table 2). LG identity and orientation were determined by comparison with the *M. truncatula* genome, as well as from the use of previously map-assigned SSRs as anchoring markers.

Comparative genome analysis

Corresponding DNA sequences were available for 310 of 458 of the mapped loci (15 EST-SSRs and 295 SNPs), of which 307 detected significant sequence similarity matches to at least one of the reference genome sequences, and 130 sequences displayed similarity to sequences in all five genomes.

Comparison of the field pea map with the chickpea genome revealed the highest number of matches (301: 97%) (Additional file 6). The syntenic relationships related each of field pea chromosomes Ps II, Ps IV, Ps V , V.2, and Ps VII to chickpea pseudomolecules Ca4, Ca7, Ca3 and Ca6, respectively. Some LGs containing blocks syntenic to more than one Ca group were also observed. Field pea - *M. truncatula* macrosynteny was observed for 292 (94%) sequences. Among *M. truncatula* chromosomes, Mt5, 1, 3, and 7 exhibited synteny and colinearity with pea linkage groups Ps I, Ps II, Ps III and Ps V respectively (Figures 2 and 3). Conversely, Mt2 and 6 contained the lowest number of field pea orthologues, revealing more complex relationships with PsLGs.

Despite a large number of matches (294) between field pea and soybean sequences, significant chromosomal rearrangements were observed between the two genomes, such that each PsLG exhibited substantial synteny with more than one soybean chromosome. Comparison with *L. japonicus* identified 226 (73%) matches with segmental syntenic blocks rather than whole chromosomal relationships. Field pea – pigeon pea synteny analysis revealed the lowest number of matches (183), short conserved regions being distributed across different chromosomes. In most instances, CcLGs were inverted in order in comparison to PsLGs, apart from CcLG 2 and 11.

The 130 common orthologous sequences were used to further analyse and confirm the degree of genome conservation (Figures 4 and 5). For most PsLGs, only one or two corresponding chromosomes were identified for chickpea and *M. truncatula*, but complex relationships were observed with *L. japonicus,* pigeon pea and soybean, consistent with the pair-wise comparisons. The exception to these general patterns was Ps VI, which displayed complex relationships in all instances.

Phenotypic analysis, QTL detection and candidate gene selection

Plant symptom response data from salinity screening of the RIL population at the seedling stage indicated that variation for tolerance was normally distributed (Additional files 7 and 8), and therefore likely to be controlled by multiple genes. The estimated narrow sense heritability (h^2) for salt index was 0.55. Two different phenotypic measurements, including salt index and mean symptom score (average of symptom scores obtained at up to 35 days) were used to detect salt tolerance QTLs (Figure 1), with LOD scores of 3.2 (salt index) and 2.5 (symptom score) as minimum significance levels. Two QTLs were identified on Ps III and Ps VII, explaining 12% and 19% of phenotypic variance (V_p) for salt index score, and 12% and 17% for the

Table 1 Total number of markers analysed, tested for polymorphism and assigned to genetic linkage map locations

Marker type	Total number of markers	Polymorphic markers	Mapped markers
Genomic DNA-derived SSR	144	54	30
EST-SSR	96	24	18
SNP	768	462	410
Total markers	1008	540	458

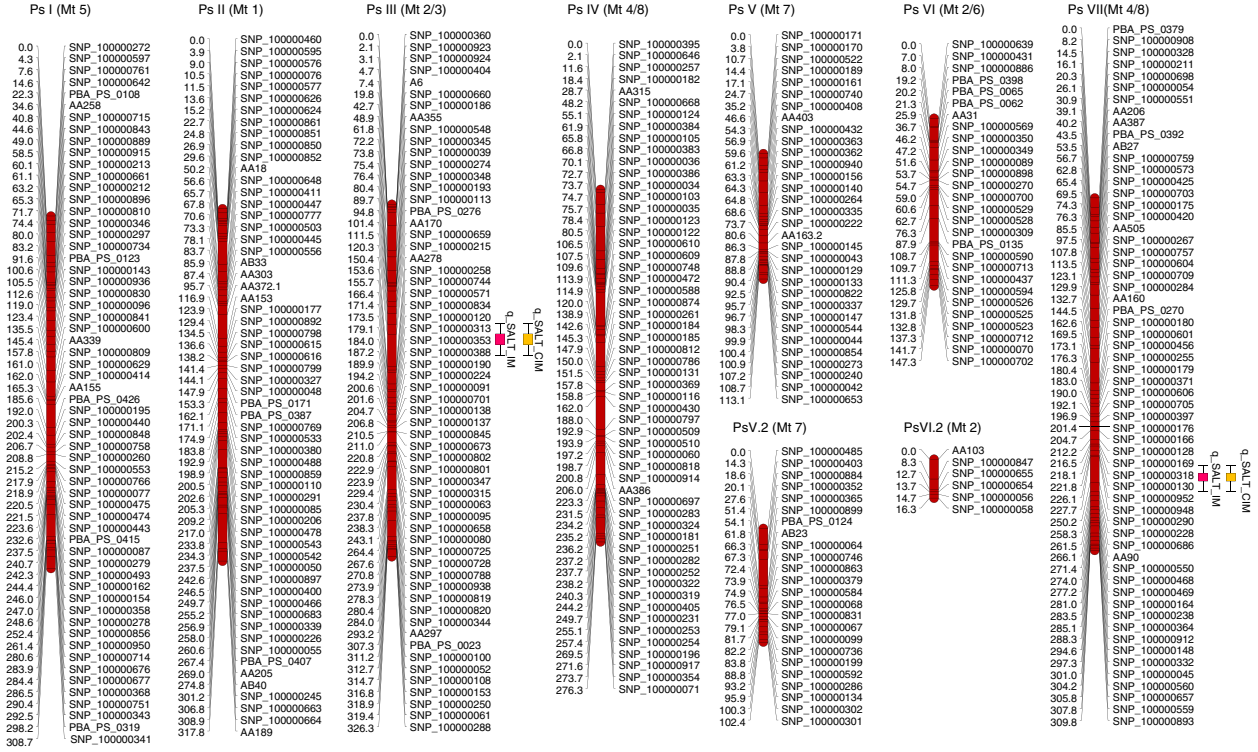

Figure 1 Genetic linkage map of the Kaspa x Parafield field pea cross, showing the location of two QTLs for salinity tolerance. The markers are shown on the right of the linkage groups and map distances between markers are indicated in cM on the left. For presentation purposes, only one of a set of co-located genetic markers are shown on the map.

symptom score, respectively (Table 3). QTL analysis was also performed using symptom scores obtained at different time points (day 7, 14, 21, 35), which identified the same QTL locations and accounted for similar proportions of V_p (data not shown). The phenotypic data for symptom scores obtained at day 42, 49, 56 deviated from normality, and was consequently not used for QTL analysis based on mean symptom score.

Table 2 Marker distribution over the LGs of the Kaspa x Parafield map

LGs	Predicted pea chromosome	Length (cM)	Number of mapped markers	Average marker density (cM)
LG 1	Ps VII	309	87	3.6
LG 2	Ps III	326	78	4.2
LG 3	Ps I	309	69	4.5
LG 4.1	Ps V	113	35	3.2
LG 4.2	Ps V	102	25	4.1
LG 5.1	Ps VI	147	29	5.1
LG 5.2	Ps VI	16	6	2.7
LG 6	Ps IV	276	63	4.4
LG 7	Ps II	318	66	4.8
Total		1916	458	4.2

Comparison of linked marker-associated sequences to the *M. truncatula* genome directly identified candidate genes with functional annotations as receptor-like protein kinase, 14-3-3-like protein, histone deacetylase and glutamine synthetase, which have been reported as being involved in the complex salt tolerance mechanisms of plants (Figure 6). In addition, regions of the *M. truncatula* genome immediately adjacent to and within the intervals between orthologues of the linked SNP-associated sequences were examined for candidate gene presence. The Medtr3g073300.1 gene was located in the interval between field pea SNP markers SNP_100000313 and SNP_100000353, in the vicinity of Ps III-QTL1, and was annotated as a salt tolerance protein.

Discussion

SNP variation in field pea

SNP frequencies in plant genomes vary significantly, depending on reproductive habit (autogamous or allogamous), diversity of populations under assessment and status (coding or non-coding) of analysed regions. The SNP frequency detected in field pea in the present study is much lower than values reported for cereal crops (16.5 SNPs per kb in wheat, 4.2 SNPs per kb in rice [55]), but similar to those for other legumes (0.9 SNPs

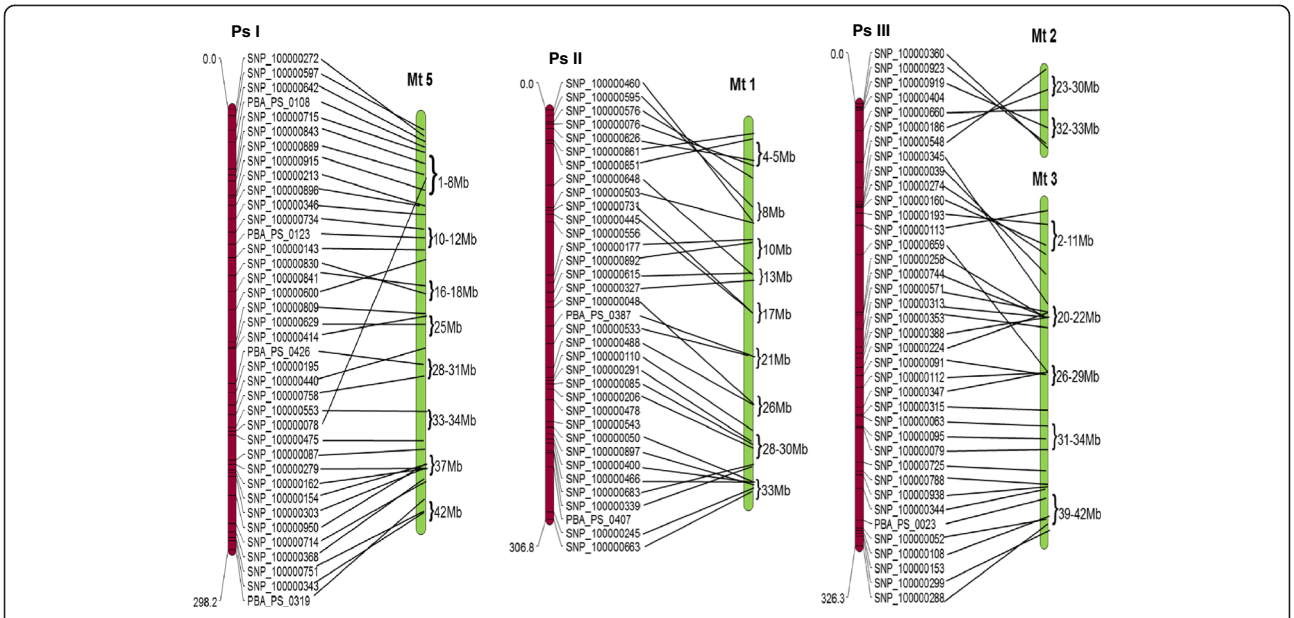

Figure 2 Schematic representation of syntenic relationships between field pea (LGs PsI - III) and the *M. truncatula* genome. LGs or chromosomes are shaded in different colours for presentation purposes. The red-shaded LGs are from field pea, and the green chromosomes are from *M. truncatula*. The lines represent the corresponding positions of orthologous sequences.

per kb in cowpea [56], 1.96 SNPs per kb in *M. truncatula* [57] and 2.06 SNPs per kb in soybean [58]).

The patterns of nucleotide substitution showed A/G and C/T to be the most common base changes, in agreement with previous studies of legume species such as white clover [59] and chickpea [60]. The high proportion of C/T transitions are likely to be partially due to deamination of 5-methylcytosine reactions, which occurs frequently over evolutionary time, particularly at CpG dinucleotides [61].

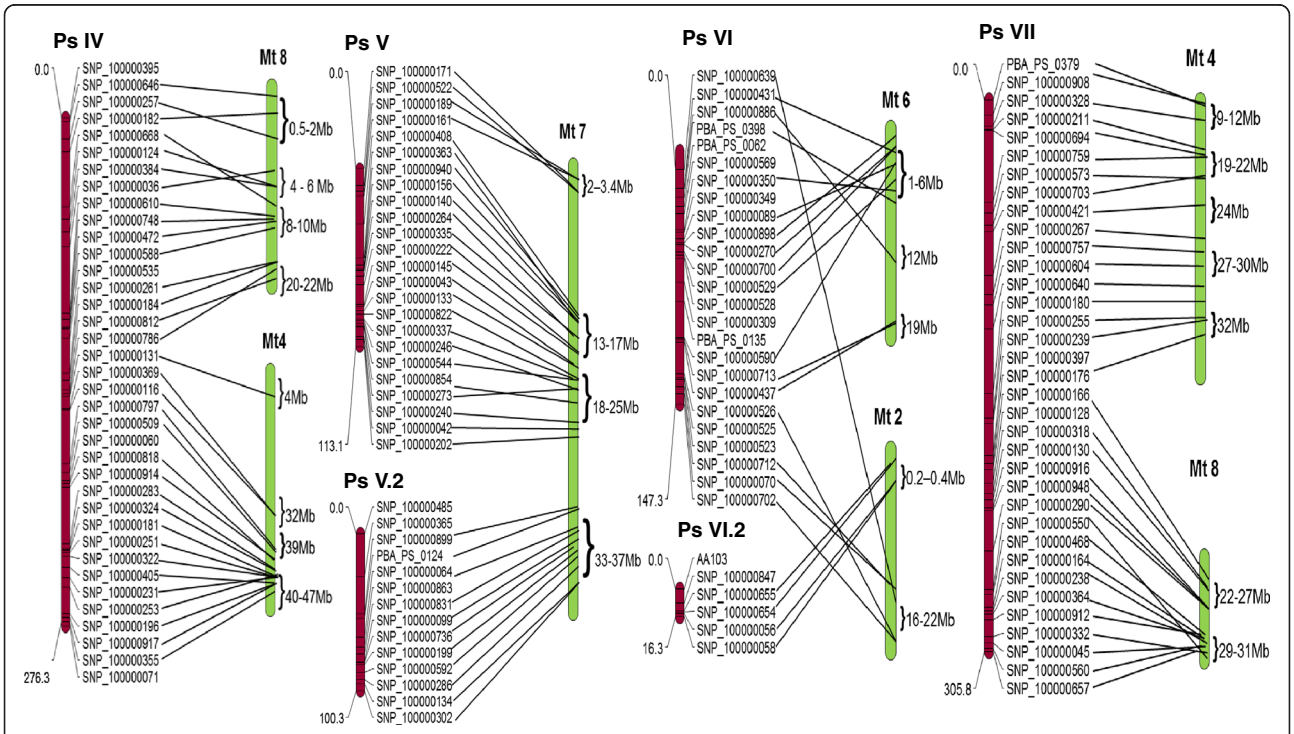

Figure 3 Schematic representation of syntenic relationships between field pea (LGs PsIV - VII) and the *M. truncatula* genome. Details are as for Figure 2.

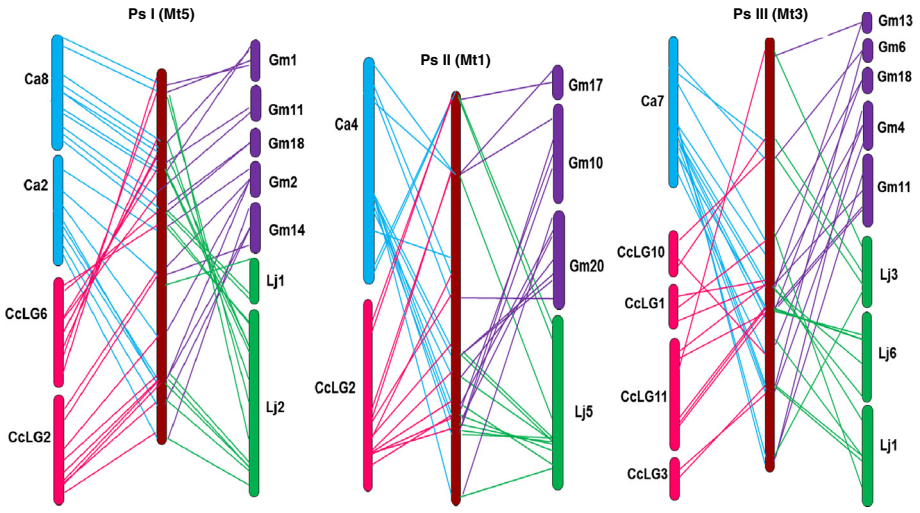

Figure 4 Syntenic relationships of field pea (LGs PsI -III) with other legume genomes. LGs or chromosomes are shaded in different colours for visualisation purposes. The details of colour codes are as follows, blue -chickpea, pink – pigeon pea, violet - soybean, green – *L. japonicus* and brown – *M. truncatula*. Coloured lines represent the corresponding positions of the orthologous sequences in field pea.

The effectiveness and suitability of GoldenGate® SNP assays for genotyping mapping populations and genetic resource collections of pea has been previously demonstrated [18]. The present study provides additional SNP markers that can be utilised for molecular breeding programs. The success rate for SNP genotyping (c. 91%) was comparable to previous observations made in pea (92.7%) [18] and chickpea (90.75%) [60]. Success of SNP genotyping depends on many factors including base variant selection, adjacent SNP frequency, presence of repetitive sequences, and finally, designability score. As field pea SNP discovery was based on transcriptome sequencing from multiple genotypes [8], it is not surprising that a substantial minority of markers (c. 35%) failed to detect polymorphism in the mapping family. However, inclusion of Kaspa and Parafield among the selected genotypes ensured a high frequency of success.

Genetic linkage mapping

Several field pea linkage maps have been previously developed with successive adoption of new molecular marker technologies [10,12-17]. The linkage map

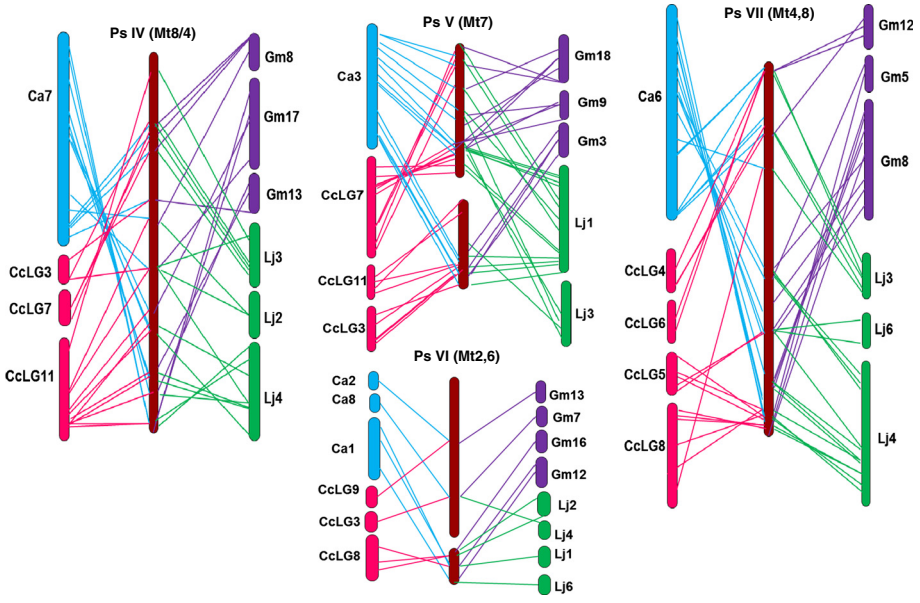

Figure 5 Syntenic relationship of field pea (LGs PsIV - VII) with other legume genomes. Details are as for Figure 4.

Table 3 Identification of QTLs for salt tolerance on the Kaspa x Parafield genetic map based on CIM

Trait	Flanking markers	Linkage group	Position (cM)	LOD threshold	Max LOD score	Phenotypic variance (%)
Salt index_QTL 1	SNP_100000313	Ps III	179 - 184	3.2	3.9	12
	SNP_100000353					
Salt index_QTL 2	SNP_100000318	Ps VII	218 - 222	3.2	4.7	19
	SNP_100000130					
Symptom score_QTL 1	SNP_100000313	Ps III	179 - 184	2.5	3.9	12
	SNP_100000353					
Symptom score_QTL 2	SNP_100000318	Ps VII	218 - 222	2.5	5.9	17
	SNP_100000130					

constructed in the present study exhibits a regular marker distribution, but a significantly longer cumulative genetic map (1916 cM) than would be expected on the basis of typical chiasma frequency (1–2 per bivalent) at meiotic prophase. Such expansions of the pea genetic linkage map were also previously reported (1700 cM [62]; 2202.7 cM [63]). Several factors may be responsible, including the genetic constitution of different mapping populations, mapping strategies, number and type of mapped loci, the choice of mapping software and ratio between number of markers and population size [64-67].

Comparative genome analysis

Extensive conservation of genome structure between field pea and both chickpea and *M. truncatula* was consistent with the closer phylogenetic relationship between these species than for the other legumes used in this study. In contrast to results of previous comparative genetic studies between chickpea and other legumes [68,69], substantial macrosynteny was observed in the present study.

Broad conservation of chromosome structure was observed between the 8 chromosomes of *M. truncatula* and 7 LGs of field pea, as well as evidence for evolutionary translocations [52,70]. A number of previous studies [52,53] have described high levels of conservation associated with comparisons to Mt1 and 5, moderate conservation of Mt3, 4, 7 and 8, and low levels of conservation for Mt2 and 6. Unlike other Mt chromosomes, Mt6 is short in length with a large number of repeats, low gene content (but a significant number of NBS-LRR disease resistance genes) and high heterochromatin content [71]. Ps VI, which matches Mt2 and 6, contained the least number of orthologous sequence queries, consistent with these prior studies. The situation may potentially be remedied by development of a larger cohort of markers from Ps VI. Despite a c. 10-fold difference in the genome size between *M. truncatula* and field pea

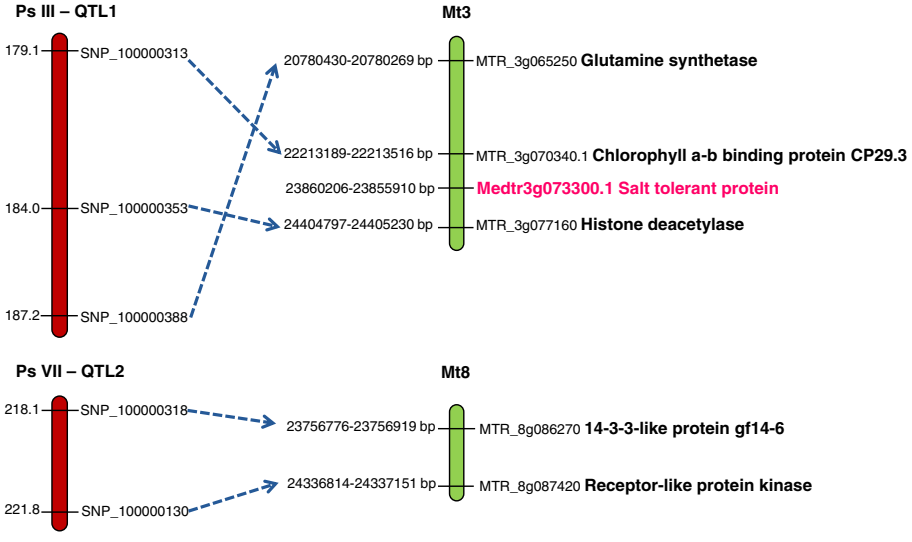

Figure 6 Syntenic relationships between salt tolerance QTL-containing regions of the field pea genetic map and the *M. truncatula* genome, indicating candidate gene locations. LGs or chromosomes are shaded in different colours for presentation purposes. The red-shaded LGs are from field pea, and the green chromosomes are from *M. truncatula*.

[72], the extensive synteny between the two genomes suggests that whole genome duplication has not occurred in the pea lineage subsequent to evolutionary divergence from c. 40 MYA [53,73]. The larger genome size of pea could be the consequence of multiple transposition events [74]. The results of the present study have substantially extended comparative knowledge of the field pea and *M. truncatula* genomes, and such information may be used for candidate gene selection for further application to breeding programs.

In contrast, large syntenic blocks spanning entire PsLGs were absent from the comparisons with the *L. japonicus,* soybean and pigeon pea genomes. The former is a member of the Galegoid clade of the Papilionoideae sub-family, but more distantly related to pea than *M. truncatula* and chickpea, while the latter two are members of the Phaseoloid clade, so the observed relationships are in accord with broad phylogenetic affinities [75]. For soybean, the more limited relationships arose despite a large number of orthologous sequences, potentially also reflecting the complex paleopolyploid genome architecture of this species [76]. The field pea – *L. japonicus* comparison revealed similarities, but was frequently interrupted by chromosomal rearrangements. Similar segmental syntenic relationships were observed between *L. japonicus* and the Galegoid forage legume white clover [77], as also inferred from comparison to *M. truncatula* [78]. Extensive chromosomal rearrangements were evident between field pea and pigeon pea, again indicating the effects of taxonomic divergence.

Phenotypic analysis, QTL detection and candidate gene selection

Plant response to salt tolerance is influenced by various physiological mechanisms, which are likely to be controlled by multiple genes across different environments [79]. The present study suggests a quantitative basis for seedling-induced salinity tolerance derived from adapted and high-yielding parental field pea genotypes, and a medium level of heritability, c. 45% of the variation being due to non-genetic factors. Two QTL loci were identified on Ps III and Ps VII, each accounting for moderate proportions of V_p. Studies of different physiological traits associated with salt tolerance in *M. truncatula* identified a total of 19 putative genomic regions, the largest number of QTLs being located on Mt8 followed by Mt5, 1, 3, 4, 7, 6, and 2 [80]. A direct comparative QTL analysis could not, however, be performed due to inaccessibility of *M. truncatula* sequences associated with markers flanking the QTL intervals. However, the comparative genome analysis revealed macrosyntenic relationships between Ps III and Mt2/3, and Ps VII and Mt4/8. It is hence possible that the QTLs identified in the present study may be conserved between the Galegoid legumes.

The present study identified candidate genes associated with salt tolerance mechanisms in field pea. Histone deacetylase and glutamine synthetase have a key role in salt stress resistance in plants [81,82], while 14-3-3 proteins regulate the activities of a wide array of targets and play an important role in responses to saline stress [83]. Receptor-like protein kinases are involved in a diverse range of processes including biotic/abiotic stress response [84]. Furthermore, the salt tolerance protein (STO) was identified as one of the gene products involved in the regulation of the internal Na^+/K^+ ratio, an essential process for salinity tolerance [85]. The genes identified within the QTL-containing regions are therefore plausible candidates, although additional studies will be required for validation.

The QTLs identified in the present study are associated with seedling growth-stage salinity tolerance. Similarly, QTLs for seedling growth tolerance have been identified in numerous grain crops, including rice [86], barley [87], soybean [88] and wheat [89]. Mechanisms related to other QTLs for growth-response occurring at germination (in tomato [90,91], rice [92], barley [93] and wheat [89]) or during reproductive development (rice [94], barley [95] and tomato [96]) are likely to be significant for field pea and warrant further investigation. The substantial variation in degree and timing of salinity-induced growth responses within and between crop species highlights complexity of the trait.

Implementation of molecular markers in MAS has rarely been achieved for physiologically complex traits such as salinity tolerance [97]. In such circumstances, breeders will need to select for varying and multiple genomic regions or response mechanisms found in different germplasm, different screening environments and within different ontogenic stages. It may therefore be necessary to quantify the adaptive nature [98] of different QTLs according to varying salinity stress, and to allocate genomic values akin to index-trait based selection. Advances in genome sequencing and genotyping capacity, especially genotyping-by-sequencing (GBS), offer the potential for genome-wide marker analysis [99] and the capacity to identify all loci contributing to a trait such as saline stress tolerance, irrespective of effect magnitude. Such data may be used to develop breeding value estimates based on all trait-linked markers, in order to identify key parental lines for targeted introgression programs.

Conclusion

The present study describes the development of a multiplexed set of EST-derived SNPs for genetic linkage map construction in field pea. Evaluation of salt

tolerance under glasshouse conditions permitted identification of two significant genomic regions. Through use of sequence-associated markers, macrosyntenic relationships were determined between field pea and five other legumes and used to predict candidate genes for salt tolerance. This information may be used for the development of linked and diagnostic polymorphisms for marker-assisted selection (MAS) of salt tolerant cultivars, based on introgression of QTL-containing genomic regions from donor to recipient germplasm. As salinity tolerance is a physiologically complex trait, future research will require evaluation in different screening environments and across varying ontogenic stages to identify additional associated genomic regions. Finally, the genetic resources generated in this study will assist other trait-dissection studies and facilitate transfer of information from related legume crops for future enhanced breeding of field pea.

Additional files

Additional file 1: Percentage of SNP base variants. This file contains a pie-chart depicting the percentages of each SNP base variant class.

Additional file 2: Details of the 768plex SNP-OPA design. This file contains names and sequence information for all SNP markers used for linkage mapping.

Additional file 3: SNP validation using Sanger sequencing. This file contains an example of an electropherogram generated by Sanger sequencing to demonstrate SNP validation, and showing the occurrence of two arising SNPs between different mapping family parents.

Additional file 4: Representative clustering patterns generated by the Illumina GoldenGate® SNP Genotyping assay. The file contains an example of clustering patterns obtained from SNP genotyping assays on two mapping populations. The data point colour codes represent: red, AA (homozygous); blue, BB (homozygous); purple, AB (heterozygous); black, no call (missing data). A) High-quality polymorphic SNP; B) Monomorphic SNP; C) SNP with a large number of heterozygous individuals; D) Failed SNP.

Additional file 5: Linkage map statistics. This file contains details of different markers (SSRs and SNPs) and their corresponding positions on different LGs.

Additional file 6: Synteny analysis statistics. This file details of field pea LGs, number of marker sequences and synteny with chickpea, *M. truncatula,* soybean, *L. japonicus* and pigeon pea chromosomes.

Additional file 7: Frequency distribution histogram. Frequency distribution for salinity index value and qualitative rating (T (tolerant), MT-T (moderately tolerant to tolerant), MS-S (Moderately sensitive to sensitive), S (sensitive), HS (high sensitivity) for Kaspa x Parafield RIL progeny following salinity treatment of 18 dS m^{-1}.

Additional file 8: Frequency distribution histogram. Frequency distribution for symptom score of Kaspa x Parafield RIL progeny at 7, 14, 21, 35, 42, 49 and 56 days post application of NaCl in watering solution at 18 dS m^{-1}.

Competing interests
The authors declare that they have no competing interests.

Authors' contributions
AL performed population development, phenotypic assessment, data interpretation and contributed to drafting the manuscript. SS performed marker discovery, map construction, QTL analysis, comparative genomics and contributed to drafting the manuscript. NC co-conceptualised the project, contributed to data interpretation and assisted in drafting the manuscript. PS and MN co-conceptualised the project and assisted in drafting the manuscript. JF, MM and SK co-conceptualised and coordinated the project and assisted in drafting the manuscript. All authors read and approved the final manuscript.

Acknowledgements
This work was supported by funding from the Victorian Department of Environment and Primary Industries and the Grains Research and Development Council, Australia.

Author details
[1]Department of Environment and Primary Industries, Biosciences Research Division, Grains Innovation Park, PMB 260, Horsham, VIC 3401, Australia. [2]Melbourne School of Land and Environment, University of Melbourne, Melbourne, VIC 3010, Australia. [3]Department of Environment and Primary Industries, Biosciences Research Division, AgriBio, Centre for AgriBioscience, 5 Ring Road, La Trobe University Research and Development Park, Bundoora, VIC 3083, Australia. [4]La Trobe University, Bundoora, VIC 3086, Australia.

Received: 26 August 2013 Accepted: 13 October 2013
Published: 17 October 2013

References
1. Kocer A, Albayrak S: **Determination of forage yield and quality of pea (*Pisum sativum* L.) mixtures with oat and barley.** *Turk J Field Crops* 2012, **17:**96–99.
2. Omokanye AT, Kelleher FM, McInnes A: **Low-input cropping systems and nitrogen fertilizer effects on crop production: Soil nitrogen dynamics and efficiency of nitrogen use in maize crop.** *American-Eurasian J Agric Environ Sci* 2011, **11:**282–295.
3. Chamarthi SK, Kumar A, Voung TD, Blair MW, Gaur PM, Nguyen HT, Varshney RK: **Trait mapping and molecular breeding**. In *Biology and Breeding of Food Legumes*. Edited by Pratap A, Kumar J. Oxfordshire: CAB International; 2011:296–313.
4. McPhee KE: **Pea**. In *Genome Mapping and Molecular Breeding in Plants: Pulses, Sugar and Tuber Crops. Volume 3*. Edited by Kole C. Berlin: Springer; 2007:33–47.
5. Redden B, Leonforte T, Ford R, Croser J, Slattery J: **Pea (*Pisum sativum* L.)**. In *Genetic Resources, Chromosome Engineering, and Crop Improvement*. Edited by Singh RJ. Florida, USA: CRC Press; 2005:49–83.
6. Kumar J, Choudhary AK, Solanki RK, Pratap AP: **Towards marker-assisted selection in pulses: a review.** *Plant Breed* 2011, **130:**297–313.
7. Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM: **Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing.** *BMC Genomics* 2011, **12:**227.
8. Kaur S, Pembleton LW, Cogan NO, Savin KW, Leonforte T, Paull J, Materne M, Forster JW: **Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers.** *BMC Genomics* 2012, **12:**265–276.
9. Jing Z, Qu Y, Yu C, Pan D, Fan Z, Chen J, Li C: **QTL analysis of yield-related traits using an advanced backcross population derived from common wild rice (*Oryza rufipogon* L).** *Mol Plant Breed* 2007, **1:**1–10.
10. Fondevilla S, Küster H, Krajinski F, Cubero JI, Rubiales D: **Identification of genes differentially expressed in a resistant reaction to *Mycosphaerella pinodes* in pea using microarray technology.** *BMC Genomics* 2011, **13:**12–28.
11. Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK: **An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts.** *Euphytica* 2005, **142:**169–196.
12. Timmerman-Vaughan GM, McCallum JA, Frew TJ, Weeden NF, Russell AC: **Linkage mapping of quantitative traits controlling seed weight in pea (*Pisum sativum* L.).** *Theor Appl Genet* 1996, **93:**431–439.
13. McCallum J, Timmerman-Vaughan G, Frew TJ, Russell AC: **Biochemical and genetic linkage analysis of green seed color in field pea (*Pisum sativum* L.).** *J Am Soc Hort Sci* 1997, **122:**218–225.

14. Weeden NF, Ellis THN, Timmerman-Vaughan GM, Swiecicki WK, Rozov SM, Berdnikov VA: **A consensus linkage map for *Pisum sativum*.** *Pisum Genet* 1998, **30:**1–4.
15. Pilet-Nayel L, Muehlbauer FJ, McGee RJ, Kraft JM, Baranger A, Coyne CJ: **Quantitative trait loci for partial resistance to Aphanomyces root rot in pea.** *Theor Appl Genet* 2002, **106:**28–39.
16. Prioul S, Frankewitz A, Deriot G, Morin G, Baranger A: **Mapping of quantativie trait locie for resistance to *Mycosopharella pinodes* in pea (*Pisum sativum*), at the seedling and adult plant stage.** *Theor Appl Genet* 2004, **108:**1322–1344.
17. Loridon K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, Rameau C, Baranger A, Coyne C, Lejeune-Hènaut I, Burstin J: **Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.).** *Theor Appl Genet* 2005, **111:**1022–1031.
18. Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Hénaut I, Burstin J, Aubert G: **Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea.** *BMC Genomics* 2010, **11:**468.
19. Krawczak M: **Informativity assessment for biallelic single nucleotide polymorphisms.** *Electrophoresis* 1999, **20:**1676–1681.
20. Xing C, Schumacher FR, Xing G, Lu Q, Wang T, Elston RC: **Comparison of microsatellites, single-nucleotide polymorphisms (SNPs) and composite markers derived from SNPs in linkage analysis.** *BMC Genet* 2005, **6:**S29.
21. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S: **Genome structure of the legume.** *Lotus japonicus DNA Res* 2008, **15:**227–239.
22. Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, N W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo MC, *et al*: **Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement.** *Nat Biotech* 2013, **31:**240–246.
23. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, *et al*: **Genome sequence of the palaeopolyploid soybean.** *Nature* 2012, **463:**178–183.
24. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, *et al*: **Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers.** *Nat Biotech* 2011, **30:**83–89.
25. Ondrasek G, Rengel Z, Veres S: **Soil Salinisation and Salt Stress in Crop Production**. In *Abiotic Stress in Plants - Mechanisms and Adaptations*. Edited by Shanker A.; ISBN: 978-953-307-394-1, InTech, doi:10.5772/22248. [http://www.intechopen.com/books/abiotic-stress-in-plants-mechanisms-and-adaptations/soil-salinisation-and-salt-stress-in-crop-production].
26. Nuttall JG, Hobson KB, Materne M, Moody DB, Munns R, Armstrong RD: **Use of genetic tolerance in grain crops to overcome subsoil constraints in alkaline cropping soils.** *Soil Research* 2009, **48:**188–199.
27. Rengasamy P: **Transient salinity and subsoil constraints to dryland farming in Australian sodic soils: an overview.** *Aust J Exp Agr* 2002, **42:**351–361.
28. Rengasamy P: **World salinization with emphasis on Australia.** *J Exp Bot* 2006, **57:**1017–1023.
29. Kabir AH, Paltridge NG, Able AJ, Paull JG, Stangoulis JC: **Natural variation for Fe-efficiency is associated with upregulation of Strategy I mechanisms and enhanced citrate and ethylene synthesis in *Pisum sativum* L.** *Planta* 2012, **235:**1409–1419.
30. Bagheri A, Paull JG, Rathjen AJ: **The response of *Pisum sativum* L. germplasm to high concentrations of soil boron.** *Euphytica* 1994, **75:**9–17.
31. Bagheri A, Paull JG, Rathjen AJ: **Genetics of tolerance to high concentrations of soil boron in peas (*Pisum sativum* L.).** *Euphytica* 1996, **87:**69–75.
32. Leonforte A, Noy D, Redden R, Enneking D: **Improving boron and salinity tolerance in field pea (*Pisum sativum* L.)**. In *Proceedings of the 14th Australasian Plant Breeding (APB) Conference and 11th Society for the Advancement of Breeding Researches in Asia and Oceania (SABRAO) Conference 2009*. Queensland, Australia: Cairns; 2009.
33. Maas EV: **Salt tolerance of plants.** *Applied Agric Res* 1986, **1:**12–26.
34. Saxena NP, Saxena MC, Ruckenbauer P, Rana RS, El-Fouly MM, Shabana R: **Screening techniques and sources of tolerance to salinity and mineral nutrient imbalances in cool season food legumes.** *Euphytica* 1994, **73:**85–93.
35. Francois LE, Maas EV: **Crop response and management on salt-affected soils**. In *Handbook of plant and crop stress*. Edited by Pessarakli M. New York: Marcel Dekker; 1994:149–181.
36. Steppuhn H, Volkmar KM, Miller PR: **Comparing canola, field pea, dry bean, and durum wheat crops grown in saline media.** *Crop Sci* 2001, **41:**1827–1833.
37. Dua RP, Sharma SK, Mishra B: **Response of broad bean (*Vicia faba*) and pea (*Pisum sativum*) varieties to salinity.** *Indian J Agr Sci* 1989, **59:**729–731.
38. Hernandez JA, Jimenez A, Mullineaux P, Sevilla F: **Tolerance of pea (*Pisum sativum* L.) to long-term salt stress is associated with induction antioxidant defences.** *Plant Cell Environ* 2000, **23:**853–862.
39. El-Hamdaoui A, Redondo-Nieto M, Rivilla R, Bonilla I, Bolaños L: **Effects of boron and calcium nutrition on the establishment of the *Rhizobium leguminosarum* – pea (*Pisum sativum*) symbiosis and nodule development under salt stress.** *Plant Cell Environ* 2003, **26:**1003–1011.
40. Gomez JM, Jiménez A, Olmos E, Sevilla F: **Location and effects of long-term NaCl stress on superoxide dismutase and ascorbate peroxidase isoenzymes of pea (*Pisum sativum* cv. Puget) chloroplasts.** *J Exp Bot* 2004, **55:**119–130.
41. Leonforte A, Forster JW, Redden RJ, Nicolas ME, Salisbury PA: **Sources of high tolerance to salinity in pea (*Pisum sativum* L.).** *Euphytica* 2013, **189:**203–216.
42. Cordovilla MP, Ligero F, Lluch C: **Influence of host genotypes on growth, symbiotic performance and nitrogen assimilation in faba bean (*Vicia faba* L.) under salt stress.** *Plant Soil* 1995, **172:**289–297.
43. Sadiki M, Rabih K: **Selection of chickpea (*Cicer arietinum*) for yield and symbiotic nitrogen fixation ability under salt stress.** *Agronomie* 2001, **21:**659–666.
44. Maher L, Armstrong R, Connor D: **Salt tolerant lentils - a possibility for the future?** In *Proceedings of the 11th Australian Agronomy Conference: Feb 2003*. Geelong, Victoria: Australian Society of Agronomy; 2003:2–6.
45. Munns R, James R, Lauchli A: **Approaches to increasing the salt tolerance of wheat and other cereals.** *J Exp Bot* 2006, **57:**1025–1043.
46. Leonforte A, Noy D, Forster JW, Salisbury P: **Evaluation for higher tolerance to NaCl in *Pisum sativum* L**. In *Proceedings of the 5th International Research Conference*. Turkey: Anatalya; 2010.
47. Schuelke M: **An economic method for the fluorescent labelling of PCR fragments.** *Nat Biotechnol* 2000, **18:**233–234.
48. van Ooijen J, Voorrips R: *JoinMap 3.0 software for the calculation of genetic linkage maps*. Wageningen, the Netherlands: Plant Research International; 2001.
49. Brown DG, Vision TJ: *MapPop version 1.0: Software for selective mapping and bin mapping*. 2000. http://www.bio.unc.edu/faculty/vision/lab/mappop/].
50. Manly KF, Cudmore RH Jr, Meer JM: **Map Manager QTX, cross-platform software for genetic mapping.** *Mamm Genome* 2001, **12:**930–932.
51. Kosambi DD: **The estimation of map distance from recombination values.** *Ann Eugen* 1944, **12:**172–175.
52. Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR: **Estimating genome conservation between crop and model legume species.** *Proc Natl Acad Sci U S A* 2004, **101:**15289–15294.
53. Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB: **Comparative mapping between *Medicago sativa* and *Pisum sativum*.** *Mol Gen Genomics* 2004, **272:**235–246.
54. Voorrips RE: **MapChart: Software for the graphical presentation of linkage maps and QTLs.** *J Hered* 2002, **93:**77–78.
55. Barker GL, Edwards KJ: **A genome-wide analysis of single nucleotide polymorphism diversity in the world's major cereal crops.** *Plant Biotechnol J* 2009, **7:**318–325.
56. Muchero W, Diop NN, Bhat PR, Fenton RD, Wanamaker S, Pottorff M, Hearne S, Cisse N, Fatokun C, Ehlers JD, Roberts PA, Close TJ: **A consensus genetic map of cowpea [*Vigna unguiculata* (L) Walp.] and synteny based on EST-derived SNPs.** *Proc Natl Acad Sci U S A* 2009, **106:**18159–18164.
57. Choi HK, Kim DJ, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR: **A sequence-**

based genetic map of *Medicago truncatula* and comparison of marker colinearity with *M. sativa*. *Genetics* 2004, **166:**1463–1502.
58. Choi IY, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon MS, Hwang EY, Yi SI, Young ND, Shoemaker RC, van Tassell CP, Specht JE, Cregan PB: **A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis.** *Genetics* 2007, **176:**685–696.
59. Hand ML, Ponting RC, Drayton MC, Lawless KA, Cogan NO, Charles Brummer E, Sawbridge TI, Spangenberg GC, Smith KF, Forster JW: **Identification of homologous, homoeologous and paralogous sequence variants in an outbreeding allopolyploid species based on comparison with progenitor taxa.** *Mol Genet Genomics* 2008, **280:**293–304.
60. Gaur R, Azam S, Jeena G, Khan AW, Choudhary S, Jain M, Yadav G, Tyagi AK, Chattopadhyay D, Bhatia S: **High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum* L.).** *DNA Res* 2012, **19:**357–373.
61. Holliday R, Grigg GW: **DNA methylation and mutation.** *Mutat Res* 1993, **285:**61–67.
62. Ellis THN, Turner L, Hellens RP, Lee D, Harker CL, Enard C, Domoney C, Davies DR: **Linkage maps in pea.** *Genetics* 1992, **130:**649–663.
63. Moreno RR: **Localization and characterization of yield component quantitative trait loci (QTLs) in Recombinant Inbred Lines (RILs) of pea, *Pisum sativum* ssp**. In *PhD thesis.* : Northern Illinois University; 2009.
64. Liu G, Bernhardt JL, Jia MH, Wamishe YA, Jia Y: **Molecular characterization of the recombinant inbred line population derived from a japonica-indica rice cross.** *Euphytica* 2008, **159:**73–82.
65. Knox MR, Ellis THN: **Map expansion in pea.** *Genetics* 2002, **162:**861–873.
66. Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM: **Development of a large SNP genotyping array and generation of high-density genetic maps in tomato.** *PLoS ONE* 2012, **7:**e40563.
67. Shirasawa K, Bertioli DJ, Varshney RK, Moretzsohn MC, Leal-Bertioli SC, Thudi M, Pandey MK, Rami JF, Foncéka D, Gowda MV, Qin H, Guo B, Hong Y, Liang X, Hirakawa H, Tabata S, Isobe S: **Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergence of the legume genomes.** *DNA Res* 2013, **20:**173–184.
68. Hiremath PJ, Kumar A, Penmetsa RV, Farmer A, Schlueter JA, Chamarthi SK, Whaley AM, Carrasquilla-Garcia N, Gaur PM, Upadhyaya HD, Kavi Kishor PB, Shah TM, Cook DR, Varshney RK: **Large-scale development of cost-effective SNP marker assays for diversity assessment and genetic mapping in chickpea and comparative mapping in legumes.** *Plant Biotechnol J* 2012, **10:**716–732.
69. Nayak SN, Zhu H, Varghese N, Datta S, Choi HK, Horres R, Jüngling R, Singh J, Kavi Kishor PB, Sivaramakrihnan S, Hoisington DA, Kahl G, Winter P, Cook DR, Varshney RK: **Integration of novel SSR and gene-based SNP marker loci in the chickpea genetic map and establishment of new anchor points with *Medicago truncatula* genome.** *Theor Appl Genet* 2010, **120:**1415–1441.
70. Aubert G, Morin J, Jacque F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Hénaut I, Huguet T, Burstin J: **Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*.** *Theor Appl Genet* 2006, **112:**1024–4.
71. Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KF, Rogers J, Quétier F, Oldroyd GE, Debellé F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, Van de Peer Y, Young ND: **Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes.** *Proc Natl Acad Sci* 2006, **103:**14959–14964.
72. Arumuganathan K, Earle ED: **Nuclear DNA content of some important plant species.** *Plant Mol Biol Rep* 1991, **9:**208–218.
73. Wojciechowski MF: **Reconstructing the phylogeny of legumes (Leguminosae): an early 21st century perspective**. In *Advances in Legume Systematics. part 10.* Edited by Klitgaard BB, Bruneau A. Royal Botanic Gardens, Kew: Higher Level Systematics; 2003:5–35.
74. Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ, Ellis THN: **Transposable elements reveal the impact of introgression, rather than rransposition, in *Pisum* diversity, evolution, and domestication.** *Mol Biol Evol* 2003, **20:**2067–2075.
75. Doyle JJ, Luckow MA: **The Rest of the Iceberg, Legume Diversity and Evolution in a Phylogenetic Context.** *Plant Physiol* 2003, **131:**900–910.
76. Saxena RK, Penmetsa RV, Upadhyaya HD, Kumar A, Carrasquilla-Garcia N, Schlueter JA, Farmer A, Whaley AM, Sarma BK, May GD, Cook DR, Varshney RK: **Large-scale development of cost-effective single-nucleotide polymorphism marker assays for genetic mapping in pigeonpea and comparative mapping in legumes.** *DNA Res* 2012, **19:**449–461.
77. Isobe SN, Hisano H, Sato S, Hirakawa H, Okumura K, Shirasawa K, Sasamoto S, Watanabe A, Wada T, Kishida Y, Tsuruoka H, Fujishiro T, Yamada M, Kohara M, Tabata S: **Comparative genetic mapping and discovery of linkage disequilibrium across linkage groups in white clover (*Trifolium repens* L.).** *G3: Genes, Genomes, Genet* 2012, **2:**607–17.
78. George J, Sawbridge TI, Cogan NOI, Gendall AR, Smith KF, Spangenberg GC, Forster JW: **Comparison of genome structure between white clover and *Medicago truncatula* supports homoeologous group nomenclature based on conserved synteny.** *Genome* 2008, **51:**905–911.
79. Foolad MR: **Recent advances in genetics of salt tolerance in tomato.** *Plant Cell Tiss Org* 2004, **76:**101–119.
80. Arraouadi S, Chardon F, Huguet T, Aouani ME, Badri M: **QTL mapping of morphological traits related to salt tolerance in *Medicago truncatula*.** *Acta Physiol Plant* 2011, **33:**917–929.
81. Yaish MW, Colasanti J, Rothstein SJ: **The role of epigenetic processes in controlling flowering time in plants exposed to stress.** *J Exp Bot* 2011, **62:**3727–3735.
82. Aydi S, Sassi S, Debouba M, Hessini K, Larrainzar E, Gouia H, Abdelly C: **Resistance of *Medicago truncatula* to salt stress is related to glutamine synthetase activity and sodium sequestration.** *J Plant Nutr Soil Sc* 2010, **173:**892–899.
83. Xu WF, Shi WM: **Expression profiling of the 14-3-3 gene family in response to salt stress and potassium and iron deficiencies in young tomato (*Solanum lycopersicum*) roots: analysis by real-time RT-PCR.** *Ann Bot (Lond)* 2006, **98:**965–974.
84. Ouyang SQ, Liu YF, Liu P, Lei G, He SJ, Ma B, Zhang WK, Zhang JS, Chen SY: **Receptor-like kinase OsSIK1 improves drought and salt stress tolerance in rice (*Oryza sativa*) plants.** *Plant J* 2010, **62:**316–329.
85. Nagaoka S, Takano T: **Salt tolerance-related protein STO binds to a Myb transcription factor homologue and confers salt tolerance in *Arabidopsis*.** *J Exp Bot* 2003, **54:**2231–2237.
86. Alam R, Rahman MS, Seraj ZI, Thomson MJ, Ismail AM, Tumimbang-Raiz E, Gregorio GB: **Investigation of seedling stage salinity tolerance QTLs using backcross lines derived from *Oryza sativa* L. Pokkali.** *Plant Breed* 2011, **130:**430–437.
87. Zhou G, Johnson P, Ryan PR, Delhaize E, Zhou M: **Quantitative trait loci for salinity tolerance in barley (*Hordeum vulgare* L.).** *Mol Breed* 2012, **29:**427–439.
88. Hamwieh AD, Tuyen D, Cong H, Benitez ER, Takahashi R, Xu DH: **Identification and validation of a major QTL for salt tolerance in soybean.** *Euphytica* 2011, **179:**451–459.
89. Ma LQ, Zhou EF, Huo NZ, Zhou RH, Wang GY, Jia JZ: **Genetic analysis of salt tolerance in a recombinant inbred population of wheat (*Triticum aestivum* L.).** *Euphytica* 2007, **153:**109–117.
90. Foolad MR, Lin GY: **Absence of a relationship between salt tolerance during germination and vegetative growth in tomato.** *Plant Breed* 1997, **116:**363–367.
91. Foolad MR, Chen FQ: **RAPD markers associated with salt tolerance in and interspecific cross of tomato (*Lycospersicon esculentum* x *L. pennelli*).** *Plant Cell Reprod* 1998, **17:**306–312.
92. Cheng HT, Jiang H, Xue DW, Guo LB, Zeng DL, Zhang GH, Qian Q: **Mapping of QTL underlying tolerance to alkali at germination and early seedling stages in rice.** *Acta Agronomica Sinica* 2008, **34:**1719–1727.
93. Mano Y, Takeda K: **Mapping quantitative trait loci for salt tolerance at germination and the seedling stage in barley (*Hordeum vulgare* L.).** *Euphytica* 1997, **94:**263–272.
94. Manneh B, Stam P, Struik PC, Bruce-Oliver S, van Eeuwijk FA: **QTL-based analysis of genotype-by-environment interaction for grain yield of rice in stress and non-stress environments.** *Euphytica* 2007, **156:**213–226.
95. Xue DW, Zhou MZ, Zhang XQ, Chen S, Wei K, Zeng FR, Mao Y, Wu FB, Zhang GP: **Identification of QTLs for yield and yield components of barley under different growth conditions.** *J Zheijang Univ Sci B* 2012, **11:**169–176.
96. Villalta I, Bernet GP, Carbonell EA, Asins MJ: **Comparative QTL analysis of salinity tolerance in terms of fruit yield using two *Solanum* populations of F-7 lines.** *Theor Appl Genet* 2007, **114:**1001–1017.

97. Ashraf M, Foolad MR: **Crop breeding for salt tolerance in the era of molecular markers and marker-assisted selection.** *Plant Breed* 2013, **132:**10–20.
98. Collins NC, Tardieu F, Tuberosa R: **Quantitative trait loci and crop performance under abiotic stress: where do we stand?** *Plant Physiol* 2008, **147:**469–486.
99. Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE: **A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species.** *PLoS One* 2011, **4:**e19379.

doi:10.1186/1471-2229-13-161
Cite this article as: Leonforte *et al.*: SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). *BMC Plant Biology* 2013 **13**:161.

CHAPTER 3

Consensus genetic map construction for field pea (*Pisum sativum* L.), trait dissection of biotic and abiotic stress tolerance and development of a diagnostic marker for the *er1* powdery mildew resistance gene

3.1 Chapter preface

This study generated genetic linkage maps for two RIL-based mapping populations (Kaspa x ps1771 and Kaspa x Yarrum) of field pea based on existing SSR and SNP markers. A composite map was generated by the merger of the Kaspa x ps1771, Kaspa x Yarrum and the previously published Kaspa x Parafield genetic linkage map. By combining this composite map with previously published consensus map, a comprehensive consensus map was generated. Trait dissection of B toxicity tolerance in the Kaspa x ps1771 population identified a single genomic region of large magnitude on Ps VI. QTL analysis for powdery mildew resistance was performed for both the Kaspa x ps1771 and Kaspa x Yarrum populations, identifying a single common genomic region on Ps VI. Resequencing of the *PsMLO1* candidate gene from powdery mildew resistant and susceptible genotypes allowed the design and validation of a putative diagnostic marker.

This chapter is presented in published format.

3.2 Publication details

Title: Consensus genetic map construction for field pea (*Pisum sativum* L.), trait dissection of biotic and abiotic stress tolerance and development of a diagnostic marker for the *er1* powdery mildew resistance gene

Journal Details: Plant Molecular Biology Reporter. 2015, 33:1391-1403. doi:10.1007/s11105-014-0837-7.

Stage of publication: Published

Authors: Shimna Sudheesh, Maria Lombardi, Antonio Leonforte, Noel OI Cogan, Michael Materne, John W Forster, Sukhjiwan Kaur

3.3 Statement of contributions of joint authorship

Shimna Sudheesh performed DNA extraction, genotyping of the mapping population, bi-parental map construction, composite and consensus map generation, QTL analysis and drafted the manuscript.

Maria Lombardi performed the PsMLO1 marker development.

Antonio Leonforte developed the populations and performed the phenotypic assessments.

Noel OI Cogan and Sukhjiwan Kaur co-conceptualised the project, contributed to data interpretation and assisted in drafting the manuscript.

Michael Materne co-conceptualised the project.

John W Forster co-conceptualised the project and assisted in drafting the manuscript.

3.4 Statement from the co-author confirming the authorship contribution of the PhD candidate

"As co-author of the manuscript 'Sudheesh *et al.* (2015) Consensus genetic map construction for field pea (*Pisum sativum* L.), trait dissection of biotic and abiotic stress tolerance and development of a diagnostic marker for the *er1* powdery mildew resistance gene. Plant Molecular Biology Reporter. 33:1391-1403', I confirm that Shimna Sudheesh has made the following contributions,

- DNA extraction of the mapping population
- SSR and SNP genotyping of the mapping populations
- Bi-parental map construction
- Composite and consensus map generation
- QTL analysis
- Data analysis and interpretation
- Writing the manuscript, critical appraisal of content and response to reviewers"

Dr Sukhjiwan Kaur

Date: 15/09/2015

Plant Mol Biol Rep (2015) 33:1391–1403
DOI 10.1007/s11105-014-0837-7

ORIGINAL PAPER

Consensus Genetic Map Construction for Field Pea (*Pisum sativum* L.), Trait Dissection of Biotic and Abiotic Stress Tolerance and Development of a Diagnostic Marker for the *er1* Powdery Mildew Resistance Gene

Shimna Sudheesh · Maria Lombardi · Antonio Leonforte · Noel O. I. Cogan · Michael Materne · John W. Forster · Sukhjiwan Kaur

Published online: 15 December 2014

Abstract Field pea (*Pisum sativum* L.) is a cool-season legume that is cultivated worldwide for both human consumption and stockfeed. Genetic improvement is essential for enhanced crop production and management of field pea, especially to deliver superior varieties adapted to various biotic and abiotic stresses. A detailed understanding of the genetic basis of such stress tolerances is hence desirable. Genetic linkage maps based on single nucleotide polymorphisms (SNP) and simple sequence repeat (SSR) markers have been developed from two recombinant inbred line (RIL) populations generated by crossing phenotypically divergent parental genotypes. The Kaspa×Yarrum map contained 428 loci across 1910 cM, while the Kaspa×ps1771 map contained 451 loci across 1545 cM. Data from these maps were combined through bridging markers with those from previously published studies to generate a consensus structure including 2028 loci distributed across seven linkage groups (LGs), with a cumulative length of 2387 cM at an average density of one marker per 1.2 cM. Trait dissection of powdery mildew resistance was performed for both RIL populations, identifying a single genomic region of large magnitude in the same genomic region on Ps VI, which were inferred to correspond to the *er1* gene. Equivalent studies of the Kaspa×ps1771 RIL population identified a major quantitative trait locus (QTL) for boron tolerance that coincided with the disease resistance-controlling locus, permitting strategies of co-selection for these desirable traits. Resequencing of the *PsMLO*1 candidate gene from resistant and susceptible genotypes allowed design and validation of a putative diagnostic marker for powdery mildew resistance. The availability of a highly saturated consensus map, linked markers for key biotic and abiotic stress tolerances and a diagnostic marker for the agronomically important *er1* gene provide important resources for field pea molecular breeding programs.

Keywords Grain legume · Simple sequence repeat · Single nucleotide polymorphism · Disease resistance · Abiotic stress tolerance · Molecular breeding

Electronic supplementary material The online version of this article (doi:10.1007/s11105-014-0837-7) contains supplementary material, which is available to authorized users.

S. Sudheesh · M. Lombardi · N. O. I. Cogan · J. W. Forster (✉) · S. Kaur
Biosciences Research Division, AgriBio, Centre for AgriBioscience, Department of Environment and Primary Industries, 5 Ring Road, Bundoora, Victoria 3083, Australia
e-mail: john.forster@depi.vic.gov.au

A. Leonforte · M. Materne
Biosciences Research Division, Grains Innovation Park, Department of Environment and Primary Industries, Horsham, Victoria 3401, Australia

S. Sudheesh · J. W. Forster
School of Applied Systems Biology, La Trobe University, Bundoora, Victoria 3086, Australia

Introduction

Field pea (*Pisum sativum* L.) is an important grain legume, which is consumed both as human food and animal feed. Field pea is the fourth most extensively cultivated pulse crop on a global basis and is widely grown in Europe (Rubiales et al. 2009; Pavan et al. 2011). Of various abiotic and biotic stresses, toxicity effects due to soil boron (B) and damaging infection by powdery mildew are among the most serious threats to pea cultivation. B toxicity is a major problem in several major production zones, including southern Australia, India, Pakistan, Iraq, Peru and the USA (Yau and Ryan 2008). High B concentrations occur at depths between 40 and

100 cm in the soil profile, inhibiting both crop growth and grain yield (Cartwright et al. 1984; Nuttall et al. 2003). In Australia, typical patterns of low rainfall exacerbate B toxicity. For instance, about 15 % of agricultural soils in Western Australia are at moderate to high risk of toxicity effects (Lacey and Davies 2009), and high B concentrations are also frequently associated with elevated salinity, levels of each stress varying seasonally according to precipitation. However, in low-rainfall environments and on clay soils, B is leached more slowly than salt (Bennett 2012). Due to these factors, B toxicity is difficult to manage by manipulation of agricultural systems, creating an incentive for solutions based on genetic improvement.

To date, limited progress has been made in understanding the genetics of B tolerance in field pea. Bagheri et al. (1996) detected two major genes for tolerance to high B concentrations, which interact with each other with incomplete dominance at each locus. However, recent studies in closely related species, such as the model legume *Medicago truncatula* Gaertn. (Bogacki et al. 2013) and cultivated lentil (Kaur et al. 2014), have indicated the effects of a single dominant locus. Several resistant cultivars have been identified within field pea breeding programs through phenotypic screening (Bennett 2012; A. Leonforte, personal communication), but no information on genomic regions associated with B tolerance in field pea is currently available.

Powdery mildew of pea, due to infection by the fungal pathogen *Erysiphe pisi* DC, causes yield losses of 25–50 % (Munjal et al. 1963; Warkentin et al. 1996; Janila and Sharma 2004; Katoch et al. 2010). The alternate pathogens *Erysiphe baeumleri* (Magnus) U, Braun & S. Takam. 2000 and *Erysiphe trifolii* Grev. have also been reported to cause powdery mildew of pea (Ondřej et al. 2005; Attanayake et al. 2010). Protection by treatment with fungicides provides a temporary control option. However, identification of resistance sources and incorporation into contemporary cultivars remain the most effective method for disease control (Janila and Sharma 2004; Katoch et al. 2010). The genetics of resistance to powdery mildew in pea is relatively well understood, three major loci (*er1*, *er2* and *Er3*) having been described (Smýkal et al. 2012). The majority of resistant cultivars rely on the presence of the recessive *er1* gene (Harland 1948; Timmerman et al. 1994; Vaid and Tyagi 1997; Janila and Sharma 2004). However, some studies have also detected a second recessive gene, *er2* (Heringa et al. 1969; Sokhi et al. 1979; Kumar and Singh 1981). Slightly differing mechanisms have been reported for the two genes, such that *er1* confers systemic resistance under field and controlled conditions, by prevention of epidermal cell penetration resulting in formation of very few haustoria or colonies. In contrast, the effects of *er2* are largely confined to leaves, in which expression is influenced by both temperature and leaf age, such that complete resistance is only observed at 25 °C or in mature leaves. The *er2*-derived resistance is primarily based on reduced penetration rate, along with post-penetration cell death in mature leaves (Fondevilla et al. 2006; Marx 1986; Tiwari et al. 1997). Only *er1* has been extensively used in breeding programs, and it is considered as stable and effective. In contrast, *Er3* is a dominant resistance locus recently characterised from a related species, tawny pea (*Pisum fulvum* Sibth. & Sm.) and is known to segregate independently of *er1* and *er2* (Fondevilla et al. 2011).

DNA-based genetic markers provide powerful tools for identification and location of genes for agronomically important characters, and subsequent selection for trait introgression in plant breeding programs. As previously noted, no such studies have yet been performed for B tolerance in field peas, but in the case of powdery mildew resistance, markers belonging to various classes such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), sequence characterised region (SCAR), sequence-tagged site (STS), SSR and SNP have been linked to *er1* on LG VI (Dirlewanger et al. 1994; Ek et al. 2005; Janila and Sharma 2004; Timmerman et al. 1994; Tiwari et al. 1998; Katoch et al. 2010; Tonguç and Weeden 2010; Pavan et al. 2013). The *er2* gene was localised to pea LG III, in linkage with various amplified fragment length polymorphism (AFLP) and SCAR markers (Tiwari et al. 1999; Katoch et al. 2010). In addition, SCAR markers in linkage with the *Er3* gene have been reported (Fondevilla et al. 2008). SNP markers are highly prevalent, usually biallelic and co-dominant in nature, and amenable to development of cost-effective and high-throughput marker systems that can provide sufficiently dense genome coverage for the dissection of key traits. In addition, discovery of SNPs from genic sequences, such as expressed sequence tags (ESTs), permits marker development from functionally associated sequences. Large-scale resources for field pea have been obtained from transcriptome sampling (Kaur et al. 2012). Consequently, SNPs provide the marker class of choice for determination of the genetic basis of agronomically important traits.

Genetic linkage maps are an essential prerequisite for the identification and localisation of genes for agronomically important characters. Several linkage maps have been constructed for pea based on different kinds of markers, including morphological markers, isoenzymes, RFLPs, RAPDs, SSRs, and SNPs (Weeden et al. 1996; Weeden and Boone 1999; Timmerman-Vaughan et al. 2000; Ellis and Poyser 2002; Loridon et al. 2005; Aubert et al. 2006; Deulvot et al. 2010). Recently, SNP-based linkage maps have been developed for field pea, which have been used for dissection of the genetic basis for salinity stress tolerance (Leonforte et al. 2013). However, individual linkage maps may suffer limitations in terms of practical application due to low marker density. Consensus linkage maps, which combine information from multiple mapping populations, have been developed for many

crop species, including pea (Cloutier et al. 2012; Millan et al. 2010; Weeden et al. 1998; Duarte et al. 2014). Consensus maps offer the following advantages: higher marker density in a single map, and hence more complete genome coverage; opportunities to determine the relative position of common markers across mapping populations; determination of conserved marker locus location; and identification of chromosomal rearrangements and degree of gene duplication (Milczarski et al. 2011; Blenda et al. 2012). Consensus maps have been constructed for several crops using software programs such as JoinMap (Shirasawa et al. 2013) and MergeMap (Gautami et al. 2012). JoinMap accounts for both sizes and structures of populations in order to estimate marker order and genetic distance using common or bridging markers (Stam 1993). In MergeMap, individual maps are first converted to directed acyclic graphs (DAGs) internally with nodes representing the mapped markers and edges defining the order of adjacent markers (Wang et al. 2011). DAGs are then merged into a consensus graph on the basis of shared vertices. Conflicts between the individual maps are apparent as cycles in the consensus graph, and MergeMap attempts to resolve such conflicts by deletion of minimum set of marker occurrences (Wu et al. 2011).

Complementary trait-specific mapping families will permit analysis of characters such as B toxicity tolerance and powdery mildew resistance, along with their mutual interactions, when these key stress tolerance traits are present in the same germplasm. SNP-based genetic maps also permit comparative analysis with related species for identification of candidate genes. In the case of cool-season legumes, species with full genome-sequence species such as the models *M. truncatula* and *Lotus japonicus* L., and the cultivated species soybean (*Glycine max* L. [Merr.]) and chickpea (*Cicer arietinum* L.) provide appropriate comparators. Candidate gene information may be further used to develop potential diagnostic markers directly applicable to selection and gene pyramiding in breeding programs. Recently, through the study of a novel chemically induced allele of the *er1* gene, co-segregation with a loss-of-function mutation at the *PsMLO1* (mildew resistance locus O) gene was reported. Analysis of the gene sequence from several known powdery mildew-resistant cultivars further supported the hypothesis that a loss-of-function in *PsMLO1* is responsible for the trait, and indicated that the molecular basis is shared with well-known powdery mildew immunity determinants in barley (*mlo*) and tomato (*ol-2*) (Pavan et al. 2010; Humphry et al. 2011).

The present study describes the generation of high-density genetic linkage maps, based exclusively on SSRs and SNPs, from two biparental field pea RIL mapping populations (Kaspa×Yarrum and Kaspa×ps1771) that exhibit variation for both B tolerance and powdery mildew resistance; development of a consensus map of field pea combining the maps from these trait-specific populations, a previously described RIL family (Leonforte et al. 2013) and an existing integrated pea map; identification of genomic regions controlling the target traits, based on phenotypic data collected from glasshouse-based nursery screens in Victoria, Australia; and development of a potential diagnostic marker for powdery mildew resistance based on resequencing analysis of the *PsMLO1* gene.

Materials and Methods

Plant Materials

Crosses were made between single genotypes selected from cultivar Kaspa (sensitive to B toxicity, susceptible to powdery mildew) with each of Yarrum (resistant to powdery mildew) and ps1771 (tolerant to B toxicity and resistant to powdery mildew) at DEPI-Horsham, Victoria, Australia, and F_2 generation progeny were produced. Single seed descent was undertaken from F_2 progeny-derived genotypes for four generations in the glasshouse, and a total of 106 F_6 RILs were subsequently generated from each of the Kaspa×Yarrum and Kaspa×ps1771 populations.

Plants were grown under glasshouse conditions at 20±2 °C under a 16/8-h (light/dark) photoperiod. Genomic DNA was extracted from young leaves using the DNeasy® 96 Plant Kit (QIAGEN, Hilden, Germany) according to manufacturer's instructions. Approximately one to two leaves per sample were used for each extraction and were ground using a Mixer Mill 300 (Retsch®, Haan, Germany). DNA was resuspended in Milli-Q water to a concentration of 50 ng/µl and stored at −20 °C until further use.

SSR and SNP Genotyping

A total of 242 genomic DNA- and EST-derived SSRs (Loridon et al. 2005; Kaur et al. 2012) were screened on the mapping parents for polymorphism detection. Primer synthesis and PCR amplifications were performed as described previously (Schuelke 2000; Kaur et al. 2012). PCR products were combined with the ABI GeneScan LIZ500 size standard and analysed using an ABI3730xl (Life Technologies Australia Pty Ltd, Victoria, Australia) capillary electrophoresis platform according to the manufacturer's instructions. Allele sizes were scored using GeneMapper® 3.7 software package (Life Technologies Australia Pty Ltd).

For SNP genotyping, a previously described set of 768 SNPs (Leonforte et al. 2013) was selected for genotyping using the GoldenGate™ oligonucleotide pooled assay (OPA). A total of 250 ng of genomic DNA from each genotype was used for amplification, after which PCR products were hybridised to bead chips via the address sequence for detection on an Illumina iSCAN Reader. On the basis of

obtained fluorescence, allele call data were viewed graphically as a scatter plot for each marker assayed using GenomeStudio software v2011.1 (Illumina) with a GeneCall threshold of 0.20.

Genetic Linkage Mapping

Data obtained from SNP and SSR genotyping were tested for goodness-of-fit to the expected Mendelian ratio of 1:1 using χ^2 analysis ($P<0.05$). The genetic linkage map was generated using Map Manager Software version QTXb19 (Manly et al. 2001). Map distances were calculated using the Kosambi mapping function (Kosambi 1944) at a threshold LOD score of 3. LGs were assigned on the basis of marker locus commonality with publicly available linkage maps of pea (Loridon et al. 2005), and by comparison with chromosomes of *M. truncatula* (Choi et al. 2004; Kalo et al. 2004). LGs were drawn using Mapchart software v 2.2 (Voorrips 2002).

Consensus Map Construction

The Kaspa×Yarrum- and Kaspa×ps1771-derived maps from the current study were combined with the Kaspa×Parafield map of Leonforte et al. (2013), which shared a high proportion of common markers, in order to generate a preliminary composite map. MergeMap (Wu et al. 2011) converted the individual maps into DAGs that were merged in a consensus graph on the basis of shared vertices (Yap et al. 2003). A comparative analysis of this preliminary composite map with the integrated pea map of Duarte et al. (2014) was performed using BLAST-based sequence analysis. Similarity searches were performed with DNA sequences underlying SNP markers assigned to the preliminary composite map against transcriptome sequencing data (DDBJ/EMBL/GenBank under the accession GAMJ00000000) with a threshold *E* value of 10^{-10}. After the identification of common sequences, the composite and integrated maps were melded into a single consensus structure using MergeMap (Wu et al. 2011). The consensus map for each LG was visualised by MapChart (Voorrips 2002).

Phenotypic Screening

Boron Toxicity

All individuals from the Kaspa×ps1771 RIL mapping populations were screened for response to B-induced stress applied at the seedling stage. Screening was undertaken by sowing three replicates of two plants each from each RIL-derived line at equidistant spacing in PVC cores using sandy clay soil. Two B concentration regimes were applied in soluble form as boric acid: 0 ppm (control) and 10 ppm. B toxicity was measured as visual assessment of leaf and stem necrotic symptoms on a 0–10 scale (Additional file 1a) performed over three different time intervals. Phenotyping data was analysed to estimate means after adjustments for any spatial effects within the trial. This was performed by calculating the averages for plant symptom score from individual plant assessments, which were then used to estimate genotype-specific average values for symptom score using residual maximum likelihood (REML) spatial row-column analysis. An index was used to quantify genotypic boron tolerance values and to describe tolerance levels according to sensitivity based on weighted symptom scores and final biomass. Models were fitted using REML as implemented in GenStat (GenStat Committee 2002 and previous releases). Best linear unbiased predictions (BLUP) analysis was used to calculate narrow-sense heritability. Means of symptom rating from each data set were used to construct distribution histograms in order to deduce Mendelian inheritance models for the trait.

Powdery Mildew Resistance

The parents and RIL progeny of the Kaspa×Yarrum and Kaspa×ps1771 populations were sown in pots under controlled environment conditions at DEPI-Horsham, Victoria, with three replicates. Plants were left in the glasshouse during the spring season to allow natural incidence of powdery mildew due to infection by *E. pisi*. The identity of the pathogen, and absence of other *Erysiphe* species (such as *E. trifolii*, which is capable of overcoming *er1* and *Er3*, but not *er2*: Fondevilla et al. (2013)), was confirmed by the Senior Pathologist, Dr. Grant Holloway (pers. comm.). For the Kaspa×Yarrum mapping population, the disease reaction was recorded on a 0–5 scale (Additional file 1b) based on the percentage of the infected foliage area. For the Kaspa×ps1771 population, whole plant symptom status was recorded as either resistant or susceptible reaction. Phenotypic assessment data were analysed to estimate means after adjustment for any spatial patterning within the trial. Models were fitted using REML as implemented in GenStat (GenStat Committee 2002 and previous releases). Means of symptom ratings from each individual of the mapping populations were used to construct distribution histograms in order to determine the mode of inheritance for the trait.

QTL Analysis

QTL detection was performed using marker regression, simple interval mapping (SIM) and composite interval mapping (CIM) in QTL Cartographer v 2.5 (Wang et al. 2012). For SIM, an arbitrary LOD threshold of 2.5 was used to determine significance, while for CIM, significance levels for LOD thresholds were determined using 1000 permutations.

Development of a Diagnostic Marker for Powdery Mildew Resistance

Primers for amplification of the *PsMLO1* genomic sequence (5′-ATGGCTGAAGAGGGAGTT-3′ and 5′- GGTAGCAG CTTGATTTGTGGATA -3′) were designed using Sequencher 4.7 (Gene Codes Corporation, USA) and OligoCalc: Oligonucleotide Properties Calculator (http://www.basic.northwestern.edu/biotools/oligocalc.html); on the basis of the published sequence (Santo et al. 2013: www.ncbi.nlm.nih.gov/nuccore/KC466597.1). PCR amplification was performed in a 20 µl reaction containing 20 ng DNA, 1× PCR buffer (Bioline), 0.4 µM of each primer, 0.2 mM of each dNTP and 0.5 U IMMOLASE (Bioline). PCR conditions included a hot start at 95 °C for 10 min, followed by 35 cycles of 94 °C for 30 s, 46–50 °C for 30 s and 72 °C for 30 s, and a final elongation step at 72 °C for 10 min.

PCR products were purified in a 15-µl reaction containing 2.5 U of shrimp alkaline phosphatase (USB-VWR International, Pennsylvania, USA) and 2.5 U of exonuclease I (at 20 U/µl) (New England Biolabs) and 5 µl of PCR product. Purified PCR products were analysed using a sequencing primer and BigDye Terminator v3.1 sequencing chemistry following the manufacturer's instructions. Final PCR products were purified using ethanol precipitation and resuspended in 12 µl Hi–Di formamide for sequence determination using an ABI3730xl (Applied Biosystems) capillary electrophoresis platform according to the manufacturer's instructions. Sequence analysis and assembly of the resulting electropherograms were performed in Sequencher 4.7 (Gene Codes Corporation, USA), allowing alignment and visual identification of sequence variants. Full-length sequence was generated using the Sanger sequencing-primer walking approach.

A specific PCR test, using the amplification primer pairs—5′-TGGTTCAATCGTCCTCACCT -3′ and 5′-TGCAAGTT GAAAGGCATTCT -3′—was designed to detect the presence of amplicon length variation due to insertion–deletion (indel) events, and the corresponding SCAR allele variation was determined in the mapping populations using PCR conditions as described for SSR analysis by Kaur et al. (2012).

Results

Polymorphic Markers for Map Construction

A total of 242 publicly available SSR markers (96 EST-SSRs and 146 genomic DNA-derived SSRs) were screened for polymorphism detection. A total of 41 (28 %) and 45 (31 %) of the genomic DNA-derived SSRs detected polymorphisms in the Kaspa×Yarrum and Kaspa×ps1771 mapping populations, respectively. Lower numbers of the EST-SSR markers detected polymorphism in the Kaspa×Yarrum (27 %) and Kaspa×ps1771 (26 %) populations (Table 1). All marker data were tested for conformity to the expected Mendelian ratio, and markers with χ^2 score >10 ($P<0.05$) were removed from further analysis. Residual sets of 41 (Kaspa×Yarrum) and 51 (Kaspa×ps1771) segregating marker alleles were used for linkage mapping.

Sub-sets of 424 and 422 SNPs from the total of 768 detected polymorphism in the Kaspa×Yarrum and Kaspa×ps1771 RILs, respectively (Table 1). A total of 314 polymorphic loci were found to be common between the two mapping populations. For each SNP, three main clusters were identified, corresponding to AA homozygotes, AB heterozygotes and BB homozygotes. The majority of the SNP markers produced two major clusters representing the homozygous genotypes, with a minor extra cluster corresponding to the heterozygous class. As both mapping populations were descended to the F_6 level, the frequency of heterozygous combinations was expected to be low, as was observed in practice (about 5 % in both populations). The χ^2 test ($P<0.05$) identified 5 % (Kaspa×Yarrum) and 4.2 % (Kaspa×ps1771) of the SNP markers that did not segregate in accordance with the expected Mendelian inheritance ratio. All markers exhibiting such significant segregation distortion were excluded from the final analysis.

Genetic Linkage Mapping

For the Kaspa×Yarrum mapping population, χ^2 analysis and missing data led to the exclusion of 48 markers (8.27 %) from further analysis. Of the remaining 443 loci (41 SSRs and 402 SNPs) (Table 1), 428 (35 SSRs and 393 SNPs) were assigned to 13 LGs (Additional file 2a). Five RILs were excluded from analysis due to marker heterozygosity levels between 25 and 35 %. For Kaspa×ps1771, 37 (6.6 %) markers were excluded from further analysis due to χ^2 values ($P<0.05$) and missing data. Of the remaining 455 loci (51 SSRs and 404 SNPs) (Table 1), 451 (50 SSRs and 401 SNPs) were assigned to 9 LGs (Additional file 2b). Comparisons between the two genetic maps revealed substantial commonality of marker order, although specific map distances were not always in similar proportion (Additional file 3a–3g).

The proportion of loci assigned to LGs was 97 and 99 % for the Kaspa×Yarrum and Kaspa×ps1771 maps, respectively, while the remaining markers were unlinked. In total, 11 LGs and 2 satellite were generated for Kaspa×Yarrum, and 7 LGs and 2 satellites were obtained for Kaspa×ps1771 (Additional file 4a and 4b). The cumulative length of the Kaspa×Yarrum map was 1910 cM, with an average distance of 4.4 cM between loci, while the Kaspa×ps1771 map spanned a total length of 1545 cM, with an average marker density of 1 locus per 3.4 cM (Table 2). As high levels of co-linearity were

Table 1 Total number of markers analysed, tested for polymorphism and assigned to genetic linkage map locations

Marker type	Total number of markers	Polymorphic markers in Kaspa×Yarrum	Mapped markers in Kaspa×Yarrum	Polymorphic markers in Kaspa×ps1771	Mapped markers in Kaspa×ps1771
Genomic DNA-derived SSR	146	41	18	45	32
EST-SSR	96	26	17	25	18
SNP	768	424	393	422	401
Total markers	1010	491	428	492	451

observed with previously published maps, most of the satellite LGs, when identity was predictable on the basis of common loci, could be reasonably expected to coalesce with the corresponding intact LG through the use of a larger number of markers (data not shown). In total, 308 markers were common between the Kaspa×Yarrum and Kaspa×ps1771 maps. Instances of markers assigned to different LGs between the two maps were rare, with the exception of Kaspa×Yarrum, for which 10 SNP markers (SNP_100000353, SNP_100000801, SNP_100000802, SNP_100000347, SNP_100000674, SNP_100000220, SNP_100000293, SNP_100000150, SNP_100000224, SNP_100000183) were located in a segment of Ps IV, while for Kaspa×ps1771, the corresponding region was on Ps III.

Consensus Linkage Map Construction

Data from three F_6 RIL populations was used to construct the preliminary composite map. The number of individual marker loci ranged from 429 (Kaspa×Yarrum) to 452 (Kaspa×ps1771) and 458 (Kaspa×Parafield). A total of 764 marker loci (680 SNPs and 84 SSRs) were assembled into 7 LGs (Additional file 5), of which 286 markers (37.4 %) were unique to single populations (Kaspa×Parafield, 157; Kaspa×Yarrum, 81; Kaspa×ps1771. 48), the remaining 478 (62.6 %) providing bridging loci between two or more maps. In total, 160 markers were common across all three maps. The highest number of common markers (308) was between the Kaspa×Yarrum and Kaspa×ps1771 maps, followed by the Kaspa×Parafield–Kaspa×ps1771 comparison with 261 markers. The total length of the preliminary composite linkage map was 2555 cM, lengths of major LGs ranging from 249 cM (Ps V) to 421 cM (Ps III), with an average density of one marker per 3.4 cM. The largest numbers of markers (135) were assigned to Ps VII, while the lowest numbers (83) were on Ps V. The marker order was largely co-linear with the three individual maps, although a few local inversions and marker rearrangements over short intervals were observed. A total of 33 markers (30 SNPs and 3 SSRs) were assigned to single loci on different LGs across mapping populations. Such

Table 2 Marker distribution over the LGs of the Kaspa×Yarrum and Kaspa×ps1771 map

Predicted pea chromosome	Map length (cM)		Number of mapped markers		Average marker density	
	Kaspa×Yarrum	Kaspa×ps1771	Kaspa×Yarrum	Kaspa×ps1771	Kaspa×Yarrum	Kaspa×ps1771
Ps I	278	189	44	47	6.3	4.0
Ps I.2	24		15		1.6	
Ps II	147	143	31	46	4.7	3.1
Ps II.2		16		12		1.3
Ps III	180	236	48	84	3.8	2.8
Ps III.2	8		3		2.7	
Ps III.3	161		30		5.4	
Ps IV	100	285	22	78	4.5	3.7
Ps IV.2	222		42		5.3	
Ps IV.3	77		24		3.2	
Ps V	184	212	47	51	3.9	4.2
Ps VI	197	178	50	51	3.9	3.4
Ps VI.2	4		4		1.0	
Ps VII	328	273	69	81	4.8	3.4
Ps VII.2		13				6.3
Total	1910	1545	429	452	4.4	3.4

loci were not considered as identical and were consequently reckoned as unique in the preliminary composite genetic map.

Based on BLASTN analysis of corresponding DNA sequences to the 768 SNPs, 767 detected significant sequence similarity matches to recently available pea transcriptome sequences (Additional file 6). This analysis supported establishment of 135 bridging loci between the field pea preliminary composite map and the available integrated map of Duarte et al. (2014). Comparison of these maps revealed the highest number of matches (29) between Ps VII and LG 7 and the lowest (7) between Ps V and LG 5. Minor discrepancies occurred in some markers (especially terminal locations), but marker order was generally co-linear. However, the global orientation was reversed for four of the seven LGs between the composite and integrated maps.

Merger of the two datasets obtained a new consensus map containing 2028 markers on 7 LGs spanning 2387 cM (Fig. 1; Table 3). A total of 535 SNP markers from the preliminary composite map were derived from transcript sequences common with the SNP markers from Duarte et al. (2014): Ps III and Ps VII containing the highest number (102) and Ps V the lowest (56) (Additional file 7). The marker order of the final consensus map was largely co-linear between the participating structures, although several inversions and local rearrangements were observed.

Phenotypic Analysis and QTL Detection

For B toxicity, in the Kaspa×ps1771 RIL population, a high degree of correlation was observed for symptom rating obtained at different time intervals (r^2=0.86). Averages for plant symptom score (calculated from individual plant assessments), and boron index were used to generate frequency distribution histograms. Although the distribution pattern was not indicative of continuous variation, it did not conform to a bimodal structure arising from a single gene effect (Additional file 8a). Narrow-sense heritability values for each measurement ranged from 0.83 to 0.85. Marker regression analysis identified two markers (AB71 and PBA_PS_0398) that were significantly associated with variation for the trait. SIM identified one genomic region on Ps VI, accounting for 58 % of the phenotypic variance (V_p) (Additional file 2b).

In the case of powdery mildew resistance, significant differences between RILs for each population were observed for both symptom score and percentage of leaf cover. Frequency distribution patterns obtained from both populations, due to bimodal structures, indicated the presence of a single gene responsible for powdery mildew resistance (Additional file 8b). The locations and magnitudes of effect for each genomic region were estimated, and for each mapping population, both marker regression analysis and SIM detected a single genomic region on Ps VI (Table 4). CIM identified the same region, accounting for 93 and 81 % of V_p for the Kaspa×Yarrum and Kaspa×ps1771 populations, respectively (Additional file 2a and 2b).

The same genomic region on the Kaspa×ps1771 map contained major QTLs for both powdery mildew resistance and boron tolerance region, indicating that the respective causal polymorphisms are closely linked in coupling phase (Fig. 2).

Development of a Diagnostic Marker for the er1 Powdery Mildew Resistance Gene

The full-length *PsMLO1* coding sequence was used a template for primer design to support amplicon generation from the Kaspa, Yarrum and ps1771 genotypes. The DNA sequence of the *PsMLO1*-specific amplicon was determined (Additional file 9), and sequence analysis revealed that the Yarrum and ps1771-derived *PsMLO1* allele containing a 2-bp insertion event in intron 11 of the gene as compared to the equivalent from Kaspa, the powdery mildew susceptible parental genotype (Fig. 3). A SCAR marker (Fig. 3) was designed to identify the allelic difference and was screened across both RIL populations. The *PsMLO1* SCAR locus was confirmed as coinciding with the major powdery mildew resistance QTLs that were identified in both populations.

Discussion

Attributes of Genetic Linkage Maps and Consensus Map Construction

During a lengthy history of genetic mapping studies in field pea, different types of genetic marker systems have been successively used (Weeden et al. 1996; Weeden and Boone 1999; Laucou et al. 1998; Timmerman-Vaughan et al. 2000; Ellis and Poyser 2002; Loridon et al. 2005; Aubert et al. 2006; Deulvot et al. 2010). However, to date, only one high-density genetic linkage map has been constructed predominantly through the use of gene-associated SSR and SNP markers (Leonforte et al. 2013). In the present study, the same marker sets have been used to generate another two maps, both of which displayed a relatively uniform marker distribution, with average marker densities of one per 4.4 cM (Kaspa×Yarrum) and 3.4 cM (Kaspa×ps1771). The former map was significantly longer (1910 cM) than the latter (1545 cM). The reason for excess map length could be due to recombination events or missing data in the Kaspa×Yarrum RIL population (Knox and Ellis 2002). Based on comparison through common markers, most LGs were consistent between the two individual maps, with a number of minor exceptions.

A preliminary composite map was constructed in this study from the three trait-specific RIL populations, which was then

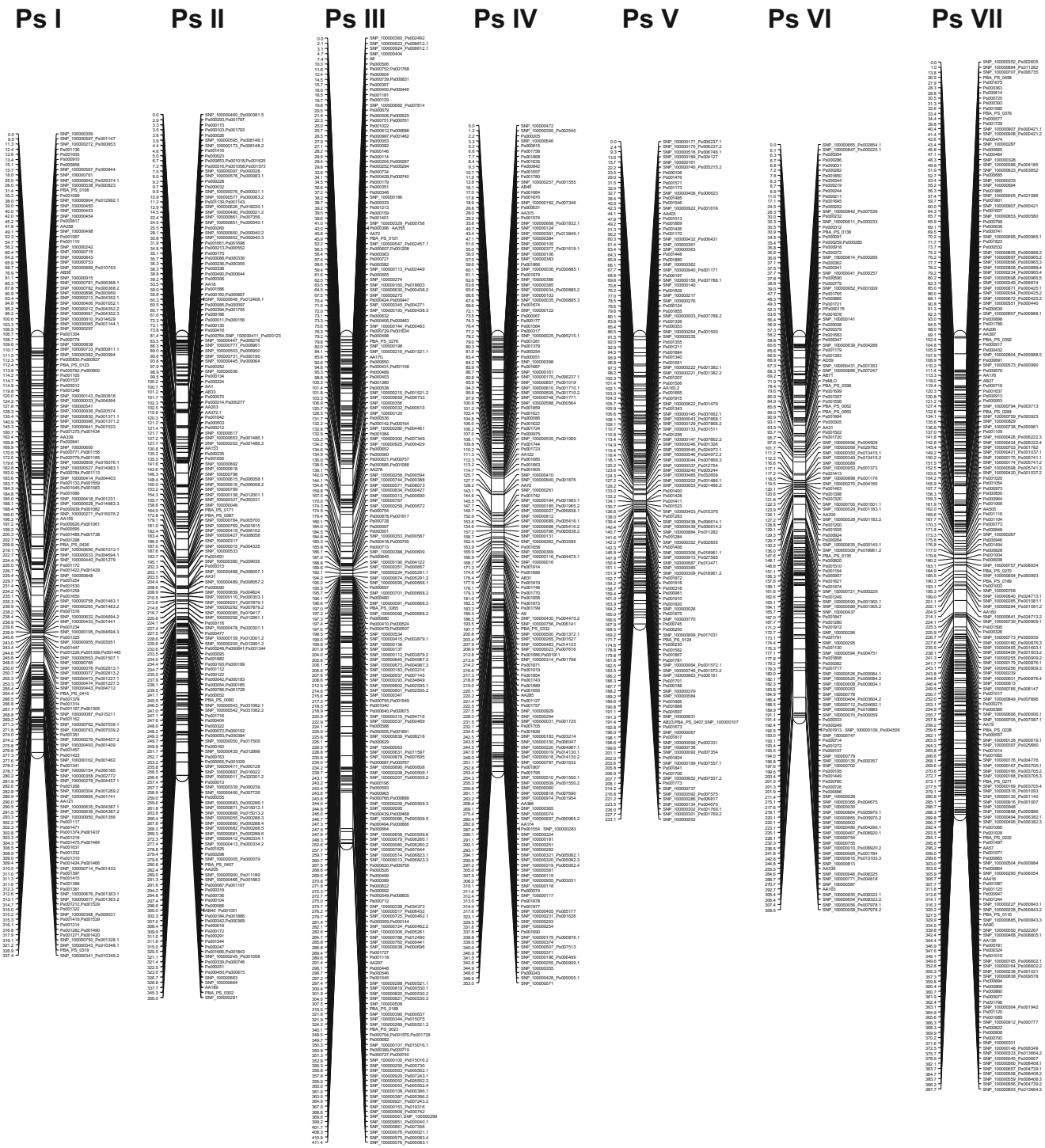

Fig. 1 Consensus map of field pea, with marker loci shown on the *right-hand side* of LGs, and distances between markers indicated in cM on the *left*. For presentation purposes, not all genetic markers are shown on the map. Specific details of marker locus identity and location are provided in Additional File 7

combined with a previously generated integrated map to obtain a final consensus. The MergeMap software that was used for this purpose has previously been used for construction of consensus maps in crops such as *Phaseolus vulgaris* (Galeano et al. 2012), *Arachis hypogaea* L. (Gautami et al. 2012) and *Vicia faba* L. (Satovic et al. 2013) based on 3, 11 and 3 populations, respectively. The preliminary composite map provided the opportunity to assign a larger number of loci than in individual maps, providing increased reliability of marker location prediction. Comparison to individual component maps revealed only minor inconsistencies of marker order, generally within the same LGs. Such discrepancies may be due to mapping errors due to missing data, limited linkage in individual maps, chromosome rearrangements or

Table 3 Characteristics of consensus map of field pea

Predicted pea chromosome	Number of mapped markers	Map length (cM)	Average marker density
Ps I	250	337	1.3
Ps II	237	356	1.5
Ps III	330	411	1.2
Ps IV	299	353	1.2
Ps V	276	233	0.8
Ps VI	242	309	1.3
Ps VII	394	388	1.0
Total	2028	2387	1.2

the influence of paralogous sequences. In general, composite maps provide one of many possible non-conflicting linear representations of the consensus DAGs (Close et al. 2009). Consequently, the order of markers in an integrated structure map may not match the order of corresponding nucleotides in a genome sequence, but marker order over longer distances should generally be preserved.

Developments in high-throughput genotyping based on SNPs facilitated comparison between the preliminary composite and integrated maps of pea. In general, small rearrangements were observed when comparing the two maps, especially near the distal ends of LGs. The observed differences in map length and marker order could be attributable to the use of different software to generate each map (Muñoz-Amatriaín et al. 2011). Comparison between the transcriptome datasets underlying each sequence-based marker led to the identification of common mRNA-encoding sequences. The newly generated consensus map for pea features integration of a large number of coding regions, providing an effective framework for downstream analyses, including comparisons between the locations of major genes for important traits or QTL positions between populations from different crosses. A high-density genetic map based on gene-based markers also provides an important foundation for QTL mapping and for anchoring of sequence scaffolds. Linkage maps have previously been used for anchoring and orientation of scaffolds in whole genome sequencing projects for many crop species, including soybean (Hyten et al. 2010), watermelon (Ren et al. 2012), grape (Jaillon et al. 2007) and cucumber (Huang et al. 2009). The linkage maps described here would hence be highly useful for the future genome assembly expected from the field pea genome sequencing consortium (http://www.coolseasonfoodlegume.org/pea_genome).

Phenotypic Analysis and Identification of QTLs for B Tolerance and Powdery Mildew Resistance

Physiological mechanisms for B tolerance are relatively well understood for some plant species such as *Arabidopsis thaliana*, and for cereal crops like wheat and barley. One major mechanism is the elimination of excess B from the root system based on transporter activity. Several different types of B transporter have been characterised in *A. thaliana* (Takano et al. 2002), and a corresponding gene has been cloned and characterised from barley (Hayes and Reid 2004; Schnurbusch et al. 2010). However, no equivalent studies have so far been conducted to determine the genetic basis of B tolerance in field pea. The frequency distributions for boron index and mean symptom score were not consistent with the contributions of a large number of genetic loci (as they did not conform to normal distributions), but equally, were not clearly attributable to single gene effects. Nonetheless, the present study identified a single genomic region of major effect on Ps VI. This finding is consistent with the outcomes of similar studies of other legume species such as lentil (Kaur et al. 2014) and *M. truncatula* (Bogacki et al. 2013), which also reported single gene models. It is possible that one or more

Table 4 Identification of QTLs for powdery mildew resistance and boron tolerance on the Kaspa×Yarrum and Kaspa×ps1771 genetic map based on CIM

Population	Trait	Flanking markers	Position	LOD threshold	Max LOD score	% V_p
Kaspa×Yarrum	Powdery mildew resistance—plant symptom	*PsMLO1* PBA_PS_0065	183–188	3.9	87	93
	Powdery mildew resistance—leaf area	*PsMLO1* PBA_PS_0065	183–188	4.3	64	89
Kaspa×ps1771	Powdery mildew resistance—plant symptom	AB71 *PsMLO1*	86–89	3.2	44	81
	Boron toxicity—symptom score	AB71 *PsMLO1*	86–89	3.4	26	58
	Boron toxicity—index	AB71 *PsMLO1*	86–89	3.1	27	56

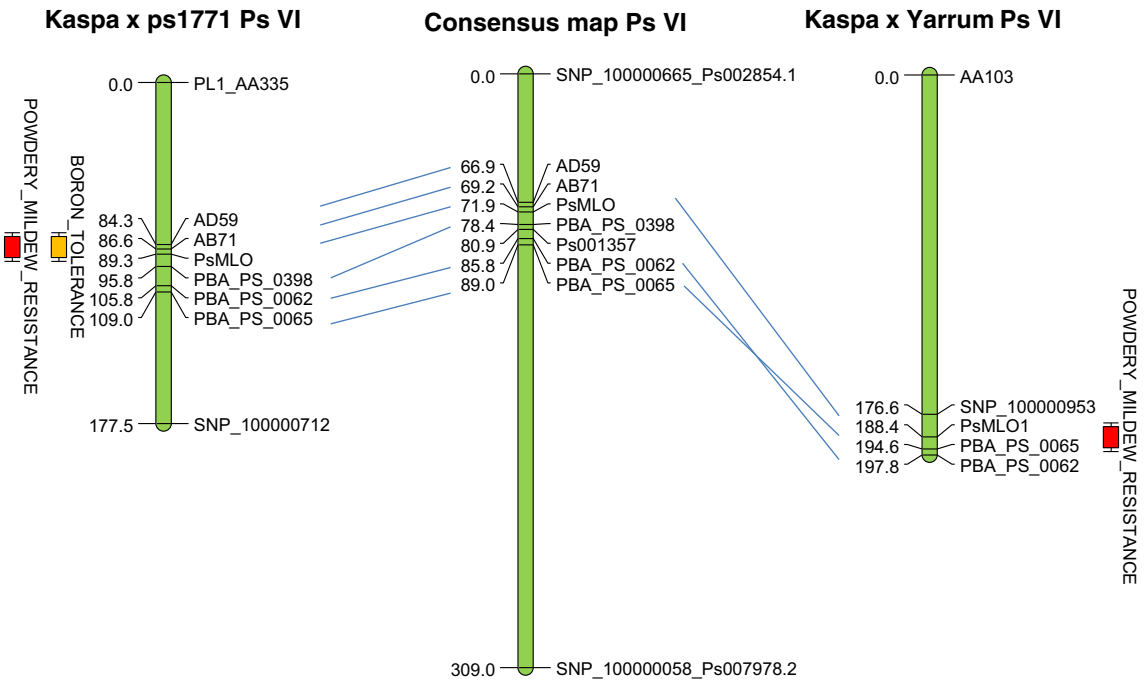

Fig. 2 Co-linearity of common markers between individual genetic linkage maps (Kaspa×ps1771 and Kaspa×Yarrum) and consensus map on Ps VI at the QTL-containing regions of the field pea. LGs or chromosomes are shaded in colours for presentation purposes. Marker loci are shown on the right of the linkage groups, and map distances between markers are indicated in cM on the left. *Coloured lines* represent the corresponding positions of common markers

minor gene also contributes to B tolerance in the Kaspa× ps1771 mapping population, but was below the threshold level for detection in the present study.

Resistance to powdery mildew in field pea is also well understood, the recessive *er1* gene being the most common cause (Timmerman et al. 1994; Vaid and Tyagi 1997; Janila and Sharma 2004. The present study obtained evidence for single gene resistance in both populations, based on both the bimodal nature of the frequency distribution data, and QTL analysis. The genomic location on Ps VI, in the vicinity of *er1*, was consistent with the outcomes of prior studies (Ek et al. 2005; Janila and Sharma 2004; Timmerman et al. 1994; Tiwari et al. 1998; Katoch et al. 2010). In a previous study, two SSR loci (AB71 and AD59) were found to be located on either side of *er1*, 4.6 and 4.3 cM distant, respectively (Loridon et al. 2005). In the present study, these loci were also identified in the QTL-containing interval. The results of the present study were not consistent with presence of *er2* in

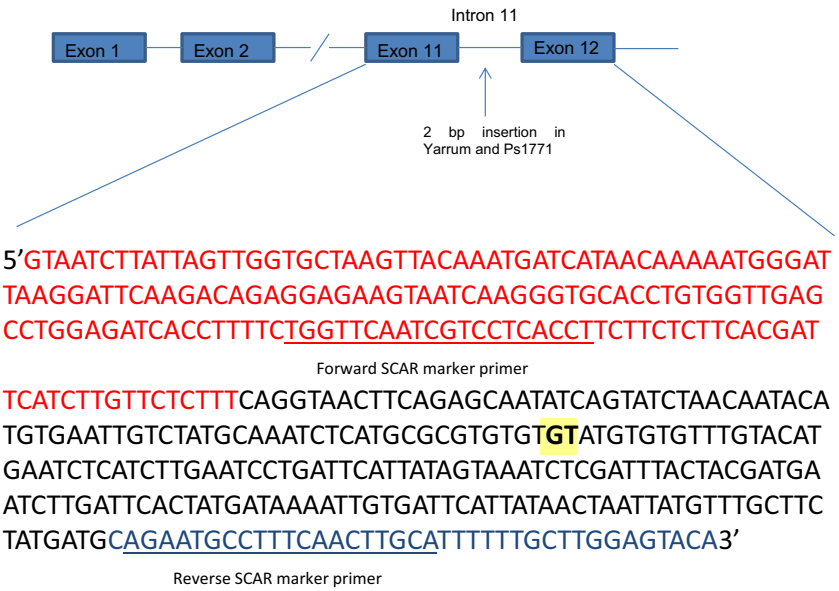

Fig. 3 Schematic diagram of intron-exon structure in the *PsMLO1* gene in the region surrounding the 2-bp indel polymorphism. The sequence of exon 11 is shown in *red text*, sequence of intron 11 shown in *black text* and exon 12 in *blue*, and the 2-bp insertion characteristic of Yarrum and Ps1771 is *highlighted*. The SCAR primer sequences and location are *underlined*

the resistant parental genotypes, as no QTL effects were identified on pea LG III, to which this locus has been previously attributed (Katoch et al. 2010).

A major implication arising from parallel assessment of B toxicity tolerance and powdery mildew resistance is that both traits in Kaspa×ps1771 population are controlled by single QTLs of large magnitude within the same interval. Although there is no reason to suspect a causal association, the respective genes are presumably in close linkage, consistent with observations from local field pea breeding trials (A. Leonforte, unpublished data). The ps1771 genotype can hence function as a common donor for the two traits. Fortunately, both the linked traits are favourable in nature, as otherwise recombination events within the QTL-containing region would be required to separate the determinants.

Development of a Diagnostic Marker for Powdery Mildew Resistance

Recent studies based on sequence analysis in four field pea accessions have reported that *PsMLO1* provides the functional basis for allelic variation of the field pea *er1* gene (Humphry et al. 2011). Loss-of-function of *PsMLO1* is further known to provide powdery mildew resistance. Following mutagenesis of the resistant genotypes Solara and Frilene with ethylnitrosourea (ENU), susceptible derivatives were shown to contain point mutations in coding sequences leading to drastic truncation of the *PsMLO1* gene product (Santo et al. 2013). In the present study, a small indel was identified that differentiates the *PsMLO1* alleles from resistant and susceptible genotypes. Conversion of the indel polymorphism into a SCAR marker demonstrated coincidence with the powdery mildew resistance QTLs reported in this study. Based on previous research, it may be reasonably concluded that pathogen resistance in Yarrum and ps1771 is due to the *er*1 gene and that the SCAR provides a diagnostic molecular marker for trait variation in this germplasm. However, the origin of functional variation in *PsMLO1* is not so clear, as, in contrast, to the outcomes of mutagenesis studies, the susceptibility-associated allele in Kaspa is not obviously impaired in terms of translation product structure, given that the 2-bp indel is located within an intron. No obvious exonic changes were identified across the full-length gene (Additional file 9). Functional variation for powdery mildew resistance must hence be due either to effects of the intron-located change on transcript splicing, or possibly the presence of a regulatory mutation outside the coding sequence, presumably in the 5′-proximal region of the gene, in linkage disequilibrum with the observed polymorphism. Further sequencing studies are required to test these hypotheses.

Implications for Field Pea Breeding Programs

Molecular markers linked to important agronomic traits have been demonstrated to be highly applicable to selection for desirable gene variants in different breeding programs. Marker-assisted selection (MAS) is cost and time efficient, non-destructive in nature and less error-prone than phenotypic selection. The major benefit of MAS for B toxicity tolerance and powdery mildew resistance in field pea would be to co-select genes for tolerance in multiple different genetic backgrounds. The Ps VI-located QTLs account for large percentages of V_p, providing the capacity for introgression into elite parental background by donor-recipient backcrossing with minimal linkage drag. In addition, the development of a diagnostic marker for powdery mildew resistance will further facilitate selection processes in field pea breeding programs by direct identification of donor genotypes in germplasm collections and hence reduce duration of the breeding cycle.

Acknowledgements This work was supported by funding from the Victorian Department of Environment and Primary Industries and the Grains Research and Development Council, Australia.

References

Attanayake RN, Glawe DA, McPhee KE, Dugan FM, Chen W (2010) *Erysiphe trifolii*—a newly recognized powdery mildew pathogen of pea. Plant Pathol 59:712–720

Aubert G, Morin J, Jacquin F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Hénaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theor Appl Genet 112:1024–1041

Bagheri A, Paull JG, Rathjen AJ (1996) Genetics of tolerance to high concentrations of soil boron in peas (*Pisum sativum* L.). Euphytica 87:69–75

Bennett SJ (2012) Early growth of field peas under saline and boron toxic soils. In Proceedings of 16th Agronomy Conference 2012, University of New England, Armidale, NSW, 14–18th October 2012

Blenda A, Fang DD, Rami JF, Garsmeur O, Luo F, Lacape JM (2012) A high density consensus genetic map of tetraploid cotton that integrates multiple component maps through molecular marker redundancy check. PLoS ONE 7:e45739

Bogacki P, Peck DM, Nair RM, Howie J, Oldach KH (2013) Genetic analysis of tolerance to boron toxicity in the legume *Medicago truncatula*. BMC Plant Biol 13:54

Cartwright B, Zarcinas BA, Mayfield AH (1984) Toxic concentrations of boron in a red-brown earth at Gladstone, South Australia. Soil Res 22:261–272

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004) Estimating genome conservation between crop and model legume species. Proc Natl Acad Sci U S A 101:15289–15294

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szűcs P, Sato K, Hayes PM, Matthews

DE, Kleinhofs A, Muehlbauer GJ, Young JD, Marshal DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genomics 10:582

Cloutier S, Ragupathy R, Miranda E, Radovanovic N, Reimer E, Walichnowski A, Ward K, Rowland G, Duguid S, Banik M (2012) Integrated consensus genetic and physical maps of flax (*Linum usitatissimum* L.). Theor Appl Genet 125:1783–1795

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Henaut I, Burstin J, Aubert G (2010) Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics 11:48

Dirlewanger E, Isaac PG, Ranade S, Belajouza M, Cousin R, de Vienne D (1994) Restriction fragment length polymorphism analysis of loci associated with disease resistance genes and developmental traits in *Pisum sativum* L. Theor Appl Genet 88:17–27

Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, Lavaud C, Lejeune-Hénaut I, Martinant J-P, Pichon J-P, Pilet-Nayel M-L, Boutet G (2014) Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. BMC Genomics 15:126

Ek M, Eklund M, von Post R, Dayteg C, Henriksson T, Weibull P, Ceptilis A, Issac P, Tuvesson S (2005) Microsatellite markers for powdery mildew resistance in pea (*Pisum sativum* L.). Hereditas 142:86–91

Ellis THN, Poyser SJ (2002) An integrated and comparative view of pea genetic and cytogenetic maps. New Phytol 153:17–25

Fondevilla S, Carver TLW, Moreno MT, Rubiales D (2006) Macroscopic and histological characterisation of genes *er1* and *er2* for powdery mildew resistance in pea. Eur J Plant Pathol 115:309–321

Fondevilla S, Rubiales D, Moreno MT, Torres AM (2008) Identification and validation of RAPD and SCAR markers linked to the gene *Er3* conferring resistance to *Erysiphe pisi* DC in pea. Mol Breed 22:193–200

Fondevilla S, Cubero JI, Rubiales D (2011) Confirmation that the Er3 gene, conferring resistance to *Erysiphe pisi* in pea, is a different gene from *er1* and *er2* genes. Plant Breed 130:281–282

Fondevilla S, Chattopadhyay C, Kahre N, Rubiales D (2013) *Erysiphe trifolii* is able to overcome *er1* and *Er3*, but not *er2* resistance genes in pea. Eur J Plant Pathol 136:557–563

Galeano CH, Cortés AJ, Fernandez AC, Soler A, Franco-Herrera N et al (2012) Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. BMC Genet 13:48

Gautami B, Foncéka D, Pandey MK, Moretzsohn MC, Sujay V et al (2012) An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.). PLoS ONE 7:e41213

GenStat Committee (2002) The Guide to GenStat Release 6.1—Part 2: Statistics. VSN International, Oxford

Harland SC (1948) Inheritance of immunity to mildew in Peruvian forms of *Pisum sativum*. Heredity 2:263–269

Hayes JE, Reid RJ (2004) Boron tolerance in barley is mediated by efflux of boron from the roots. Plant Physiol 136:3376–3382

Heringa RJ, Van Nore A, Tazelaar MF (1969) Resistance to powdery mildew (*Erysiphe polygoni* D.C.) in peas (*Pisum sativum* L.). Euphytica 18:163–169

Huang S, Li R, Zhang Z, Li L, Gu X, Fan W, Lucas WJ, Wang X, Xie B, Ni P, Ren Y, Zhu H, Li J, Lin K, Jin W, Fei Z, Li G, Staub J, Kilian A, van der Vossen EAG, Wu Y, Guo J, He J, Jia Z, Ren Y, Tian G, Lu Y, Ruan J, Qian W, Wang M, Huang Q, Li B, Xuan Z, Cao J, Asan WZ, Zhang J, Cai Q, Bai Y, Zhao B, Han Y, Li Y, Li X, Wang S, Shi Q, Liu S, Cho WK, Kim J-Y, Xu Y, Heller-Uszynska K, Miao H, Cheng Z, Zhang S, Wu J, Yang Y, Kang H, Li M, Liang H, Ren X, Shi Z, Wen M, Jian M, Yang H, Zhang G, Yang Z, Chen R, Liu S, Li J, Ma L, Liu H, Zhou Y, Zhao J, Fang X, Li G, Fang L, Li Y, Liu D, Zheng H, Zhang Y, Qin N, Li Z, Yang G, Yang S, Bolund L, Kristiansen K, Zheng H, Li S, Zhang X, Yang H, Wang J, Sun R, Zhang B, Jiang S, Wang J, Du Y, Li S (2009) The genome of the cucumber, *Cucumis sativus* L. Nat Genet 41:1275–1281

Humphry M, Reinstädler A, Ivanov S, Bisseling T, Panstruga R (2011) Durable broad-spectrum powdery mildew resistance in pea *er1* plants is conferred by natural loss-of-function mutations in *PsMLO1*. Mol Plant Pathol 12:866–878

Hyten DL, Cannon SB, Song Q, Weeks N, Fickus EW, Shoemaker RC, Specht JE, Farmer AD, May GD, Cregan PB (2010) High-throughput SNP discovery through deep resequencing of a reduced representation library to anchor and orient scaffolds in the soybean whole genome sequence. BMC Genomics 11:38

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyere C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pe ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quetier F, Wincker P, French-Italian P (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Janila P, Sharma B (2004) RAPD and SCAR markers for powdery mildew resistance gene *er* in pea. Plant Breed 123:271–274

Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Mol Gen Genomics 272:235–246

Katoch V, Sharma S, Pathania S, Banayal DK, Sharma SK, Rathour R (2010) Molecular mapping of pea powdery mildew resistance gene *er2* to pea linkage group III. Mol Breed 25:229–237

Kaur S, Pembleton LW, Cogan NOI, Savin KW, Leonforte T, Paul J, Materne M, Forster JW (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics 13:104

Kaur S, Cogan NOI, Stephens A, Noy D, Butsch M, Forster JW, Materne M (2014) EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance. Theor Appl Genet 127:703–713

Knox MR, Ellis THN (2002) Map expansion in pea. Genetics 162:861–873

Kosambi DD (1944) The estimation of map distance from recombination values. Ann Eugen 12:172–175

Kumar H, Singh RB (1981) Genetic analysis of adult plant resistance to powdery mildew in pea (*Pisum sativum* L.). Euphytica 30:147–154

Lacey A, Davies S (2009) Boron toxicity in WA soils. Farm note 388, Department of Agriculture and food, Western Australia

Laucou V, Haurogne K, Ellis N, Rameau C (1998) Genetic mapping in pea: I. RAPD-based linkage map of *Pisum sativum*. Theor Appl Genet 97:905–915

Leonforte A, Sudheesh S, Cogan NOI, Salisbury PA, Nicolas ME, Materne M, Forster JW, Kaur S (2013) SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). BMC Plant Biol 13:161

Loridon K, Mcphee K, Morin J (2005) Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L). Theor Appl Genet 111:1022–1031

Manly KF, Cudmore RH Jr, Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. Mamm Genome 12:930–932

Marx GA (1986) Location of *er* proving elusive. Pisum Newsl 18:39–41

Milczarski P, Bolibok-Bragoszewska H, Myśków B, Stojałowski S, Heller-Uszyńska K, Góralska M, Brągoszewski P, Uszyński G,

Kilian A, Rakoczy-Trojanowska M (2011) A high density consensus map of rye (*Secale cereale* L.) based on DArT markers. PLoS ONE 7:e28495

Millan T, Winter P, Jüngling R, Gil J, Rubio J, Cho S, Cobos MJ, Iruela M, Rajesh PN, Tekeoglu M, Kahl G, Muehlbauer FJ (2010) A consensus genetic map of chickpea (*Cicer arietinum* L.) based on 10 mapping populations. Euphytica 175:175–189

Munjal RL, Chenulu VV, Hora TS (1963) Assessment of losses due to powdery mildew (*Erysiphe polygoni*) on pea. Indian Phytopathol 19:260–267

Muñoz-Amatriaín M, Moscou MJ, Bhat PR, Svensson JT, Bartoš J, Suchánková P, Šimková H, Endo TR, Fenton RD, Lonardi S, Castillo AM, Chao S, Cistué L, Cuesta-Marcos A, Forrest KL, Hayden MJ, Hayes PM, Horsley RD, Makoto K, Moody D, Sato K, Vallés MP, Wulff BBH, Muehlbauer GJ, Doležel J, Close TJ (2011) An improved consensus linkage map of barley based on flow-sorted chromosomes and single nucleotide polymorphism markers. Plant Genome 4:238–249

Nuttall JG, Armstrong RD, Connor DJ, Matassa VJ (2003) Interrelationships between edaphic factors potentially limiting cereal growth on alkaline soils in north-western Victoria. Soil Res 41:277–292

Ondřej M, Dostálová R, Odstrčilová L (2005) Response of *Pisum sativum* germplasm resistant to *Erysiphe pisi* to inoculation with *Erysiphe baeumleri*, a new pathogen of peas. Plant Prot Sci 41:95–103

Pavan S, Jacobsen E, Visser RGF, Bai Y (2010) Loss of susceptibility as a novel breeding strategy for durable and broad-spectrum resistance. Mol Breed 25:1–12

Pavan S, Schiavull A, Appiano M, Marcotrigiano AR, Cillo F, Visser RGF, Bai Y, Lotti C, Ricciardi L (2011) Pea powdery mildew *er1* resistance is associated to loss-of-function mutations at a MLO homologous locus. Theor Appl Genet 123: 1425–1431

Pavan S, Schiavulli A, Appiano M, Miacola C, Visser RG, Bai Y, Lotti C, Ricciardi L (2013) Identification of a complete set of functional markers for the selection of *er1* powdery mildew resistance in *Pisum sativum* L. Mol Breed 31:247–253

Ren Y, Zhao H, Kou Q, Jiang J, Guo S, Zhang H, Hou W, Zou X, Sun H, Gong G, Levi A, Xu Y (2012) A high resolution genetic map anchoring scaffolds of the sequenced watermelon genome. PLoS ONE 7:e29453

Rubiales D, Fernandez-Aparicio M, Moral A, Barilli E, Sillero JC, Fondevilla S (2009) Disease resistance in pea (*Pisum sativum* L.) types for autumn sowings in Mediterranean environments. Czech J Genet Plant Breed 45:135–142

Santo T, Rashkova M, Alabaça C, Leitão J (2013) The ENU-induced powdery mildew resistant mutant pea (*Pisum sativum* L.) lines S (*er1mut1*) and F (*er1mut2*) harbour early stop codons in the *PsMLO1* gene. Mol Breed 32:723–727

Satovic Z, Avila CM, Cruz-Izquierdo S, Díaz-Ruíz R, García-Ruíz GM, Palomino C, Gutiérrez N, Vitale S, Ocaña-Moral S, Gutiérrez MV, Cubero JI, Torres AM (2013) A reference consensus genetic map for molecular markers and economically important traits in faba bean (Vicia faba L). BMC Genomics 14:932

Schnurbusch T, Hayes J, Hrmova M, Baumann U, Ramesh SA, Tyerman SD, Langridge P, Sutton T (2010) Boron toxicity tolerance in barley through reduced expression of the multifunctional aquaporin HvNIP2; 1. Plant Physiol 153:1706–1715

Schuelke M (2000) An economic method for the fluorescent labelling of PCR fragments. Nat Biotechnol 18:233–234

Shirasawa K, Bertioli DJ, Moretzsohn MC, Leal-Bertioli SCM, Varshney RK et al (2013) An integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of *Arachis* and divergences with other legume genomes. DNA Res 20: 173–184

Smýkal P, Aubert G, Burstin J, Coyne CJ, Ellis NTH, Flavell AJ, Ford R, Hýbl M, Macas J, Neumann P, McPhee KE, Redden RJ, Rubiales D, Weller JL, Warkentin TD (2012) Pea (*Pisum sativum* L.) in the genomic era. Agronomy 2:74–115

Sokhi SS, Jhooty JS, Bains SS (1979) Resistance in pea against powdery mildew. Indian Phytopathol 32:571–574

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join Map. Plant J 3:739–744

Takano J, Noguchi K, Yasumori M, Kobayashi M, Gajdos Z, Miwa K, Hayashi H, Yoneyama T, Fujiwara T (2002) *Arabidopsis* boron transporter for xylem loading. Nature 420:337–340

Timmerman GM, Frew TJ, Weeden NF, Miller AL, Goulden DS (1994) Linkage analysis of *er-1*, a recessive *Pisum sativum* gene for resistance to powdery mildew fungus (*Erysiphe pisi* D.C.). Theor Appl Genet 88:1050–1055

Timmerman-Vaughan G, Frew TJ, Weeden NF (2000) Characterization and linkage mapping of R-gene analogous DNA sequences in pea (*Pisum sativum* L.). Theor Appl Genet 101:241–247

Tiwari KR, Penner GA, Warkentin TD (1997) Inheritance of powdery mildew resistance in pea. Can J Plant Sci 77:307–310

Tiwari KR, Penner GA, Warkentin TD (1998) Identification of coupling and repulsion phase RAPD markers for powdery mildew resistance gene *er1* in pea. Genome 41:440–444

Tiwari KR, Penner GA, Warkentin TD (1999) Identification of AFLP markers for powdery mildew resistance gene *er2* in pea. Pisum Genet 31:27–29

Tonguç M, Weeden NF (2010) Identification and mapping of molecular markers linked to er1 gene in pea. J Plant Mol Biol Biotechnol 1:1–5

Vaid A, Tyagi P (1997) Genetics of powdery mildew resistance in pea. Euphytica 96:203–206

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:7–78

Wang J, Lydiate DJ, Parkin IA, Falentin C, Delourme R et al (2011) Integration of linkage maps for the Amphidiploid *Brassica napus* and comparative mapping with *Arabidopsis* and *Brassica rapa*. BMC Genomics 12:101

Wang S, Basten CJ, Zeng ZB (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. (http://statgen.ncsu.edu/qtlcart/WQTLCart.htm)

Warkentin TD, Rashid KY, Xue AG (1996) Fungicidal control of powdery mildew in field pea. Can J Plant Sci 76:933–935

Weeden NF, Boone WE (1999) Mapping the *Rb* locus on linkage group III using long PCR followed by endonuclease digestion. Pisum Genet 31:36

Weeden NF, Swiecicki WK, Timmerman-Vaughan GM, Ellis THN, Ambrose M (1996) The current pea linkage map. Pisum Genet 28: 1–4

Weeden NF, Ellis THN, Timmerman-Vaughan GM, Swiecicki WK, Rozov SM, Berdnikov VA (1998) A consensus linkage map for *Pisum sativum*. Pisum Genet 30:1–3

Wu Y, Close TJ, Lonardi S (2011) Accurate construction of consensus genetic maps via integer linear programming. IEEE/ACM Trans Comput Biol Bioinforma 8:381–394. doi:10.1109/TCBB.2010.35

Yap IV, Schneider D, Kleinberg J, Matthews D, Cartinhour S, McCouch SR (2003) A graph-theoretic approach to comparing and integrating genetic, physical and sequence-based maps. Genetics 165:2235–2247

Yau SK, Ryan J (2008) Boron toxicity tolerance in crops: a viable alternative to soil amelioration. Crop Sci 48:854–865

CHAPTER 4

De novo assembly and characterisation of the field pea transcriptome using RNA-Seq

4.1 Chapter preface

This chapter describes the generation of a comprehensive transcriptome data sets from two genotypes (Kaspa and Parafield) of field pea that differ in terms of seed and plant morphological characteristics through the use of RNA-Seq, followed by *de novo* assembly, comparison to gene complements in related species, sequence annotation and assessment of tissue-specific expression.

This chapter is presented in published format.

4.2 Publication details

Title: *De novo* assembly and characterisation of the field pea transcriptome using RNA-Seq

Journal Details: BMC Genomics. 2015, 16:611. doi:10.1186/s12864-015-1815-7.

Stage of publication: Published

Authors: Shimna Sudheesh, Timothy I Sawbridge, Noel OI Cogan, Peter Kennedy, John W Forster, Sukhjiwan Kaur

This paper has also been presented at the following conference, and appears in the conference proceedings as:

Sudheesh S, Lombardi M, Kennedy P, Cogan NOI, Forster JW, Kaur S (2014) Transcriptome atlas of field pea (*Pisum sativum* L.). In: *Proceedings of the 6th International Food Legume Research Conference and the 7th International Conference on Legume Genetics and Genomics, Saskatoon, Saskatchewan, Canada.* 7th-11th July 2014, P72.

4.3 Statement of contributions of joint authorship

Shimna Sudheesh prepared plant materials, performed RNA extraction, library preparation, sequencing, data analysis and drafted the manuscript.

Timothy I Sawbridge contributed to data interpretation and assisted in drafting the manuscript.

Noel OI Cogan and Sukhjiwan Kaur co-conceptualised the project, participated in experimental design, contributed to data interpretation and assisted in drafting the manuscript.

Peter Kennedy co-conceptualised the project.

John W Forster co-conceptualised the project and assisted in drafting the manuscript.

4.4 Statement from the co-author confirming the authorship contribution of the PhD candidate

"As co-author of the manuscript 'Sudheesh *et al.* (2015) *De novo* assembly and characterisation of the field pea transcriptome using RNA-Seq. BMC Genomics. 16:611', I confirm that Shimna Sudheesh has made the following contributions,

- Preparation of plant materials
- RNA extraction

- Library preparation and sequencing
- Data analysis and interpretation of findings
- Writing the manuscript, critical appraisal of content and response to reviewers”

Dr Sukhjiwan Kaur

Date: 15/09/2015

Sudheesh et al. BMC Genomics (2015) 16:611
DOI 10.1186/s12864-015-1815-7

BMC Genomics

RESEARCH ARTICLE Open Access

De novo assembly and characterisation of the field pea transcriptome using RNA-Seq

Shimna Sudheesh[1,3], Timothy I. Sawbridge[1,3], Noel OI Cogan[1], Peter Kennedy[2], John W. Forster[1,3*] and Sukhjiwan Kaur[1]

Abstract

Background: Field pea (*Pisum sativum* L.) is a cool-season grain legume that is cultivated world-wide for both human consumption and stock-feed purposes. Enhancement of genetic and genomic resources for field pea will permit improved understanding of the control of traits relevant to crop productivity and quality. Advances in second-generation sequencing and associated bioinformatics analysis now provide unprecedented opportunities for the development of such resources. The objective of this study was to perform transcriptome sequencing and characterisation from two genotypes of field pea that differ in terms of seed and plant morphological characteristics.

Results: Transcriptome sequencing was performed with RNA templates from multiple tissues of the field pea genotypes Kaspa and Parafield. Tissue samples were collected at various growth stages, and a total of 23 cDNA libraries were sequenced using Illumina high-throughput sequencing platforms. A total of 407 and 352 million paired-end reads from the Kaspa and Parafield transcriptomes, respectively were assembled into 129,282 and 149,272 contigs, which were filtered on the basis of known gene annotations, presence of open reading frames (ORFs), reciprocal matches and degree of coverage. Totals of 126,335 contigs from Kaspa and 145,730 from Parafield were subsequently selected as the reference set. Reciprocal sequence analysis revealed that c. 87 % of contigs were expressed in both cultivars, while a small proportion were unique to each genotype. Reads from different libraries were aligned to the genotype-specific assemblies in order to identify and characterise expression of contigs on a tissue-specific basis, of which 87 % were expressed in more than one tissue, while others showed distinct expression patterns in specific tissues, providing unique transcriptome signatures.

Conclusion: This study provided a comprehensive assembled and annotated transcriptome set for field pea that can be used for development of genetic markers, in order to assess genetic diversity, construct linkage maps, perform trait-dissection and implement whole-genome selection strategies in varietal improvement programs, as well to identify target genes for genetic modification approaches on the basis of annotation and expression analysis. In addition, the reference field pea transcriptome will prove highly valuable for comparative genomics studies and construction of a finalised genome sequence.

Keywords: Grain legume, Second-generation DNA sequencing, *De novo* assembly, Sequence annotation, Tissue-specific gene expression, Plant breeding

* Correspondence: john.forster@ecodev.vic.gov.au
[1]Department of Economic Development, Jobs, Transport and Resources, Biosciences Research Division, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, VIC 3083, Australia
[3]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3086, Australia
Full list of author information is available at the end of the article

Background

Field pea is a member of the Galegoid clade of the Papilionoideae sub-family of the Fabaceae family, and is a cool-season grain legume which is cultivated world-wide (6.4 million hectares per year) for both human consumption and stock-feed purposes [1]. Pea is a self-pollinated diploid species (2n = 2× = 14) with a genome size of c. 4,300 Mb, which is approximately 10-fold larger than that of the most closely related model legume species, *Medicago truncatula* Gaertn. (c. 500 Mb). This expansion is largely due to a substantial quantity of repetitive DNA (c. 50-70 % of the nuclear genome complement) composed of various families of mobile genetic elements [2, 3]. As a consequence, the exomic component (gene space) of the pea genome constitutes a much lower proportion of total genomic DNA than for legume species such as *M. truncatula* [4], *Lotus japonicus* L. (c. 472 Mb) [5], and chickpea (*Cicer arietinum* L.) (c. 740 Mb) [6], which have been the subjects of whole genome sequencing activities.

Enrichment of genetic resources for field pea is essential in order to provide effective tools for molecular breeding, with the aim of improving both productivity and quality of the crop, and sustainability of farming practices. Genic regions, which provide the primary targets for such activities, may be obtained by direct sampling of genomic sequences. However, given the relatively low proportion of such regions within the field pea genome, a more attractive current option is indirect sampling through access to the transcriptome, the latter being the actively transcribed sub-component of the genome in a given cell type at any particular stage of the life-cycle.

The increasingly high-throughput nature and declining costs of second-generation DNA sequencing have provided a durable solution for transcriptome analysis based on direct sequence evaluation through transcript discovery, identification of the transcriptional structure of a gene, detection of alternate splicing patterns, and quantification of expression levels [7, 8]. RNA sequencing (RNA-Seq), has been demonstrated to be superior to earlier methods such as microarrays for detection of low abundance transcripts, differentiation of biologically critical isoforms and identification of genetic variants such as alternative alleles [9]. Extensive transcript expression profiling has previously been performed in order to provide insight into the roles of different functional developmental modules (for *Arabidopsis thaliana*) [10], or different cell types and developmental processes (*Oryza sativa*) [11]. For legumes, specific transcriptional activities of genes across tissues and/or between organs such as nodule [12], seed [13] and flower [14], and their respective developmental stages, have been identified. In the recent past, RNA-Seq has been performed for genome-wide transcriptome characterisation of both model and non-model plant species including maize (*Zea mays*) [15, 16], perennial ryegrass (*Lolium perenne* L.) [17] and soybean (*Glycine max* [L.] Merr.) [18]. RNA-Seq has also been employed to understand transcriptomic dynamics during plant responses to different biotic [19, 20] and abiotic stresses [21, 22].

Previous transcriptome analysis studies of field pea have been performed using second-generation DNA sequencing technologies, specifically the Roche 454 pyrosequencing system, mainly for use in the development of genetic markers [23–25]. Considerable progress has been made in the development of pair-cross specific [25–28] and consensus genetic linkage maps [25, 27, 28], based on use of such markers. Advances in genomics technologies now provide opportunities to develop substantially enriched genomic resources for field pea in order to assist accelerated delivery of improved cultivars. As the genome sequences of a number of legume species are now available, including *M. truncatula* [4], *L. japonicus* [5], chickpea [6], soybean [29] and pigeonpea (*Cajanus cajan* [L.] Millsp.) [30], corresponding data may be exploited for interpretation of transcriptome resources from less-well characterised taxa, such as field pea, to support gene annotation and comparative genome analysis.

In the present study, comprehensive transcriptome sets were generated from two genotypes of field pea that differ in terms of seed and plant morphological characteristics through use of RNA-Seq, followed by assembly, comparison to gene complements in related species, sequence annotation and assessment of tissue-specific expression. The resulting data provides a large-scale resource for the development of tools for molecular breeding of this important grain legume species.

Methods

Plant materials

Six plants from each of the two field pea cultivars, Kaspa (semi-leafless variety with medium height and produces spherical medium sized dun-type grain) and Parafield (conventional plant morphology and produces large sized dun-type grain) were maintained in glasshouses at 22 ± 2 °C under a 16/8-h (light/dark) photoperiod in individual pots filled with standard potting mix at the premises of DEDJTR - Bundoora, Victoria, Australia. Leaf, stipule, stem, tendril tissues from multiple nodes (at different developmental stages) as well as the root and root-tip tissues were collected from 4 weeks-old plants (three replicates per genotype). Fully open flowers, stamens, pistils, immature pods (10–14 days after flowering), immature seeds (20–25 days after flowering) and nodules (from 3 months old plants) were collected from three replicates of the two genotypes. To collect seedlings (7 days old), Kaspa and Parafield seeds (three replicates) were germinated on moist Whatman

filter paper in falcon tubes and maintained in growth chambers at 22 ± 2 °C under a 16/8-h (light/dark) photoperiodic regime. After harvest, tissues were frozen immediately in liquid nitrogen and stored at −80 °C until required. For RNA extraction, replicates from individual tissues were pooled in equal proportion (by weight).

Total RNA was extracted using the RNeasy® Plant Mini Kit (QIAGEN, Hilden, Germany) following manufacturer's instructions. A slight modification of the procedure was performed in order to extract RNA from immature seeds, which involved the addition of polyvinylpyrrolidone (PVP-40) (Sigma-Aldrich, Missouri, USA) to 450 μl of Buffer RLT containing 10 μl/ml β-mercaptoethanol (Sigma-Aldrich) to a final proportion of 2 % (w/v), and the remainder of the protocol was implemented according to manufacturer's instructions. Aliquots of purified RNA were stored at −80 °C. The concentration of RNA was confirmed using a spectrophotometer (Thermo-Scientific, Delaware, USA) at the two wavelength ratios of A260/230 and A260/280 nm. The integrity of total RNA was determined by electrophoretic separation on 1.2 % (w/v) denaturing agarose gels.

Library preparation

The polyA-containing (mRNA) fraction was isolated from total RNA (1 μg) using Dynabeads® with oligo $(dT)_{25}$ residues covalently coupled to the surface (Life Technologies Australia Pty Ltd, Victoria, Australia). After purification, mRNA was fragmented by random shearing using heat treatment in the presence of Mg^{2+} ions. The resulting small fragments were used as templates to synthesise first-strand cDNA using random hexamer priming and the Bioline Tetro cDNA Synthesis Kit (Bioline US Inc., Massachusetts, USA). Second strand synthesis was performed in a solution containing NEBuffer 2 (New England Biolabs Ltd., Hitchin, United Kingdom), dNTPs (Bioline US Inc.), RNaseH (New England Biolabs) and DNA polymerase I (Thermo-Scientific). Subsequently, the ends of nascent cDNAs were polished by the addition of T4 DNA Polymerase (New England Biolabs) and Klenow DNA Polymerase (New England Biolabs) followed by paired-end adapter ligation to both termini of the cDNA fragments. Templates were then subjected to amplification using Phusion (Thermo-Scientific) DNA polymerase. The amplified libraries were pooled in equimolar amount and assessed by loading of a 1 μl aliquot on an Agilent Bioanalyzer 1000 DNA chip according to the manufacturer's instructions. Library quantification was performed using the KAPA library quantification kit (KAPA Biosystems, Boston, USA). All reads were pair-end sequenced using the HiSeq 2000 and MiSeq platforms (Illumina Inc., San Diego, USA).

De novo *transcriptome sequence assembly*

After removal of adaptor sequences along with low quality reads, sequence reads from each cultivar were *de novo* assembled using two transcriptome assemblers, Trinity (Trinityrnaseq_r20131110) [31] with the default settings including a fixed *k*-mer size of 25 and SOAPdenovo-Trans v1.03 [32], with different *k*-mer sizes (35, 45 and 55). To evaluate the quality of the assemblies, N50 statistics, contig counts and contig length distributions were assessed. The assemblies of the two cultivars were subsequently labelled as being derived from Kaspa and Parafield, respectively. Contigs from the assembly were further combined using CAP3 assembler with 95 % identity and minimum of 50 bp overlap to produce longer, more complete consensus sequences [33]. The CAP3 software removed the redundancies generated within assembly by consolidating the transcripts using overlap-layout-consensus (OLC) approach [33].

Functional annotation and classification of the transcriptome

A workflow detailing the process of annotation and classification of the field pea transcriptome is shown in Fig. 1. All assembled contigs were searched against the non-redundant (nr) protein database maintained by NCBI using BLASTX [34] under the threshold parameter of E-value $< 10^{-10}$. Any contigs that showed significant matches to non-plant databases were excluded from further analysis. For further assembly annotation, the contigs were utilised for similarity searches against the NCBI nucleotide (nt) database, genomes and coding DNA sequences (CDS) of *M. truncatula* (medicago v3.5) [4], chickpea [35] and soybean [36] using BLASTN with a threshold E-value of $< 10^{-10}$ to capture any genomic sequences that may have been missed by BLASTX analysis. In order to obtain the final transcriptome set, results from the nr database were preferentially selected followed by those from the nt and other legume databases, where necessary. Both Kaspa and Parafield contigs were queried by BLASTN analysis against the pea chloroplast genome sequence (NCBI RefSeq NC_014057.1) in order to identify chloroplast-derived sequences.

Reads from each cultivar were also reciprocally compared to the assembly from the alternate cultivar as a reference using Bowtie 2 [37], in order to obtain estimates of common genic content between the two cultivars. The read counts of contigs that had no significant hit to the reference (other field pea genotype) were also examined, as these may represent contigs that were not assembled or expressed in the other transcriptome.

Subsequently, any contigs that were not annotated in the above mentioned procedures were searched for the presence of open reading frames (ORFs) using the 'getorf' function in the EMBOSS package [38] with

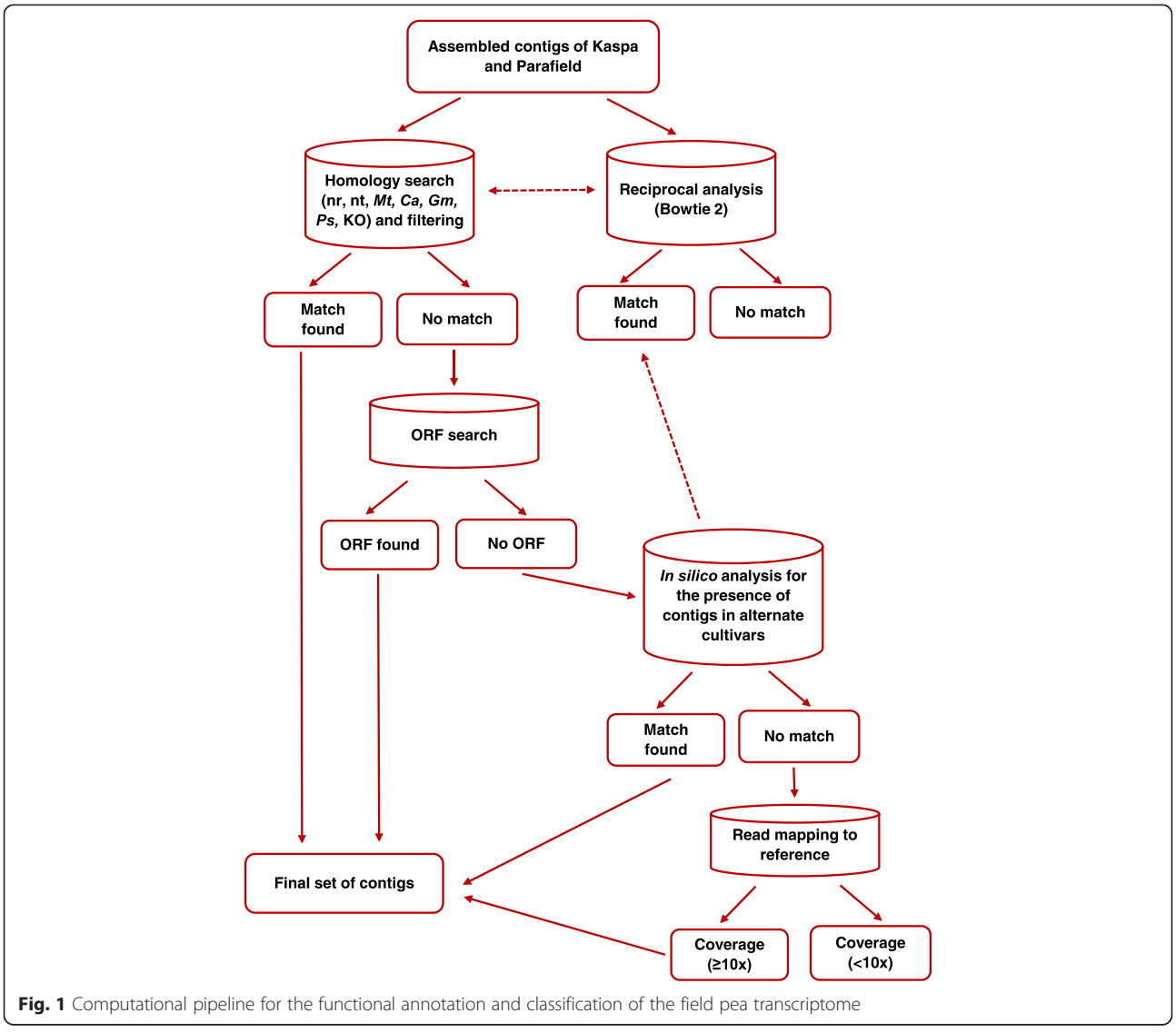

Fig. 1 Computational pipeline for the functional annotation and classification of the field pea transcriptome

minimum nucleotide size of 100 between START and STOP codons.

All annotated contigs were compared to the Kyoto Encyclopedia of Genes and Genomes (KEGG) database based on BLASTX queries. The KEGG pathway annotation was performed in the KEGG Automatic Annotation Server (KAAS) [39, 40] to further characterise the assembly.

To validate the current assembly, unigene sequences from previous pea transcriptome sequencing studies [23–25, 41] were aligned to the transcriptome dataset generated in the current study using BLASTN with an E-value < 10^{-10}.

Tissue-specific expression analysis

The trimmed reads from each library were aligned to the genotype-specific transcriptome through the use of Bowtie 2 [37], to obtain tissue-specific gene expression data. Relative expression based on read counts was used for this purpose, as the individual libraries varied in terms of read numbers. Normalisation of read counts from individual libraries of each cultivars were performed in MS Excel, based on the 75th percentile value. For this, read counts from each libraries were multiplied by library specific scaling factor. This factor was calculated by dividing the maximum 75th percentile value among different libraries by 75th percentile value of the particular library (read normalisation = read counts × (the maximum 75th percentile value/75th percentile value of particular library). Previously, experiments were conducted comparing the reads normalisation as described above using MS Excel and RPKMs in "R Software", revealing a 99 % correlation between the two methodologies (unpublished data). Read counts from different

tissues were grouped into three clusters: reproductive tissues (flower, immature pod and immature seed); subterranean tissues (root, root-tip and nodule); and vegetative tissues (leaf, stem, stipule, seedling and tendril). The transcript expression profile was analysed in each instance.

The trimmed reads from root, root-tip and nodule libraries were further BLASTN analysed against various fungi and bacterial sequence (downloaded from NCBI) collections in order to estimate the possible presence of bacterial genes present within these tissues.

Results

De novo *sequence assembly of field pea transcriptome*

To generate a comprehensive transcriptome dataset for field pea, a total of 23 cDNA libraries were generated from the various target tissues of the two cultivars, and were sequenced using both the HiSeq 2000 and MiSeq platforms. For cv. Kaspa, a total of 432 million paired-end reads with an average read length of 100 bp were obtained from the HiSeq 2000, as compared to 4 million paired-end reads with average read length of 250 bp from the MiSeq. The comparable figures from cv. Parafield were 372 million paired end reads with an average read length of 100 bp from the HiSeq 2000, and 3.7 million paired end reads with an average read length of 250 bp from the MiSeq. Details of the sequencing outcomes for each tissue-specific library of both varieties are provided in Additional file 1. An average of 35.2 million reads were generated per tissue type. After strict quality filtering, 408 million and 352 million reads (Table 1) from Kaspa and Parafield, respectively, were used for *de novo* assembly. Trinity assemblies were selected for further analysis, which produced 201,317 transcripts with N50 of 781 bp (Kaspa) and 226,701 transcripts with N50 of 772 bp (Parafield) (Table 2). Further CAP3 assembly in the former resulted in 129,282 contigs, while the latter constituted 149,272 contigs (Table 2). The contig length distribution from both assemblies is shown in Fig. 2.

Table 1 Details of the reads used for *de novo* transcriptome assembly

Cultivar	Source tissue	Number of reads used for assembly
Kaspa	Flower	49,261,131
	Immature pod	47,420,170
	Immature seed	33,879,162
	Nodule	31,101,358
	Pistil	25,767,536
	Root	33,938,588
	Root-tip	34,409,507
	Seedling	33,974,057
	Stamen	19,487,248
	Stem	35,122,274
	Stipule	33,995,906
	Tendril	29,401,978
	Total	407,758,914
Parafield	Flower	47,432,143
	Immature pod	18,675,349
	Immature seed	13,481,117
	Leaf	21,943,059
	Nodule	20,744,191
	Root	27,353,394
	Root-tip	27,865,916
	Seedling	38,346,614
	Stem	41,295,663
	Stipule	65,339,912
	Tendril	29,495,107
	Total	351,972,464

Functional annotation and classification of field pea transcriptome

In order to annotate the transcriptomes, all contigs were BLASTX analysed against the nr database of GenBank. For the Kaspa transcriptome, BLASTX analysis (Additional file 2) revealed 60,808 sequences (47 %) with significant matches, which were then filtered to remove non-plant sequences. This process resulted in a set of 59,229 sequences corresponding to 27,145 unique gene clusters. The length of the annotated sequences varied from 201 to 7,802 bp, with an average of 809 bp, and N50 of 1,106 bp. There were 34,452 (59 %) annotated sequences ≥ 500 bp, in which 15,867 sequences were longer than 1,000 bp, and the remaining 41 % of sequences were 201–500 bp in size. The E-value distribution of significant hits revealed that 48 % of matched sequences exhibited high levels of similarity (E-value lower than 10^{-50}) to other legume genomes (Additional file 3, Figure A). For the Parafield transcriptome, 64,727 (43 %) of sequences exhibited significant BLASTX hits

Table 2 Overview of sequencing outputs and assembly

	Kaspa	Parafield
Total raw reads	436,282,428	374,354,188
Total clean reads	407,758,914	351,972,464
Trinity		
Total number of contigs	201,317	226,701
N50	781	772
CAP3		
Total number of contigs	129,282	149,272
N50	757	717

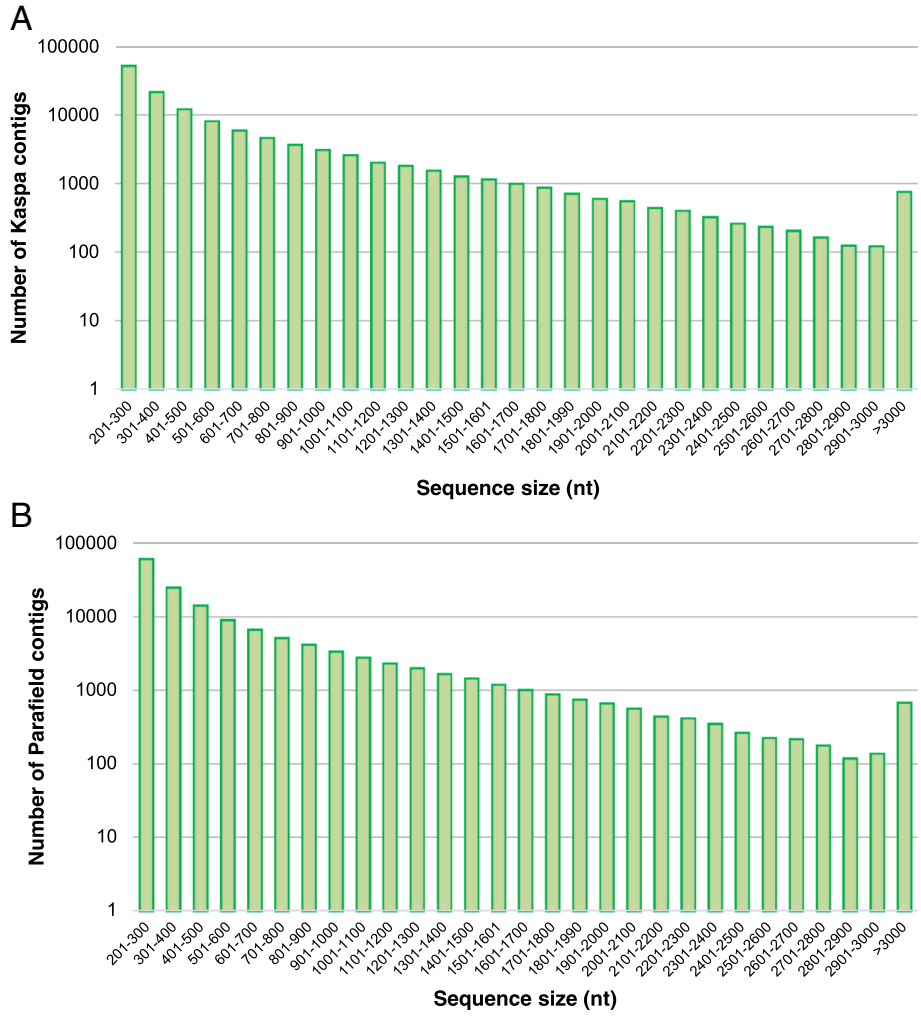

Fig. 2 Length distribution of contigs from the (**a**) Kaspa-specific assembly, and (**b**) Parafield-specific assembly

(Additional file 2), and after the removal of the non-plant sequences, 63,843 sequences (N50 of 1,083 bp and average 797 bp) remained, corresponding to 27,655 unique genes. Among the annotated sequences, 36,979 (58 %) were greater than 500 bp in length, whereas 26,863 sequences were 201–500 bp in length. The distribution of significant hits for the Parafield contigs showed that 48 % of the sequences displayed E-values less than 10^{-50}, while the other matching sequences were located in the value range between 10^{-50} and 10^{-10} (Additional file 3, Figure A). The annotated contigs were also examined for the presence of repetitive elements, and c. 1 % of the contigs were annotated as repeat elements such as retrotransposons, *gag* polyprotein-encoding etc. The distribution of gene annotations based on BLASTX analysis exhibited a highest number of hits against sequences of *M. truncatula*, followed by soybean, and so-far published pea protein sequences within the nr database of NCBI (Additional file 3, Figure B). The BLASTN analysis of transcriptome contigs (Additional file 4) identified a higher number of matches (Fig. 3) to the NCBI nt database as compared to BLASTX analysis against nr. However, most of these additional matches were annotated as retrotransposons and hypothetical proteins, without well-characterised functions. The BLASTN analysis of transcriptome contigs (Additional file 4) against the pea chloroplast genome identified up to 0.17 % of contigs to be chloroplast-derived.

Contigs from the assembly were also BLASTN analysed against both the genomes and CDS of chickpea, *M. truncatula* and soybean. A total of 72,651 (56 %) Kaspa contigs (Additional file 5A and B) and 73,621 (49 %) Parafield contigs (Additional file 5C and D) could be mapped to any of these reference species. Of the total 72,651 Kaspa contigs, 38,924 (53 %) contigs were found to have common matches between chickpea, *M. truncatula* and soybean, while for Parafield out of 73,621 contigs 42,581 (57 %) were found to be common between

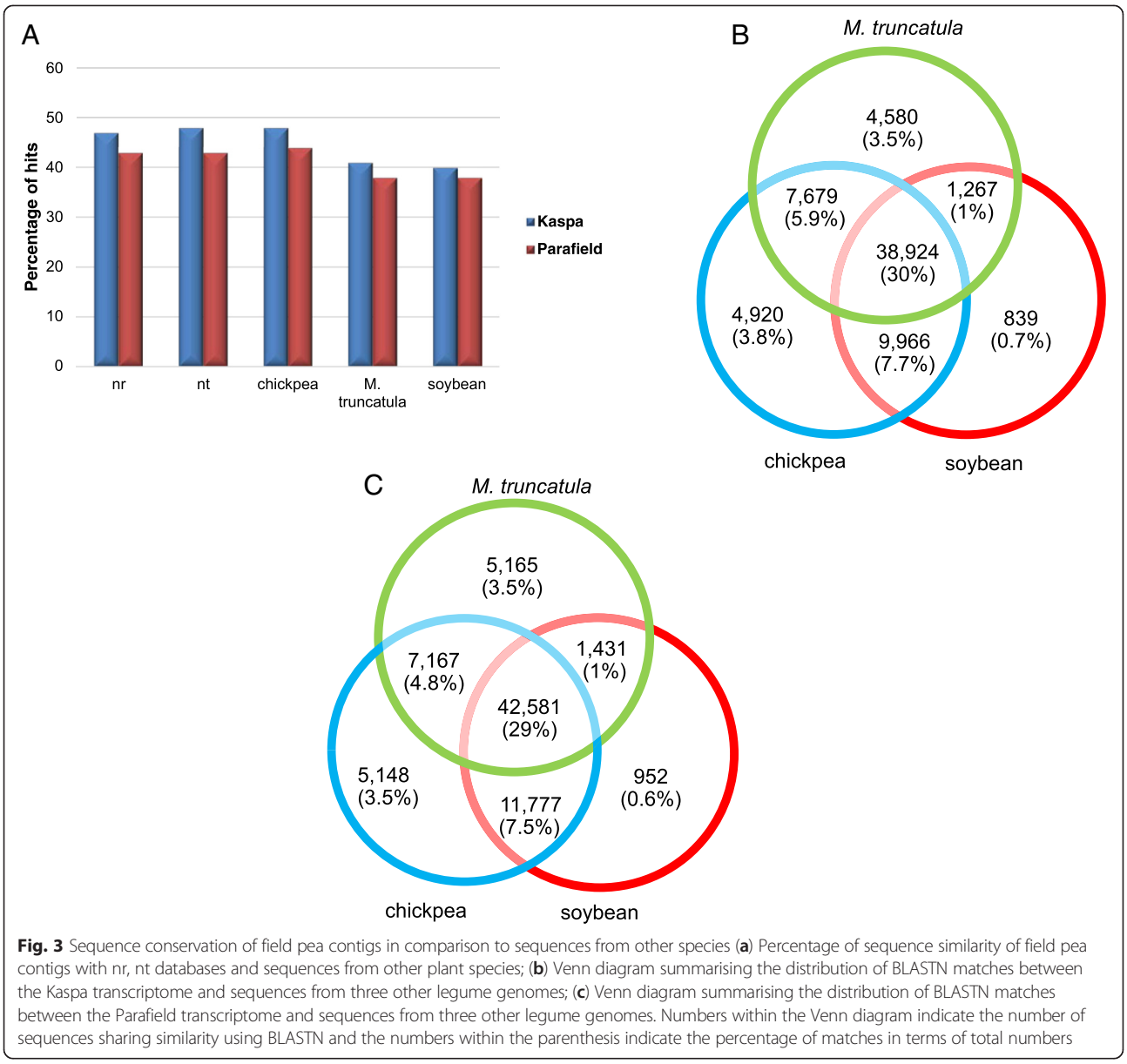

Fig. 3 Sequence conservation of field pea contigs in comparison to sequences from other species (**a**) Percentage of sequence similarity of field pea contigs with nr, nt databases and sequences from other plant species; (**b**) Venn diagram summarising the distribution of BLASTN matches between the Kaspa transcriptome and sequences from three other legume genomes; (**c**) Venn diagram summarising the distribution of BLASTN matches between the Parafield transcriptome and sequences from three other legume genomes. Numbers within the Venn diagram indicate the number of sequences sharing similarity using BLASTN and the numbers within the parenthesis indicate the percentage of matches in terms of total numbers

all reference species. Other contigs were either common between any two of the three references, or specific to each reference (3.5 % to chickpea and *M. truncatula*, 0.7 % to soybean) (Fig. 3).

Reciprocal reference read mapping indicated a large number of contigs that showed matches to the other genotype (87.2 % of Kaspa contigs matched to Parafield reads, and 82.7 % of Parafield contigs matched to Kaspa reads). Among the shared contigs, specific genes known to be essential for plant development and function were identified, including but not limited to chlorophyll a-b binding protein AB80, cytochrome P450, dehydrin-cognate and seed albumin PA1. The contigs with no significant match to the other genotype were also examined and identified as hypothetical proteins, disease resistance genes, stress-related proteins etc.

In order to characterise the assembled contigs and identify active biological processes, annotated sequences were mapped to the reference biochemical pathways in the KEGG database using eudicot species such as *Arabidopsis thaliana* (L.) Heynh., cocao (*Theobroma cacao* L.), soybean, alpine strawberry (*Fragaria vesca* L.), grapevine (*Vitis vinifera* L.), potato (*Solanum lycopersicum* L.) and rice (*Oryza sativa* sp. *japonica*) as references. In total, 22,056 (37.3 %) contigs from Kaspa and 23,692 (37.1 %) contigs from Parafield were mapped to 157 KEGG pathways corresponding to five modules; metabolism, cellular processes, genetic information processing, environmental information processing and organismal

systems (Additional file 6). Metabolic pathways were well represented, most of which were associated with biosynthesis of secondary metabolites, carbohydrate metabolism and amino acid metabolism. Furthermore, mapping of contigs against the glycolysis/gluconeogenesis pathway revealed that all of the genes involved in this pathway were present in the dataset. Another important pathway (nitrogen metabolism), which is crucial to legume species, was also analysed and revealed the presence of all known genes (Fig. 4). In addition, genes for all key enzymes required for the legume-specific isoflavonoid biosynthesis pathway were identified, using *M. truncatula* and chickpea as references.

Comparison of *de novo* assembled contigs from each genotype to previously published pea transcriptome databases indicated that the current activity captured majority of the previously described contigs (99 % of the contigs described by [24], 96 % contigs from [25], 98 % contigs from [23] and 72 % contigs from [40]), representing 33-36 % (46,631 Kaspa contigs and 49,424 Parafield contigs) of the current assemblies (Additional file 7A and B).

In summary, a total of 80,592 contigs from Kaspa and 88,487 from Parafield were annotated and characterised using the similarity searches as described. However, a large proportion of contigs from both cultivars (47,058 from Kaspa and 60,280 from Parafield) still remained uncharacterised. These sub-sets were further evaluated and searched for the presence of ORFs. This process identified an additional 23,800 contigs from Kaspa (Additional file 8A) and 28,047 from Parafield (Additional file 8B) which contained a START and STOP codon with minimum

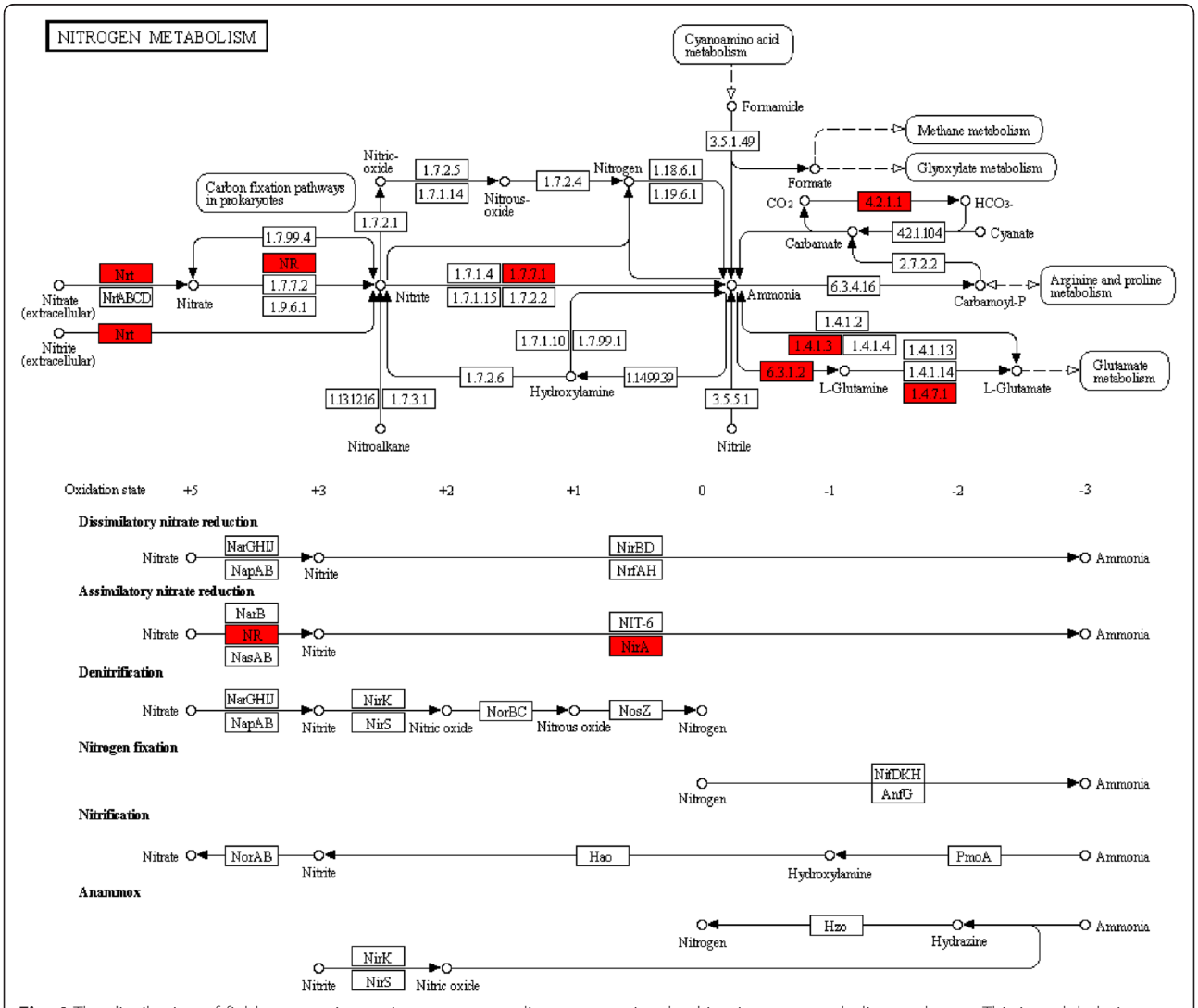

Fig. 4 The distribution of field pea contigs against genes encoding enzymes involved in nitrogen metabolism pathways. This is a global nitrogen metabolism pathway map in which a red colour indicates genes identified in data from the present study, all of the known nitrogen metabolism genes in legumes having been identified

sequence length of 100 bp. An additional 16,178 contigs from Kaspa and 20,602 from Parafield were identified in the reciprocal searches (Fig. 5). A final set of contigs (126,335 contigs in Kaspa and 145,730 contigs in Parafield) were compiled after further selection of contigs from the remaining sub-set based on level of coverage (≥10×), although this threshold requirement prevented discovery of lowly expressed novel contigs in field pea. All filtered contigs from both Kaspa and Parafield assemblies were deposited in DDBJ/EMBL/GenBank (accession numbers, Kaspa - GCMF00000000, GCMG00000000, GCMH00000000, GCMI00000000, GCMJ00000000, GCMK00000000 and GCML00000000 and Parafield - GCKA00000000, GCMM00000000, GCMN00000000, GCMO00000000, GCMP00000000 and GCMQ00000000).

Tissue-specific expression analysis

Reads from the individual (tissue-specific) libraries were aligned to genotype-specific assemblies. Most of the tissues showed expression of a similar number of contigs, with the exception of immature seeds for which a relatively lower number was observed (Fig. 6a). Expression of contigs from reproductive tissues, subterranean tissues and vegetative tissues of the two genotypes was compared through use of Venn diagrams (Figs. 6b and c). A total of 62 % of contigs were common between the three groups.

Analysis of tissue-specific expression revealed that roots expressed the maximum number of tissue-specific contigs (Fig. 6d). Flowers and nodules expressed more tissue-specific contigs than immature pods and leaves, while there were very few contigs that were expressed exclusively in immature seed. Detailed contig expression lists for each tissue are provided in Additional file 9A and B. Approximately 87 % of the contigs (109,281 in Kaspa and 130,297 in Parafield) were expressed in more than one tissue in both genotypes. Only a small proportion (0.2 % in immature seed - 5 % in root) of contigs

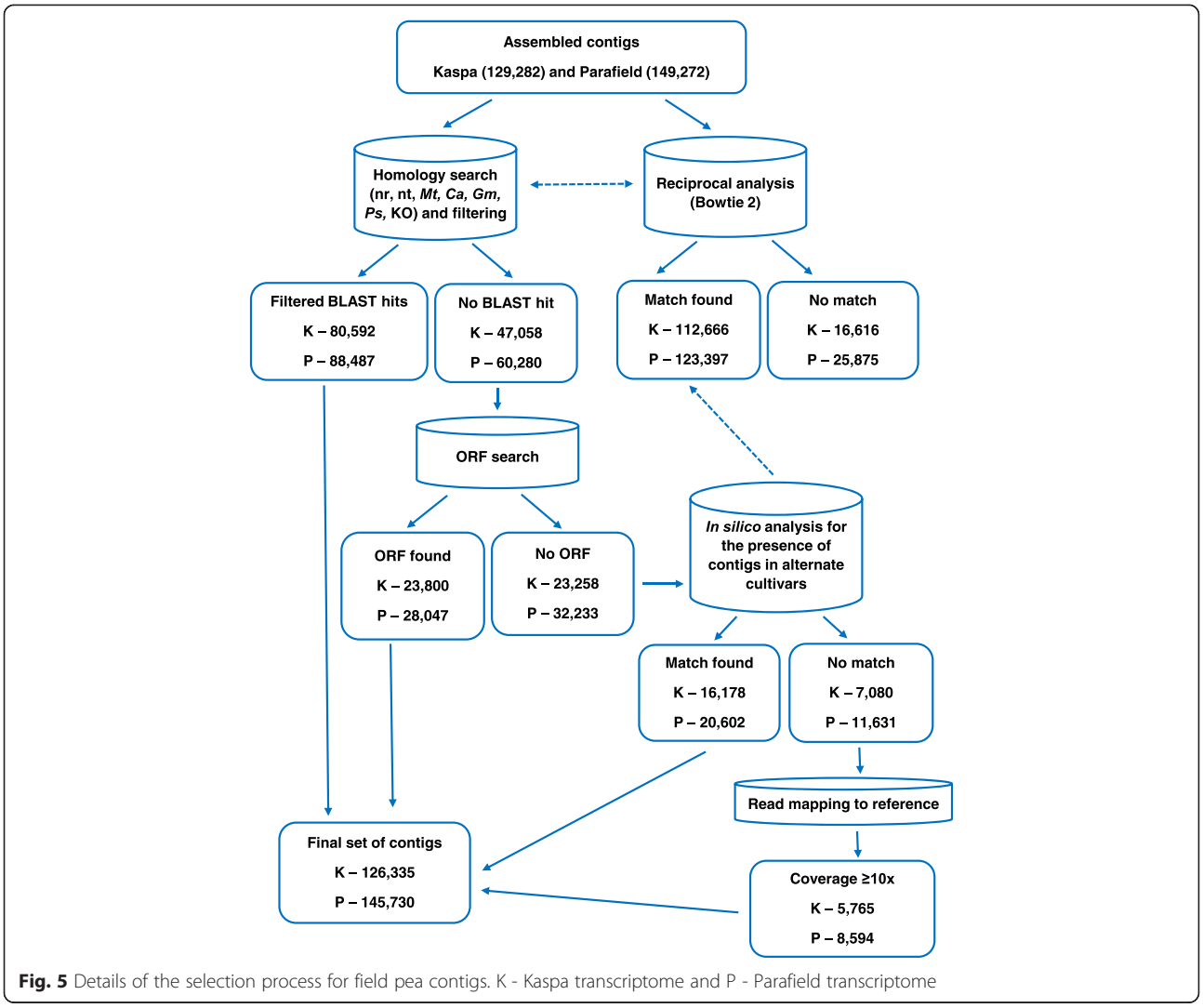

Fig. 5 Details of the selection process for field pea contigs. K - Kaspa transcriptome and P - Parafield transcriptome

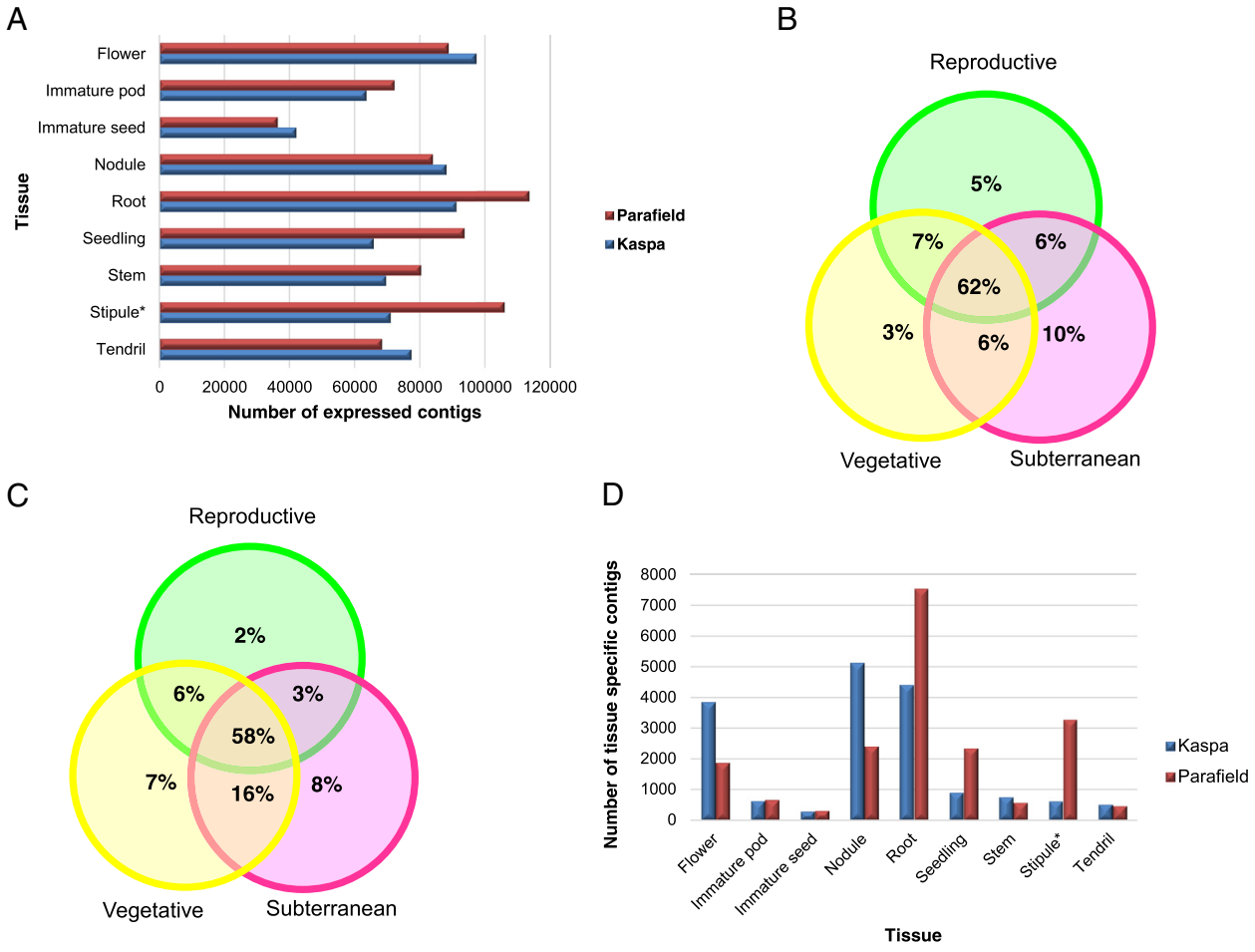

Fig. 6 Expression patterns in different tissue samples: (**a**) Number of contigs expressed in each tissue sample; (**b**) Percentage of shared and specific expression profiles of contigs expressed in Kaspa; (**c**) Percentage of shared and specific expression profiles of contigs expressed in Parafield; (**d**) Number of tissue-specific contigs. *For Parafield, stipule and leaflet tissue-derived read counts were merged, while Kaspa contributed only stipule tissue-derived reads

were specific to a particular tissue type (Fig. 6d). Using the read counts obtained from mapping, expression levels of commonly expressed contigs were assessed, revealing that different tissues displayed variable expression levels for the common contigs. For example, storage proteins such as albumin and vicilin were highly expressed in immature seeds compared to other tissues. Similarly, the genes for the small sub-unit of ribulose bisphosphate carboxylase-oxygenase (RUBISCO) and the light-harvesting chlorophyll-a/b binding (CAB) protein Lhcb1 were expressed at a much higher level in leaves when compared to other tissues. Assessment of the annotation of tissue-specific contigs indicated predominant involvement in functions particular to that tissue type (Additional file 10).

A high level of expression was noted from root, root-tip and nodule tissues from read mapping. Further analysis of contigs from these tissues to fungal and bacterial sequences revealed that only a small percentage of reads showed matches to non-plant references, of which *Rhizobium* was one of commonly represented species, particularly (as expected) in nodule-derived assemblies (0.5 % of nodule, 0.05 % of root-tip and 0.01 % of root mapped reads from Kaspa exhibited similarity to *Rhizobium*, as compared to 0.3 % of nodule and 0.01 % of root tissues mapped reads from Parafield).

Discussion

De novo *sequence assembly and functional annotation of the field pea transcriptome*

Legume species such as peas are economically important as sources of food for humans, feed for livestock and contributors to sustainable agriculture due to the ability to fix atmospheric nitrogen in symbiotic association with *Rhizobium* bacteria, hence providing crop plants with a free and renewable source of nitrogen [42]. A fundamental understanding of the field pea transcriptome will provide an overview of the genes, regulatory networks and

functional roles that govern these key biological processes [43]. Additional genomic and transcriptomic resources may permit discovery of novel genes associated with multiple agronomic traits useful for plant breeding purposes.

The RNA-Seq approach has a wide range of applications, including investigation of different biological processes at the tissue or cell level [44], description of the entire transcriptome of a given organism [17], and assessment of genetic diversity on an evolutionary time scale [45]. RNA-Seq technology has previously been used to characterise the transcriptomes of a number of plant species, including maize and soybean [15, 16, 18], and decreasing costs of DNA sequencing technologies will provide opportunities for the generation of equivalent information for many more species in future. Due to developments in sequence analysis software, transcriptome studies are now possible for species that do not as yet have a reference genome [46, 47], of which field pea is an example.

In order to characterise the field pea transcriptome, two commonly cultivated Australian field pea cultivars (Kaspa and Parafield) were subjected to sequencing of cDNA samples. A range of different tissues were included in order to obtain effective sampling of transcript complexity and to maximise the probability of detecting mRNA of low abundance. In total, c. 408 million and 352 million high quality reads were used for *de novo* assembly, generating 129,282 and 149,272 contigs for Kaspa and Parafield respectively. The number of contigs generated in the current study is comparable to that from other studies [17], which used Illumina sequencing data and Trinity assembler. Moreover, SOAP-denovo-Trans assembler also generated a similar number of contigs, with a lower N50 value and therefore not used for further analysis.

Approximately 50 % of the contigs were annotated by comparison with the NCBI nr database. The majority of contigs exhibited significant matches to *M. truncatula* sequences, followed by those from soybean. However, only 3.3 % of the BLASTX-mediated matches were to pea-derived protein sequences, probably due to the limited number of proteins currently available in the NCBI database (3,689), as opposed to failure to recognise homologous sequences. As a consequence, the number of observed matches reflects not only the degree of relatedness between comparator species, but also the scope of available sequence data. Nonetheless, the proportion of similarity of field pea transcriptome sequences against the NCBI nr database is comparable to results from other species, such as sweet potato (*Ipomoea batata*) [44].

The field pea transcriptome assembly contained small sequences as well as unigenes containing more than one sequence. Trinity assembler generates high numbers of putative transcripts, including alternatively spliced isoforms and transcripts from recently duplicated genes which lead to the generation of similar transcripts [48]. Moreover, Trinity-derived transcripts are not scaffolded across sequencing gaps, which may also lead to generation of a large number of small transcripts. The small-sized sequences (in the range 200–300 bp) may be too small for BLASTX analysis, and may hence have failed to detect similarity to any known proteins [49]. Alternatively, the small sequences may encode novel proteins, or be derived from untranslated regions (UTRs) or non-coding RNAs (ncRNAs) [17]. The multiple contigs that were assembled into unigene clusters may represent transcription variants, allelic variants, closely related paralogous sequences, misassembled transcripts, or transcripts that were fragmented due to low coverage [46].

Sequence similarity of field pea contigs to the genomes and transcriptomes of other legume species was determined using BLASTN analysis, revealing levels of conservation up to 56 %. Comparable results were observed for the chickpea transcriptome in similarity searches against other legumes [6]. Moreover, up to 57 % of the annotated contigs were common to all of the legume genomes used in the present study. BLAST analysis revealed a highest level of similarity to sequences from chickpea, followed by *M. truncatula*, and more distantly, by soybean. Field pea, chickpea and *M. truncatula* all belong to the Galegoid clade of cool-season legumes, and are hence mutually more closely related than to soybean, which belongs to the Phaseoloid clade of warm-season legume species within the Papilionoideae [50]. A comparatively lower level of similarity to soybean genomic sequences, is hence not unexpected, considering the evolutionary divergence of the various species. However, the rank order of similarity within the Galegoid clade is not so easy to rationalise, as the Vicieae and Trifolieae tribes (to which field pea and *M. truncatula* belong) are thought to share a common ancestor more recently in evolutionary time than with the Cicereae tribe, to which chickpea belongs [51]. A proportion of the unmatched field pea contigs may genuinely represent species-specific components of the field pea transcriptome. Similarity searches against the genomes of the legume comparators identified matches to field pea contigs which did not display matches to the CDS, possibly due to incomplete annotation of those genomes. Previous studies have also reported that clusters of reads from transcriptome sequencing were mapped to the unannotated regions of the genome [52].

Within the KEGG analysis, well-represented pathways in the field pea transcriptome included those involved in carbohydrate metabolism, biosynthesis of secondary metabolites, amino acid metabolism, lipid metabolism and

energy metabolism. All of the expected genes involved in the isoflavonoid biosynthesis pathway were identified, though expression was lower than that of the genes involved in synthesis of other secondary metabolites such as phenylpropanoids and flavonoids. In addition, activities associated with genetic information processing (spliceosome, ribosome and RNA transport functions), plant-pathogen interactions and plant hormone signal transduction were identified.

The two field pea genotypes used in this study differ in terms of morphology and resistance to different abiotic and biotic stresses. Kaspa is a high-yielding, late flowering field pea variety with excellent pod-shatter resistance, good lodging tolerance, resistance to downy mildew and improved resistance to black spot. The Parafield cultivar is mid- to late-season flowering, with moderate resistance to pod-shattering, moderate resistance to bacterial blight and tolerance to saline soil toxicity. The observed differences in gene expression between genotypes may account for some of the performance differences between cultivars. Reciprocal sequence analysis identified 23-27 % of contigs which displayed no significant match to any transcript in the other genotype. Those contigs may not be present in the other genome, or much more likely, have not been expressed at a sufficiently high level to undergo assembly. Based on sequence annotation, the major differences in gene expression between the two genotypes were identified as being associated with abiotic/biotic stress tolerance (at low levels), transcription factors and signal transducers. Conversely, the genes held in common between the two genotypes included those encoding proteins known to be necessary for development and function, such as chlorophyll a-b binding protein AB80, dehydrin-cognate, cytochrome P450, disease resistance proteins and ABC transporters.

For both the Kaspa and Parafield transcriptomes, the proportion of sequences that are present in previous pea transcriptome datasets [23–25, 41] were assessed, revealing that from 72 to 99 % of the unigene sets from those assemblies were regenerated in the current study. Based on this comparisons, it would appear that the current study was able to reconstruct a higher number of assembled contigs than those obtained from other assembly processes.

After step-by-step annotation and classification of the field pea transcriptome, totals of 126,335 contigs in Kaspa and 145,730 contigs in Parafield were obtained, representing 71,014,518 bp and 79,440,852 bp of cumulative sequence, respectively. These sets include contigs that may represent alternatively spliced forms of the same gene locus. Although this sampling process has been highly effective, a determination of the exact composition of the pea transcriptome will require a corresponding genomic sequence assembly, permitting annotation and classification of a broader range of transcripts. As c. 40 % of the sequences generated in the present study lacked significant similarity to genes of known function, alternative computational means were used to identify more sequences, such as identification of ORFs, use of reciprocal analysis and relative coverage in the transcriptome. The results of these analyses were used to annotate and classify the remaining sequences, but these processes are still be prone to exclude contigs with low levels of expression. Limited sequencing depth of lowly expressed contigs can cause sequencing biases, resulting in the partial assembly of contigs which may fail to be classified by comparison to known gene annotations. A high proportion of these lowly expressed contigs may be derived from pea-specific repetitive elements, belonging to several sub-families which are highly variable in sequence and hence individually present in relatively lower copy number [2].

Tissue-specific expression analysis

In order to identify and characterise expression of contigs on a tissue-specific basis, reads from different libraries were aligned to the genotype-specific assemblies. The number of contigs detected was similar for most samples, with the exception of those from seeds, despite generation of a similar number of reads. A similar observation was reported in a previous study of seed-specific transcription in *A. thaliana* [10]. Despite the similar number of active contigs in each sample, expression dynamics varied considerably between tissues, the largest number of contigs showing preferential expression in root. Substantial overlap in expressed contigs was identified between reproductive, vegetative and subterranean tissue-derived clusters, c. 62 % of contigs being attributed to a generic expression profile, while smaller cohorts displayed tissue-specific expression. The root tissue-derived group displayed the most diverse transcriptome, as compared to the reproductive tissue-derived group which may be associated with regulation of root apical meristem cells, pathogen resistance, symbiosis and immune responses [53, 54]. A larger number of vegetative tissue-specific contigs were identified in the Parafield transcriptome as compared to that from Kaspa, possibly because Parafield contributed sample from leaf tissue, in addition to stipule-derived contigs which were common to both.

The analysis was performed in more detail at the level of individual organ types, which demonstrated that roots contributed the largest number of tissue-specific contigs, followed by flower and nodule, while immature seed contributed the least. The identity of organ types that contribute the largest number of tissue-specific contigs varies between legume species, from nodules and flowers

in *L. japonicus* [14] and soybean [55], to flower bud and immature pod in chickpea [56]. Assessment of the annotation of some of the contigs that are expressed only in nodule tissue (such as those for nodule inception protein and nodulin) revealed involvement in nodule developmental processes [57, 58]. Likewise, some of the contigs expressed specifically in roots (such as those for hyoscyamine 6-dioxygenase, acyltransferase, MtN19 protein and germin-like protein) were associated with root development, root wax formation, and defence/wounding-related process that are also implicated in the legume-Rhizobium symbiosis [53, 54]. The MADS box protein PIM, which is represented only among flower contigs regulates floral meristem identity in pea [59]. Many immature seed-specific contigs represented transcription factors, including the BZIP transcription factor that is involved in seed maturation [60]. The results of tissue-specific analysis indicate that different tissues express distinct contigs, many of which are clearly related to biological functions, providing a unique transcriptome signature for that tissue. These tissue-specific contigs may provide further insight into specialised organ-specific biochemical, physiological, and developmental processes.

The contigs that were expressed at very low levels without annotation and classification, and also without any reciprocal match to the transcriptome of the other cultivar were further analysed. The lowly expressed cultivar-specific contigs were preferentially associated with nodule and root tissues in both transcriptomes. Field pea root and nodule tissues may hence possess novel contigs associated with specific functions such as rhizobial symbiosis and nitrogen fixation, or these contigs could represent sequences from novel bacteria, although additional studies will be required for validation of this hypothesis.

Conclusions

The present study has demonstrated that RNA-Seq technology provides an efficient method for transcriptome analysis of non-model plant organisms, delivering a valuable resource of gene expression data for further analysis. Gene annotation and understanding of potential pathways provides the basis for investigation of specific processes, biological functions, gene interactions and mechanisms involved in different agronomic traits. The transcript expression patterns were generally similar between different tissues, but the tissue-specific contigs from different libraries displayed signatures which were consistent with biological expectations. The combined transcriptomes of two contrasted varieties provide a key resource for identification of DNA sequence variants for use in genomics-assisted breeding of field pea. In conclusion, the present study has substantially increased the transcriptome resources that are available for use in varietal improvement of this important grain legume species.

Availability of supporting data

The data sets supporting the results of this article are included within the article and its additional files. The sequence data has been deposited at DDBJ/EMBL/GenBank under the SRA accessions - SRR1910794, SRR1910804-SRR1910826, SRR1913075, SRR1913256 and SRR1913731-SRR1913750. Transcriptome Shotgun Assembly project has been deposited at DDBJ/EMBL/GenBank under the accessions, Kaspa - GCMF00000000, GCMG00000000, GCMH00000000, GCMI00000000, GCMJ00000000, GCMK00000000 and GCML00000000 and Parafield - GCKA00000000, GCMM00000000, GCMN00000000, GCMO00000000, GCMP00000000 and GCMQ00000000. The version described in this paper is the first version. BioProject ID for Kaspa transcriptome is PRJNA277074 and for Parafield transcriptome is PRJNA277076.

Additional files

Additional file 1: Summary of sequencing outputs. This file contains a summary of the sequencing outcomes for each tissue-specific library of both varieties, Kaspa and Parafield (total number of raw reads from both sequencing platforms, number of reads used for assembly). (XLSX 11 kb)

Additional file 2: Bioinformatic annotation (BLAST) of Kaspa and Parafield contigs against the nr database. The information includes the BLASTX results obtained as a result of comparison of Kaspa and Parafield contigs against the nr database at a threshold E value < 10^{-10}. (XLSX 10037 kb)

Additional file 3: BLASTX alignment details. (A) This figure shows the E-value distribution of significant hits for Kaspa and Parafield transcriptomes; (B) This figure represents the BLASTX top-hit species distribution of contig annotations. (PDF 34 kb)

Additional file 4: Bioinformatic annotation (BLASTN) of Kaspa and Parafield contigs against the nt database and the pea chloroplast genome. The information includes the BLASTN results obtained as a result of comparison of Kaspa and Parafield contigs against the nt database and the pea chloroplast genome at a threshold E value < 10^{-10}. (XLSX 10124 kb)

Additional file 5: (A) Bioinformatic annotation (BLASTN) of Kaspa contigs against the chickpea, *M. truncatula* and soybean CDS. The information includes the BLASTN results obtained as a result of comparison of Kaspa contigs against the chickpea, *M. truncatula* and soybean CDSs at a threshold E value < 10^{-10}. (B) Bioinformatic annotation (BLASTN) of Kaspa contigs against the chickpea, *M. truncatula* and soybean genome: The information includes the BLASTN results obtained as a result of comparison of Kaspa contigs against the chickpea, *M. truncatula* and soybean genomes at a threshold E value < 10^{-10}. (C) Bioinformatic annotation (BLASTN) of Parafield contigs against the chickpea, *M. truncatula* and soybean CDS: The information includes the BLASTN results obtained as a result of comparison of Parafield contigs against the chickpea, *M. truncatula* and soybean CDSs at a threshold E value < 10^{-10}. (D) Bioinformatic annotation (BLASTN) of Parafield contigs against the chickpea, *M. truncatula* and soybean genome: The information includes the BLASTN results obtained as a result of comparison of Parafield contigs against the chickpea, *M. truncatula* and soybean genomes at a threshold E value < 10^{-10}. (ZIP 52127 kb)

Additional file 6: Pathway assignment based on KEGG. This file summarises the properties of Kaspa and Parafield contigs annotated to the reference pathways in KEGG database, and the contigs with KO annotations. (XLSX 915 kb)

Additional file 7: (A). Bioinformatic comparison (BLASTN) of Kaspa contigs against the pea transcriptome datasets. The information

includes the BLASTN results obtained as a result of comparison of Kaspa contigs against four other transcriptome datasets of pea at a threshold E value < 10^{-10}. (B) Bioinformatic comparison (BLASTN) of Parafield contigs against the pea transcriptome datasets: The information includes the BLASTN results obtained as a result of comparison of Parafield contigs against four other transcriptome datasets of pea at a threshold E value < 10^{-10}. (ZIP 31071 kb)

Additional file 8: ORF prediction of the unannotated contigs. (A) This file contains information on ORFs prediction for the unannotated Kaspa-derived contigs. (B) This file contains information on ORFs prediction for the unannotated Parafield-derived contigs. (ZIP 7244 kb)

Additional file 9: Transcript expression for each tissue. (A) This file details contigs expression for the individual (tissue-specific) libraries after alignment to the Kaspa assembly. (B) This file details contigs expression for the individual (tissue-specific) libraries after alignment to the Parafield assembly. (ZIP 20813 kb)

Additional file 10: Annotation of tissue-specific contigs. This file contains information on the annotation of some of the tissue-specific contigs. (XLSX 10 kb)

Abbreviations

ABC transporters: ATP-binding cassette transporters; bp: Base pair; BLAST: Basic Local Alignment Search Tool; BZIP: Basic leucine zipper; CAB: Chlorophyll-a/b binding; cDNA: Complementary DNA; CDS: Coding DNA sequences; DNA: Deoxyribonucleic acid; Gb: Giga base pair; KEGG: Kyoto Encyclopedia of Genes and Genomes; Mb: Mega base pair; mRNA: messenger RNA; NCBI: National Center for Biotechnology Information; ncRNAs: non-coding RNAs; ORF: Open reading frame; PVP: Polyvinylpyrrolidone; RNA: Ribonucleic acid; RNA-Seq: RNA sequencing technology; RUBISCO: Ribulose bisphosphate carboxylase-oxygenase; UTRs: Untranslated regions; W/V: Weight/volume.

Competing Interests

The authors declare that they have no competing interests.

Author's contributions

SS prepared plant materials, performed RNA extraction, library preparation, sequencing, data analysis and drafted the manuscript. TIS contributed to data interpretation and assisted in drafting of the manuscript. SK and NOIC co-conceptualised the project, participated in experimental design, contributed to data interpretation and assisted in drafting of the manuscript. PK co-conceptualised the project. JWF co-conceptualised the project and assisted in drafting of the manuscript. All authors read and approved the final manuscript.

Acknowledgements

This work was supported by funding from the Victorian Department of Economic Development, Jobs, Transport and Resources and the Grains Research and Development Council, Australia.

Author details

[1]Department of Economic Development, Jobs, Transport and Resources, Biosciences Research Division, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, VIC 3083, Australia. [2]Department of Economic Development, Jobs, Transport and Resources, Biosciences Research Division, Grains Innovation Park, Horsham, VIC 3401, Australia. [3]School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3086, Australia.

Received: 12 January 2015 Accepted: 15 May 2015
Published online: 16 August 2015

References

1. FAOSTAT. [http://faostat3.fao.org] Accessed 8 January 2015.
2. Macas J, Neumann P, Navrátilová A. Repetitive DNA in the pea (Pisum sativum L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. BMC Genomics. 2007;8:427.
3. Murray MG, Cuellar RE, Thompson WF. DNA sequence organisation in the pea genome. Biochemistry. 1978;17:5781–90.
4. *Medicago Truncatula* Genome Project. [http://jcvi.org/medicago/] Accessed 8 January 2015.
5. Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, et al. Genome structure of the legume, *Lotus japonicus*. DNA Res. 2008;15:227–39.
6. Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, et al. Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nat Biotech. 2013;31:240–6.
7. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev. 2009;10:57–63.
8. Marguerat S, Wilhelm BT, Bähler J. Next-generation sequencing: applications beyond genomes. Biochem Soc Trans. 2008;36:1091–6.
9. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. PLOS ONE. 2014;9:1–13. doi:10.1371/journal.pone.0078644.
10. Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, et al. A gene expression map of *Arabidopsis thaliana* development. Nat Genet. 2005;37:501–6. doi:10.1038/ng1543.
11. Jiao Y, Tausta SL, Gandotra N, Sun N, Liu T, et al. A transcriptome atlas of rice cell types uncovers cellular, functional and developmental hierarchies. Nat Genet. 2009;41:258–63. doi:10.1038/ng.282.
12. Moreau S, Verdenaud M, Ott T, Letort S, de Billy F, Niebel A, et al. Transcription reprogramming during root nodule development in *Medicago truncatula*. PLoS ONE. 2011;6, e16463. doi:10.1371/journal.pone.0016463.
13. Verdier J, Dessaint F, Schneider C, Abirached-Darmency M. A combined histology and transcriptome analysis unravels novel questions on *Medicago truncatula* seed coat. J Exp Bot. 2013;64:459–70. doi:10.1093/jxb/ers304.
14. Verdier J, Torres-Jerez I, Wang M, Andriankaja A, Allen SN, He J, et al. Establishment of the *Lotus japonicus* gene expression atlas (LjGEA) and its use to explore legume seed maturation. Plant J. 2013;74:351–62.
15. Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, et al. Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE. 2013;8, e61005. doi:10.1371/journal.pone.0061005.
16. Sekhon RS, Briskine R, Hirsch CN, Myers CL, Springer NM, et al. Correction: Maize gene atlas developed by RNA sequencing and comparative evaluation of transcriptomes based on RNA sequencing and microarrays. PLoS ONE. 2014;9:10. doi:10.1371/annotation/0444d495-9231-4097-abe0-4750b9045971.
17. Farrell JD, Byrne S, Paina C, Asp T. *De novo* assembly of the perennial ryegrass transcriptome using an RNA-Seq strategy. PLoS ONE. 2014;9, e103567. doi:10.1371/journal.pone.0103567.
18. Severin AJ, Woody JL, Bolon YT, Joseph B, Diers BW, Farmer AD, et al. RNA-Seq atlas of *Glycine max*: a guide to the soybean transcriptome. BMC Plant Biol. 2010;14:160.
19. Zhu QH, Stephen S, Kazan K, Jin G, Fan L, Taylor J, et al. Characterization of the defense transcriptome responsive to *Fusarium oxysporum*-infection in Arabidopsis using RNA-seq. Gene. 2013;512:259–66.
20. Gao L, Tu ZJ, Millett BP, Bradeen JM. Insights into organ-specific pathogen defense responses in plants: RNA-seq analysis of potato tuber-*Phytophthora infestans* interactions. BMC Genomics. 2013;14:340.
21. Raney JA, Reynolds DJ, Elzinga DB, Page J, Udall JA, Jellen EN, et al. Transcriptome analysis of drought induced stress in *Chenopodium quinoa*. Am J Plant Sci. 2014;5:338–57. doi:10.4236/ajps.2014.53047.
22. Belamkar V, Weeks NT, Bharti AK, Farmer AD, Graham MA, Cannon SB. Comprehensive characterization and RNA-Seq profiling of the HD-Zip transcription factor family in soybean (*Glycine max*) during dehydration and salt stress. BMC Genomics. 2014;15:950. doi:10.1186/1471-2164-15-950.
23. Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM. Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. BMC Genomics. 2011;12:227.
24. Kaur S, Pembleton L, Cogan NOI, Savin KW, Leonforte T, Paull J, et al. Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics. 2012;13:104.
25. Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, et al. Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. BMC Genomics. 2014;15:126.
26. Leonforte A, Sudheesh S, Cogan NOI, Salisbury PA, Nicolas ME, Materne M, et al. SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (Pisum sativum L.). BMC Plant Biol. 2013;13:161.
27. Sudheesh S, Lombardi M, Leonforte A, Cogan NOI, Materne M, Forster JW, et al. Consensus genetic map construction for field pea (*Pisum sativum* L.),

trait dissection of biotic and abiotic stress tolerance and development of a diagnostic marker for the *er1* powdery mildew resistance gene. Plant Mol Biol Rep. 2014, doi: 10.1007/s11105-014-0837-7.
28. Sindhu A, Ramsay L, Sanderson L-A, Stonehouse R, Li R, Condie J, et al. Gene-based SNP discovery and genetic mapping in pea. Theor Appl Genet. 2014;127:2225–41.
29. Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, et al. Genome sequence of the palaeopolyploid soybean. Nature. 2012;463:178–83.
30. Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, et al. Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nat Biotech. 2011;30:83–9.
31. Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, et al. *De novo* transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. Nat Protoc. 2013;8:1494–512.
32. Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, et al. SOAPdenovo-Trans: *De novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics. 2014;30:1660–6. doi:10.1093/bioinformatics/btu077.
33. Huang, Madan A. CAP3: A DNA sequence assembly program. Genome Res. 1999;9:868–77.
34. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res. 1997;25:3389–402. doi:10.1093/nar/25.17.3389.
35. International Chickpea Genetics and Genomics Consortium (ICGGC). [http://www.icrisat.org/gt-bt/ICGGC/GenomeManuscript.htm] Accessed 8 January 2015.
36. PlantGDB. [http://www.plantgdb.org/XGDB/phplib/download.php?GDB=Gm] Accessed 8 January 2015.
37. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9:357–9.
38. Rice P, Longden I, Bleasby A. EMBOSS: the European Molecular Biology Open Software Suite. Trends Genet. 2000;16:276–7.
39. KEGG Automatic Annotation Server (KAAS). [http:/www.genome.jp/tools/kaas/] Accessed 8 January 2015.
40. Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. KAAS: an automatic genome annotation and pathway reconstruction server. Nucleic Acids Res. 2007;35:182–5.
41. Cool Season Food Legume Genome Database. [http://www.coolseasonfoodlegume.org/sativum_unigene_v2] Accessed 27 February 2015.
42. Graham PH, Vance CP. Legumes: importance and constraints to greater use. Plant Physiol. 2003;131:872–7.
43. Shakoor N, Nair R, Crasta O, Morris G, Feltus A, Kresovich S. A *Sorghum bicolor* expression atlas reveals dynamic genotype-specific expression profiles for vegetative tissues of grain, sweet and bioenergy sorghums. BMC Plant Biol. 2014;14:35–48.
44. Wang Z, Fang B, Chen J, Zhang X, Luo Z, Huang L, et al. *De novo* assembly and characterization of root transcriptome using Illumina paired-end sequencing and development of cSSR markers in sweetpotato (*Ipomoea batatas*). BMC Genomics. 2010;11:726.
45. Martin AR, Costa HA, Lappalainen T, Henn BM, Kidd JM, Yee M-C, et al. Transcriptome sequencing from diverse human populations reveals differentiated regulatory architecture. PLoS Genet. 2014;10, e1004549. doi:10.1371/journal.pgen.1004549.
46. Liang C, Liu X, Yiu S-M, Lim BL. *De novo* assembly and characterization of *Camelina sativa* transcriptome by paired-end sequencing. BMC Genomics. 2013;14:146.
47. Zhu L, Zhang Y, Guo W, Xu XJ, Wang Q. *De Novo* assembly and characterization of *Sophora japonica* transcriptome using RNA-Seq. Biomed Res Int. 2014;2014:750961.
48. Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, et al. Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nat Biotechnol. 2011;29:644–52.
49. Li D, Deng Z, Qin B, Liu X, Men Z. De novo assembly and characterization of bark transcriptome using Illumina sequencing and development of EST-SSR markers in rubber tree (Hevea brasiliensis Muell. Arg.). BMC Genomics. 2012;13:192.
50. Doyle JJ, Luckow MA. The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiol. 2003;131:900–10.
51. Steele KP, Wojciechowski MF. Phylogenetic analyses of tribes Trifolieae and Vicieae, based on sequences of the plastid gene, *mat K* (Papilionoideae: Leguminosae). In: Klitgaard BB, Bruneau A, editors. Advances in Legume Systematics, part 10. Kew: Higher Level Systematics, Royal Botanic Gardens; 2003. p. 355–70.
52. Crawford JE, Guelbeogo WM, Sanou A, Traoré A, Vernick KD, Sagnon NF, et al. *De novo* transcriptome sequencing in *Anopheles funestus* using Illumina RNA-seq technology. PLoS one. 2010;5, e14202.
53. Zolala J, Farsi M, Gordan HR, Mahmoodnia M. Producing a high scopolamine hairy root clone in *Hyoscyamus muticus* through transformation by *Agrobacterium rhizogenes*. J Agric Sci Technol. 2007;9:327–39.
54. Gamas P, de Carvalho NF, Lescure N, Cullimore JV. Use of a subtractive hybridisation approach to identify new *Medicago truncatula* genes induced during nodule development. Mol Plant Microbe In. 1996;9:233–42.
55. Libault M, Farmer A, Joshi T, Takahashi K, Langley RJ, Franklin LD, et al. An integrated transcriptome atlas of the crop model *Glycine max*, and its use in comparative analyses in plants. Plant J. 2010;63:86–99.
56. Garg R, Patel RK, Jhanwar S, Priya P, Bhattacharjee A, Yadav G, et al. Gene discovery and tissue-specific transcriptome analysis in chickpea with massively parallel pyrosequencing and web resource development. Plant Physiol. 2011;56:1661–78.
57. Borisov AY, Madsen LH, Tsyganov VE, Umehara Y, Voroshilova VA, Batagov AO, et al. The *Sym35* gene required for root nodule development in pea is an ortholog of Nin from *Lotus japonicus*. Plant Physiol. 2003;131:1009–17.
58. Schauser L, Roussis A, Stiller J, Stougaard J. A plant regulator controlling development of symbiotic root nodules. Nature. 1999;402:191–5.
59. Taylor SA, Hofer JM, Murfet IC, Sollinger JD, Singer SR, Knox MR, et al. PROLIFERATING INFLORESCENCE MERISTEM, a MADS-box gene that regulates floral meristem identity in pea. Plant Physiol. 2002;129:1150–9.
60. Jakoby M, Weisshaar B, Dröge-Laser W, Vicente-Carbajosa J, Tiedemann J, Kroj T, et al. bZIP transcription factors in Arabidopsis. Trends Plant Sci. 2002;7:106–11.

CHAPTER 5

Construction of an integrated linkage map and trait dissection for bacterial blight resistance in field pea (*Pisum sativum* L.)

5.1 Chapter preface

This chapter describes the generation of a genetic linkage map for the Kaspa x PBA Oura RIL-based mapping population based on the use of SSR and SNP markers. A consensus map was obtained by combining this map with three previously published bi-parental maps, leading eventually to the generation of a comprehensive integrated structure for pea. Trait dissection of resistance to bacterial blight caused by *P. syringae* pv. *syringae* identified 4 QTL-containing regions on the Kaspa x Parafield genetic map, and 2 regions on the Kaspa x PBA Oura genetic map. Resistance to *P. syringae* pv. *pisi* was controlled by a single genomic region on both the Kaspa x Parafield and Kaspa x PBA Oura maps.

This chapter is presented in published format.

5.2 Publication details

Title: Construction of an integrated linkage map and trait dissection for bacterial blight resistance in field pea (*Pisum sativum* L.)

Journal Details: Molecular Breeding. 2015, 35:185. doi:10.1007/s11032-015-0376-4.

Stage of publication: Published

Authors: Shimna Sudheesh, Matthew Rodda, Peter Kennedy, Preeti Verma, Antonio Leonforte, Noel OI Cogan, Michael Materne, John W Forster, Sukhjiwan Kaur

5.3 Statement of contributions of joint authorship

Shimna Sudheesh carried out the DNA extraction, SSR and SNP genotyping of the mapping population, linkage and integrated map construction, phenotypic screening of the RILs, QTL analysis and manuscript preparation.

Matthew Rodda and Peter Kennedy performed phenotypic screening of the RILs.

Preeti Verma assisted in DNA extraction of the mapping population RILs.

Antonio Leonforte and Michael Materne co-conceptualised the project.

Noel OI Cogan and John W Forster co-conceptualised the project and assisted in drafting the manuscript.

Sukhjiwan Kaur co-conceptualised the project, participated in experimental design, contributed to data interpretation and assisted in drafting the manuscript.

5.4 Statement from the co-author confirming the authorship contribution of the PhD candidate

“As co-author of the manuscript ‘Sudheesh *et al.* (2015) Construction of an integrated linkage map and trait dissection for bacterial blight resistance in field pea (*Pisum sativum* L.). Molecular Breeding. 35:185’, I confirm that Shimna Sudheesh has made the following contributions,

- DNA extraction of the mapping population
- Genotyping of the mapping population
- Bi-parental and integrated map generation
- Phenotypic screening of the RILs
- QTL analysis
- Interpretation of findings

- Writing the manuscript, critical appraisal of content and response to reviewers”

Dr Sukhjiwan Kaur

Date: 15/09/2015

Mol Breeding (2015)35:185
DOI 10.1007/s11032-015-0376-4

Construction of an integrated linkage map and trait dissection for bacterial blight resistance in field pea (*Pisum sativum* L.)

Shimna Sudheesh · Matthew Rodda · Peter Kennedy · Preeti Verma · Antonio Leonforte · Noel O. I. Cogan · Michael Materne · John W. Forster · Sukhjiwan Kaur

Received: 27 May 2015 / Accepted: 27 August 2015

Abstract Field pea (*Pisum sativum* L.) is a grain legume crop that is cultivated for either human or livestock consumption. Development of varieties adapted to damaging abiotic and biotic stresses is a major objective for field pea breeding. Bacterial blight is a serious disease caused by the pathogenic agents *Pseudomonas syringae* pv. *syringae* and *Pseudomonas syringae* pv. *pisi*. A recombinant inbred line (RIL) genetic mapping population was generated by crossing the susceptible genotype Kaspa to the resistant genotype PBA Oura. Previously described sets of single-nucleotide polymorphism and simple sequence repeat markers were assigned to a genetic linkage map of the Kaspa × PBA Oura population, which contained 358 markers spanning 1070 cM with an average marker density of 1 locus per 3.0 cM. Combination with multiple previously published maps (including that of Kaspa × Parafield) subsequently generated an integrated structure with much higher marker density of 1 locus per 0.85 cM. The Kaspa × PBA Oura and Kaspa × Parafield RILs were screened at the seedling stage for resistance to both pathovars. Totals of four and two QTLs for resistance to infection by *P. syringae* pv. *syringae* were detected in the Kaspa × Parafield and Kaspa × PBA Oura populations, respectively. A single common genomic region associated with *P. syringae* pv. *pisi* resistance was identified in both mapping populations. To integrate information on bacterial blight resistance from various QTL mapping studies, the relevant regions were extrapolated on to the integrated map through use of common flanking markers. The resources generated in this study will support map enhancement, identification of marker-trait associations, genomics-assisted breeding, map-based gene isolation and comparative genetics.

Keywords Grain legume · Simple sequence repeat · Single-nucleotide polymorphism · Bacterial disease resistance · Molecular breeding

Electronic supplementary material The online version of this article (doi:10.1007/s11032-015-0376-4) contains supplementary material, which is available to authorized users.

S. Sudheesh · P. Verma · N. O. I. Cogan · J. W. Forster (✉) · S. Kaur
Biosciences Research Division, AgriBio, Centre for AgriBioscience, 5 Ring Road, Bundoora, VIC 3083, Australia
e-mail: john.forster@ecodev.vic.gov.au

S. Sudheesh · J. W. Forster
School of Applied Systems Biology, La Trobe University, Bundoora, VIC 3086, Australia

M. Rodda · P. Kennedy · A. Leonforte · M. Materne
Biosciences Research, Grains Innovation Park, Horsham, VIC 3401, Australia

Introduction

Field pea (*Pisum sativum* L.) is an annual cool-season legume crop with a global annual production of 11 million tonnes from 6.4 million hectares (FAOSTAT

Published online: 04 September 2015

2014). In many agricultural systems, legumes are complementary to cereals due to contributions from biological nitrogen fixation. Moreover, seeds of legumes contain valuable macro- and micro-nutrients, including proteins, providing a highly nutritious food resource (Duranti 2006). Production of field pea is constrained by the impact of a number of biotic and abiotic stresses, requiring improvement of levels of protection against major fungal, bacterial and viral diseases to increase yield and quality of the crop (Allen and Lenne 1998).

Bacterial blight, caused by the pathogenic agents *P. syringae* pv. *syringae* and *P. syringae* pv. *pisi*, is a sporadic and damaging disease of field pea. Incidence of bacterial blight has been reported in most of the pea-growing areas of the world. Prevalence of each pathogen varies between regions (depending on the predominant cultivar within that region), and they may occur in the field separately, or in combination (Martín-Sanz et al. 2011; Richardson and Hollaway 2011). A total of eight different pathogenic races of *P. syringae* pv. *pisi* have been described, based on their differential interaction with specific pea genotypes (Taylor et al. 1989; Martín-Sanz et al. 2011). In Australia, race 3 (64 %), followed by race 6 (31 %), is the most common of these races (Hollaway et al. 2007). *P. syringae* pv. *pisi* is the prevalent bacterial blight pathogen in Australian production zones, but the incidence of disease caused by *P. syringae* pv. *syringae* has increased since the introduction of semi-leafless cultivars (such as Kaspa, Excell and Snow-peak) in the south-eastern region (Richardson and Hollaway 2011).

Development of disease-resistant cultivars is the most effective method for control of bacterial blight, as chemical control is ineffective against both pathogens in the infected fields. Resistance to *P. syringae* pv. *pisi* is conferred by a specific gene-for-gene interaction mechanism, with six putative resistance–avirulence (R–A) gene pairs (Bevan et al. 1995). In contrast, resistance to *P. syringae* pv. *syringae* under controlled and field conditions exhibits continuous variation, suggesting contributions by a number of genes with lower magnitudes of effect, and leading to quantitative inheritance (Martín-Sanz et al. 2012). QTL studies for resistance to *P. syringae* pv. *pisi* in pea have identified several of these R genes (*Ppi1* [*R1*] for race 1, *Ppi2* [*R2*] for race 2, *Ppi3* [*R3*] for race 3 and *Ppi4* [*R4*] for race 4), which have been assigned to different genetic map locations (*Ppi1* on linkage group [LG] VI, *Ppi2* on LG VII, *Ppi3* and *Ppi4* probably on LG II [Hunter et al. 2001]). For *P. syringae* pv. *pisi* race 6, no source of resistance has so far been identified. In contrast, relatively few QTL studies have been conducted in order to identify mechanisms of resistance to bacterial blight caused by *P. syringae* pv. *syringae*, possibly due to limited availability of resistant genetic backgrounds (Fondevilla et al. 2012). A RIL population-based QTL analysis study identified quantitative variation for resistance, through detection of two QTLs accounting for 22 and 8 % of the phenotypic variation (V_p), respectively (Fondevilla et al. 2012).

Molecular genetic tools including high-throughput, robust molecular marker technologies such as SSR and SNP markers, high-density genetic linkage maps and consensus maps (Leonforte et al. 2013; Duarte et al. 2014; Sindhu et al. 2014; Sudheesh et al. 2014), and significant genomic resources such as transcriptome sequencing datasets (Franssen et al. 2011; Kaur et al. 2012; Duarte et al. 2014; Sudheesh et al. 2015) have been developed for pea, with an aim of crop improvement through enhancement of breeding programs. Genetic linkage maps are essential for QTL identification, marker-assisted selection, comparative genomics and map-based cloning (Semagn et al. 2006; McConnell et al. 2010), and also play a crucial role in anchoring physical maps and ordering and orientation of scaffolds during genome assemblies. Integration of multiple individual linkage maps into consensus structures permits saturation of marker content, with enhanced value across a broad range of germplasm sources, and ultimately allowing identification of the genetic factors that control agronomically important traits, including bacterial blight resistance, through QTL mapping and candidate gene identification.

The objectives of the present study were to develop a genetic linkage map for the Kaspa × PBA Oura RIL mapping population using SSR and SNP markers; to develop an integrated map of field pea by combining the Kaspa × PBA Oura map with previously described maps (Leonforte et al. 2013; Sudheesh et al. 2014; Duarte et al. 2014; Sindhu et al. 2014); and to detect QTLs associated with bacterial blight resistance in two RIL populations (Kaspa × Parafield and Kaspa × PBA Oura) that exhibit variation for the trait, using phenotypic data collected from glasshouse-based nursery screens.

Materials and methods

Plant materials

A segregating genetic mapping population was developed by crossing single genotypes from cultivar Kaspa (susceptible to both *P. syringae* pv. *syringae* and *P. syringae* pv. *pisi* race 3) to PBA Oura (moderately resistant to *P. syringae* pv. *syringae* and resistant to *P. syringae* pv. *pisi* race 3) (Richardson and Hollaway 2011) at the premises of Biosciences Research, Horsham, Victoria, Australia, and F_2 progeny were obtained. Single-seed descent was performed for four subsequent generations in the glasshouse, and a total of 110 F_6 RILs were subsequently generated. The plants were maintained in glasshouses under ambient conditions of 22 °C with a 16-/8-h (light/dark) photoperiod. Genomic DNA was extracted from leaves using the DNeasy® 96 Plant Kit (QIAGEN, Hilden, Germany) according to manufacturer's instructions. The DNA was resuspended in MilliQ water and quantified using a Nanodrop spectrometer (Thermo Scientific, Massachusetts, USA), followed by storage at −20 °C.

SSR and SNP genotyping and genetic linkage mapping

SSR (Loridon et al. 2005; Kaur et al. 2012) and SNP (Leonforte et al. 2013) genotyping, followed by genetic linkage mapping and visualisation of the linkage groups of Kaspa × PBA Oura RILs, was performed as described previously (Leonforte et al. 2013; Sudheesh et al. 2014).

Integrated map construction

The Kaspa × PBA Oura genetic linkage map generated in this study was combined with the pre-existing Kaspa × Parafield (Leonforte et al. 2013), Kaspa × Yarrum and Kaspa × ps1771 (Sudheesh et al. 2014) maps to generate an intermediate structure, using MergeMap (Wu et al. 2011). Similarity searches of the DNA sequences underlying map-assigned SNP markers were performed against the transcriptome sequencing data (DDBJ/EMBL/GenBank under the accession GAMJ00000000 and KnowPulse—http://knowpulse.usask.ca/portal/search/sequences) underpinning the SNP markers on integrated maps of Duarte et al. (2014) and Sindhu et al. (2014) with a threshold *E*-value of 10^{-10}. After identification of common sequences, data from all pre-existing maps were combined into a single integrated map structure through use of MergeMap. The linkage groups were visualised by use of MapChart (Voorrips 2002).

Phenotypic evaluation

Bacterial isolates

Bacterial isolates (*P. syringae* pv. *syringae* and *P. syringae* pv. *pisi* race 3) that were used in this study were obtained from Biosciences Research, Horsham, Victoria, Australia. The identities of the bacterial isolates were confirmed (using PCR), and bacterial isolates were maintained as described by Rodda et al. (2015).

Disease assessment

Screening for bacterial blight resistance in both mapping populations was performed under controlled environmental conditions at Biosciences Research, Horsham, Victoria, Australia. The experimental design was a randomised split block with three replicates of three plants each from each RIL-derived line in a pot. Two types of inoculation procedure (spray inoculation or stab inoculation) were performed on three- to four-week-old seedlings. *P. syringae* inoculum preparation and the inoculation procedures were performed as described previously [spray inoculation method—Rodda et al. (2015); stab inoculation method—Hollaway and Bretag (1995)]. Kaspa × Parafield RIL-derived plants were stab-inoculated with *P. syringae* pv. *pisi* race 3. Spray inoculation was performed for the other three combinations (Kaspa × PBA Oura and Kaspa × Parafield with *P. syringae* pv. *syringae*; Kaspa × PBA Oura with *P. syringae* pv. *pisi* race 3).

Following spray inoculation of *P. syringae* pv. *syringae* on both mapping populations, disease incidence scores were recorded as percentages of affected plant tissue. Following stab inoculation of the Kaspa × Parafield RIL population with *P. syringae* pv. *pisi* race 3, disease incidences were recorded on a numerical scale, with score 1—inoculation point dried, score 2—lesion growth arrested and score 3—water-soaked lesion. Means for disease scores were

calculated, and if the value exceeded 2, a susceptible rating was assigned. In contrast, if the value was less than or equal to 2, a resistant rating was given. Following spray inoculation of the Kaspa × PBA Oura mapping population with *P. syringae* pv. *pisi* race 3, plants were scored on a scale of 1–4 (1—fully resistant; 2—moderately resistant; 3—moderately susceptible; and 4—fully susceptible). Mean scores for each individual from the mapping populations were used to construct frequency distribution histograms in order to determine mode of inheritance for each trait.

QTL analysis and identification of sequences associated with flanking genetic markers

The Kaspa × PBA Oura genetic linkage map that was developed in this study and Kaspa × Parafield genetic linkage map from Leonforte et al. (2013) were used for QTL analysis. QTL detection was performed as described by Sudheesh et al. (2014) using QTL Cartographer v 2.5 (Wang et al. 2012). QTLs identified in the current study and from a previous study (Fondevilla et al. 2012) were placed on the integrated map on the basis of common flanking markers.

BLASTN analysis of the sequences underlying SNP markers flanking the QTL-containing intervals was performed against the coding sequences (CDS) of *M. truncatula*, v3.5 (http://jcvi.org/medicago/), with a threshold *E*-value of 10^{-10}. The sequences of SNP loci flanking the QTL-containing intervals were BLASTN analysed against the *M. truncatula* genome v3.5 (http://jcvi.org/medicago/), to identify genomic regions containing putative candidate genes.

Results

Genotyping and genetic linkage mapping

A total of 240 genomic DNA derived and EST-SSRs from different sources were screened to detect polymorphism between the parental lines of the Kaspa × PBA Oura mapping population, identifying 50 informative marker assays (40 genomic DNA derived and 10 EST-SSRs), of which 40 were successfully screened on 110 RILs. SNP genotyping was performed with 768 markers, of which 350 (46 %) revealed clear polymorphism and were hence processed for genetic linkage map construction. From the selected markers, a total of 23, including 9 SSRs and 14 SNPs, were excluded from linkage map construction due to skewed segregation ratios. The remaining 367 markers (31 SSRs and 336 SNPs) (Supplementary file 1) were used for linkage mapping, of which 358 (31 SSRs and 327 SNPs) were assigned to ten LGs (Supplementary files 2 and 3). The map spanned a total length of 1070 cM, with an average marker density of 1 locus per 3.0 cM (Supplementary file 4). The number of markers assigned to LGs varied from 5 (Ps VI) to 69 (Ps VII), with an average of 36 markers per LG. The identity and orientation of LGs were determined by comparison with the *M. truncatula* genome, as well as from previously map-assigned SSRs as anchoring markers. The largest numbers of SSR marker loci that were common between maps were on Ps IV and Ps VII, and the smallest was on Ps VI.

Comparative analysis of the four population-specific genetic maps (Kaspa × PBA Oura, Kaspa × Parafield, Kaspa × Yarrum and Kaspa × ps1771) revealed high levels of co-linearity, although specific map distances were not always comparable. A total of 123 markers were common across all four maps. The highest number of common markers (308) was between the Kaspa × Yarrum and Kaspa × ps1771 maps, followed by the comparison between the Kaspa × PBA Oura and Kaspa × ps1771 maps, with 281 markers. The lowest number of common markers was between the Kaspa × Yarrum and Kaspa × Parafield maps. Comparison of the common markers revealed co-linearity in locus order for Ps I, Ps III, Ps IV, Ps V and Ps VII. Alternatively, Ps II and Ps VI exhibited different genomic configurations, although co-linearity of gene order was maintained within the respective translocated segments. LGs containing large (≥20 cM) intervals devoid of any markers were observed for all four maps. The total number of gaps varied, the highest number (13) being for the Kaspa × Yarrum map, and the lowest (7) for the Kaspa × ps1771 and Kaspa × PBA Oura maps.

A composite structure was generated by merging information from the four maps, which were based on differing numbers of individual markers (358 in Kaspa × PBA Oura, 429 in Kaspa × Yarrum, 452 in Kaspa × ps1771 and 458 in Kaspa × Parafield). The composite map consisted of 782 markers (694 SNPs and 88 SSRs) which were assembled into seven LGs (Supplementary file 3), with a cumulative map length of 2601 cM. The lengths of LGs ranged from

293 cM (Ps VI) to 430 cM (Ps IV), with an average density of one marker per 3.4 cM. In some instances (37 in total, including three SSRs), the same marker assay detected multiple loci, which were consequently considered as unique markers. A total of 145 markers (19 %) were common to all four maps, while 234 (30 %) were unique to single populations (Kaspa × Parafield—128; Kaspa × Yarrum—63; Kaspa × ps1771—28 and Kaspa × PBA Oura—15), the remaining 548 (70 %) providing bridging loci between two or more maps. Comparison of the composite map with individual maps was performed to verify marker order in LGs, revealing high co-linearity in locus order (Supplementary file 5), apart from Ps II and Ps VI.

The DNA sequences underlying the 768 SNP loci were used for sequence similarity searches (Supplementary file 6) against pea transcriptome sequences underpinning those SNP markers assigned to previously published integrated maps. This process identified 203 bridging loci between the field pea composite map and the integrated maps. A comprehensive integrated structure was generated by merger of all available maps, permitting placement of 2857 markers on seven LGs spanning 2427 cM (Fig. 1; Table 1). Map density varied from 0.67 cM (Ps VI) to 1.09 cM (Ps II), with an average of 0.85 cM. Of the seven LGs, Ps VII contained the largest number of markers (525), followed by Ps III and Ps IV with 459, 392 markers, respectively, while Ps VI contained only 360 markers (Supplementary file 3). Some intervals devoid of any markers (>10 cM) were observed: single gaps on each of Ps I and Ps VII; three gaps on each of Ps II, Ps III and Ps IV, mainly towards the distal end; and a couple of gaps in the middle of each LG. Markers from each of the maps contributed unique loci to the integrated structure, and detailed analysis indicated that some of these loci were localised to specific regions. For instance, eight markers spanning 22 cM on the upper part of Ps V were derived from the composite structure of population-specific genetic maps. Marker order was well conserved in comparison with LGs of the individual maps, although several inversions and local rearrangements were observed.

Disease assessment and QTL detection

Significant differences between disease scores of whole plants following infection with *P. syringae* pv. *syringae* were observed between RILs of each population (Supplementary file 7). Frequency distribution patterns obtained from both populations indicated the presence of multiple genes contributing to disease resistance (Fig. 2). Evidence was also obtained for transgressive segregation patterns, as the phenotypic scores of some individual lines were located outside the range of the parental genotypes (being both more and less susceptible). The locations and magnitudes of effect for each QTL were estimated using both simple interval mapping (SIM) and composite interval mapping (CIM). On the Kaspa × PBA Oura map, two QTLs were detected (on Ps III and Ps VII), accounting for *circa* (c.) 15 % (Psy_KO1) and 11 % (Psy_KO2) of V_p, respectively (Fig. 3; Table 2). On the Kaspa × Parafield map, CIM detected four QTLs, two (Psy_KP2 and 3) of which were located on Ps III, accounting for c. 13 and 23 % of V_p, respectively (Table 2). The other two QTLs were detected on Ps II (Psy_KP4) and Ps VII (Psy_KP1), accounting for c. 9 and 8 % of V_p, respectively (Fig. 3, Table 2). For Psy_KP2, resistance determinants were derived from the susceptible parent (Kaspa), while the genetic factors at the other QTL regions were associated with the moderately resistant parent (Parafield).

For the resistance to *P. syringae* pv. *pisi* race 3, only a single genomic region with moderate magnitude of effect was detected, located on Ps III of both the Kaspa × Parafield and Kaspa × PBA Oura maps, with maximum logarithm of odds (LOD) scores of 9.5 and 12.1, respectively, using CIM. The Ps III-located QTL accounted for c. 26 % (Kaspa × Parafield) and 38 % (Kaspa × PBA Oura) of V_p (Fig. 3, Table 2). The LOD peak for the QTL in both instances coincided with the marker SNP_100000801, further supporting the inference that a single common genomic region is responsible for conferring resistance.

All QTLs identified in the current study, as well those from a previous study (Fondevilla et al. 2012), were extrapolated onto the integrated map. This process obtained seven loci distributed across three LGs, with number of loci per LG varying from 1 to 4 (Fig. 3). A single correspondence was observed with the previous study, based on the location of the PsBB2_Psy from the Kaspa × Parafield population.

Annotation of the flanking markers

Comparison of DNA sequences associated with markers linked to the *P. syringae* pv. *syringae*


Ps I
Ps II
Ps III
Ps IV
Ps V
Ps VI
Ps VII

◀**Fig. 1** Integrated map of field pea, with markers shown on the *right* of LGs, and distances between markers indicated in cM on the *left*. For presentation purposes, only one of a set of co-located genetic markers is shown on the map

resistance QTLs directly identified *M. truncatula* CDS with functional annotations (supplementary file 8) such as serine/threonine-protein kinase, monothiol glutaredoxin-S15, polyadenylate-binding protein 2, cysteine proteinase 3 and serine hydroxymethyltransferase gene products, some of which have been reported to be involved in disease resistance mechanisms of plants. Similarly, the sequences associated with SNP markers flanking the *P. syringae* pv. *pisi* race 3 resistance QTL-containing intervals included annotations (supplementary file 8) as ascorbate peroxidase, rhodanese-like and caffeic acid O-methyltransferase gene products.

Moreover, regions of the *M. truncatula* genome near and within the QTL-containing intervals were examined for candidate gene presence. Multiple genes located in the vicinity of QTLs (Psy_KP3, Psy_KO1, Psy_KO 2 and Ppi3) on Ps III (Mt III) and Ps VII (Mt IV and Mt VIII) were annotated as being involved in disease resistance (Supplementary file 9).

Discussion

Genotyping and linkage mapping

In the recent past, several field pea genetic linkage maps have been developed, mostly through the use of SNP and SSR marker technologies (Leonforte et al. 2013; Duarte et al. 2014; Sindhu et al. 2014; Sudheesh et al. 2014). The Kaspa × PBA Oura linkage map constructed in the present study exhibits an average marker density of 1 per 3.0 cM and a cumulative length of 1070 cM. This value is comparatively smaller than those of the maps from Leonforte et al. (2013) and Sudheesh et al. (2014), possibly due to the effects of a lower number of map-assigned markers in the present study (358). Comparisons with previously constructed maps (Leonforte et al. 2013; Sudheesh et al. 2014) revealed highly conserved marker orders, although the corresponding distances were not always in similar proportion.

Individual genetic linkage maps exhibit limitations such as genotype-specific chromosomal rearrangements (duplications, translocations, inversions), restricted access to population-specific information and limitations of marker content and density (Yu et al. 2010). Combination of data from multiple linkage maps into a single integrated structure can address these shortcomings and facilitate the comparison of QTL regions identified in various genetic backgrounds. In the present study, this objective was accomplished in a two-step process, initially by merger of four population-specific genetic linkage maps and then by combination with two other integrated maps. An integrated map has a higher level of relevance for positioning of the order of markers than to determine absolute distances between markers. The recombination frequencies of maps from different mapping populations are not fully consistent, although the MergeMap product provided a more accurate description of marker order than a linear map derived using approximations based on recombination values (Close et al. 2009). Moreover, the consensus map based on

Table 1 Characteristics of the composite and integrated maps of field pea

Predicted pea LG	Number of mapped markers		Map length (cM)		Average marker density	
	Composite map	Integrated map	Composite map	Integrated map	Composite map	Integrated map
Ps I	105	375	344	317	3.3	0.85
Ps II	97	368	392	401	4.0	1.09
Ps III	134	459	425	399	3.2	0.87
Ps IV	132	392	430	400	3.3	1.02
Ps V	87	378	310	287	3.6	0.76
Ps VI	87	360	293	242	3.4	0.67
Ps VII	140	525	407	381	2.9	0.73
Total	782	2857	2601	2427	3.4	0.85

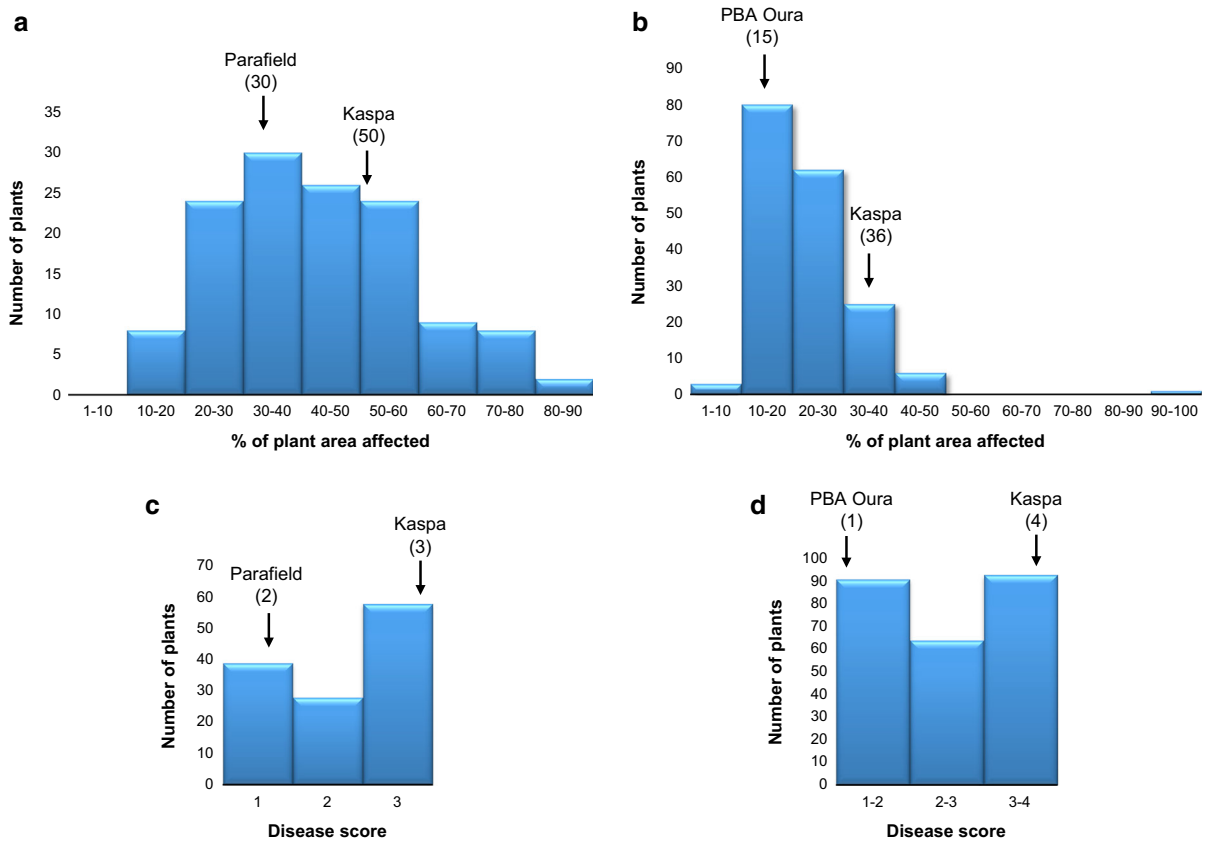

Fig. 2 Frequency distribution histograms generated from bacterial blight phenotyping using **a** percentage of infected area—*P. syringae* pv. *syringae* for Kaspa × Parafield; **b** percentage of infected area—*P. syringae* pv. *syringae* for Kaspa × PBA Oura; **c** disease scores—*P. syringae* pv. *pisi* race 3 for Kaspa × Parafield; and **d** disease scores—*P. syringae* pv. *pisi* race 3 for Kaspa × PBA Oura mapping populations

merging different maps is simply one of many possible non-conflicting linear representations of the consensus DAGs. MergeMap software, which was used for this purpose, has been used in construction of consensus maps in crops such as common bean (*Phaseolus vulgaris* L.) (Galeano et al. 2012), peanut (*Arachis hypogaea* L.) (Gautami et al. 2012) and faba bean (*Vicia faba* L.) (Satovic et al. 2013). The integrated map contained 829 more markers than a previously constructed consensus map (Sudheesh et al. 2014). Although high levels of co-linearity were observed between the two maps, addition of new markers generated a small number of fine-scale differences.

Disease assessment, QTL detection and annotation of the flanking markers

The present study used two different methodologies for phenotypic assessment of disease resistance. Previously, the stab inoculation method has been used for bacterial inoculation (Hollaway and Bretag 1995), but is a slow and laborious process and not suitable for use with *P. syringae* pv. *syringae*. The spray inoculation method more closely reflects the natural means of infection than stab inoculation, and is also more efficient and reproducible (Rodda et al. 2015). Only minor differences in disease symptoms due to different inoculation methods were observed.

In the present study, phenotypic screening for bacterial blight resistance was conducted under controlled conditions in the glasshouse using a single *P. syringae* pv. *syringae* isolate. A significant challenge arises from the presence of partial resistance effects which require quantitative assessment, in contrast to major R genes that generally give rise to qualitative effects on the phenotype. Only limited studies to date have been conducted to analyse the resistance of pea to *P. syringae* pv. *syringae*, under both controlled and

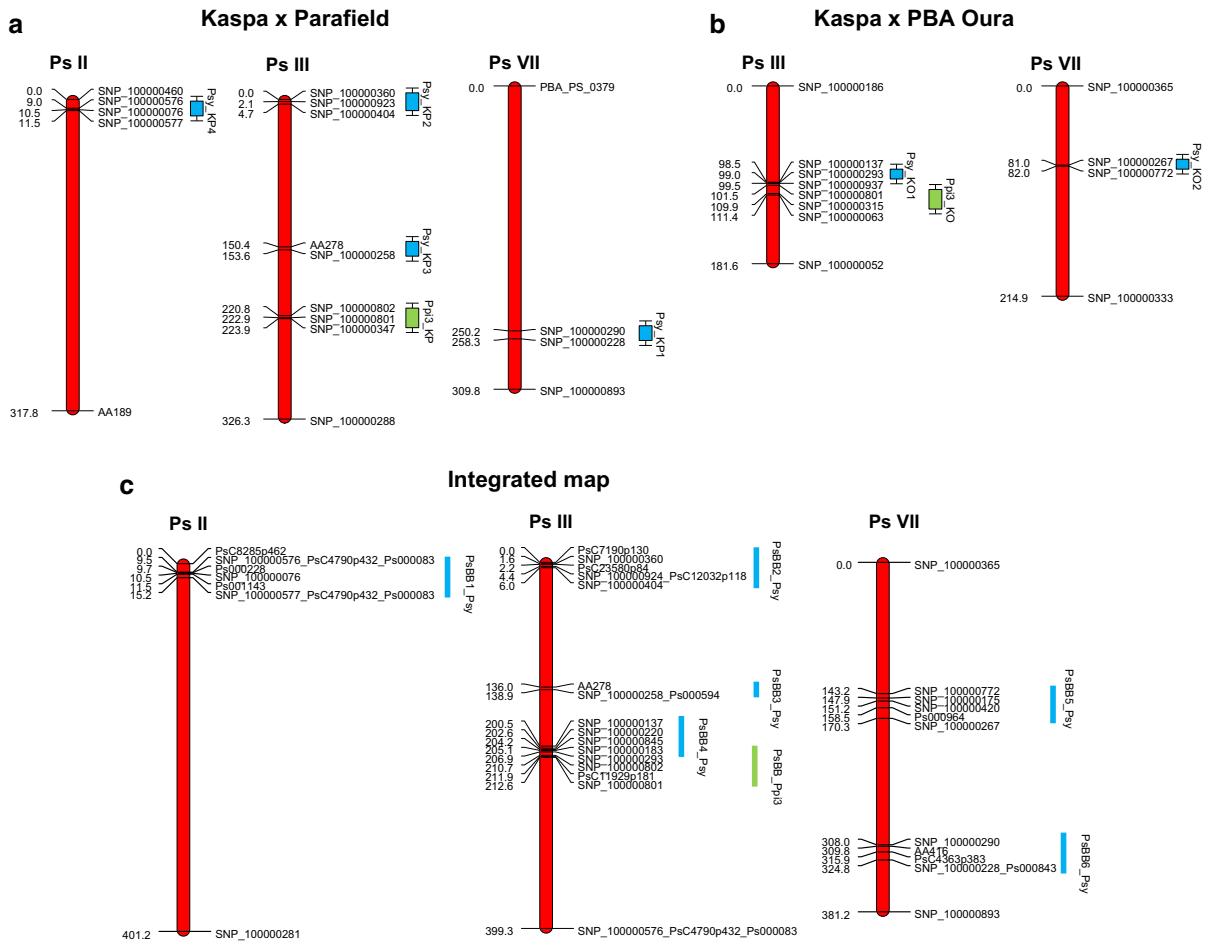

Fig. 3 Localisation of QTLs associated with resistance to *P. syringae* pv. *syringae* and *P. syringae* pv. *pisi* race 3 on the **a** Kaspa × Parafield-derived genetic map; **b** Kaspa × PBA Oura-derived genetic map; and **c** integrated map. The name is provided at the *top* of each LG. Distances of the markers (cM) are shown to the *left*, and names of markers are shown to the *right* side of LGs. For presentation purposes, only selected markers are shown on the map

field conditions (Martín-Sanz et al. 2011; Richardson and Hollaway 2011). A continuous frequency distribution of resistance scores in response to *P. syringae* pv. *syringae* infection was observed for both RIL populations in the present study, suggesting quantitative inheritance. An earlier study in pea also reported a continuous distribution, supporting this hypothesis (Fondevilla et al. 2012).

CIM analysis detected totals of 4 and 2 QTLs for resistance to bacterial blight caused by *P. syringae* pv. *syringae* in the Kaspa × Parafield and Kaspa × PBA Oura mapping populations, respectively. This discrepancy may be due to the differences of genetic background. For the Kaspa × Parafield population, both parents contributed alleles for resistance. The inheritance of such favourable alleles from susceptible parents has been reported previously (Tar'an et al. 2004). For the Kaspa × PBA Oura RIL population, only the resistant parent (PBA Oura) contributed alleles for bacterial blight resistance. It is possible that such differences may have arisen from gene interaction effects, especially due to a predominant effect of genetic contributions of a more resistant parent (PBA Oura).

The parental lines of the two mapping populations exhibited relatively narrow differences in resistance level, and this may have contributed to the identification of QTLs accounting for low-to-moderate proportions of V_p, rather than major gene effects, which would generally be manifested through highly

Table 2 Summary of information on QTLs for bacterial blight resistance

Mapping population	QTL name	LG	Flanking markers	Position cM	LOD threshold	Maximum LOD threshold	% Phenotypic variance	Additive effect
Kaspa × Parafield	Psy_KP1	Ps VII	SNP_100000290	250–258	3.5	3.6	8	−4.86
			SNP_100000228					
	Psy_KP2	Ps III	SNP_100000360	0–5	3.5	5.4	13	6.18
			SNP_100000923					
			SNP_100000404					
	Psy_KP3	Ps III	AA278	150–154	3.5	5.7	23	−8.17
			SNP_100000258					
	Psy_KP4	Ps II	SNP_100000576	9–12	3.5	3.7	9	−5.18
			SNP_100000076					
			SNP_100000577					
	Ppi3_KP	Ps III	SNP_100000802	220–223	3.3	9.5	26	−0.29
			SNP_100000801					
			SNP_100000347					
Kaspa × PBA Oura	Psy_KO1	Ps III	SNP_100000137	98–99	2.8	4.2	15	−3.41
			SNP_100000293					
			SNP_100000937					
	Psy_KO2	Ps VII	SNP_100000267	81–82	2.8	3.4	11	−2.88
			SNP_100000772					
	Ppi3_KO	Ps III	SNP_100000801	101–111	2.5	12.1	38	−0.31
			SNP_100000315					
			SNP_100000063					

Psy, *P. syringae* pv. *syringae*; Ppi3, *P. syringae* pv. *pisi* race 3

divergent performance of parental genotypes. In both mapping populations, transgressive lines with higher resistance than the parents were observed. This observation indicates the potential to select such lines, which may provide a novel source of resistance in breeding programs.

QTL analysis for *P. syringae* pv. *pisi* resistance revealed a single locus of moderate effect on the genetic map of the Kaspa × Parafield mapping population. A similar result was obtained for the Kaspa × PBA Oura population. The detailed genetic map structure and marker content vary between the two populations, but the presence of a common flanking marker strongly suggests that the same genomic region is involved. Identification of the same genomic region in both mapping populations supports the view that differences between inoculation methodologies did not contribute to variability of the disease development in the RILs. Differing proportions of V_p accounted for the Ppi3 QTLs detected from the two mapping populations may be due to a number of contributory factors, including differences in genetic background between the two populations; difference in levels of resistance directly conferred by the causal gene(s) that contribute the QTLs, possibly due to independent allelic effects; and methodological differences of phenotypic screening and assessments. In support of these possibilities, there is no commonality of genetic background between the two resistant parents, Parafield and PBA Oura.

Previous studies to evaluate responses to *P. syringae* pv. *pisi* revealed that race-specific resistance under controlled conditions for races 2, 3 and 4 was also observed under field conditions (Martín-Sanz et al. 2012). This outcome suggests that the corresponding resistance genes (*R2*, *R3* and *R4*) are effective under genuine field conditions. Moreover, the gene combination (*R2* + *R3* + *R4*) provided a more effective resistance than a single gene in controlling most *P. syringae* pv. *pisi* races, including

the newly identified race 8, but was ineffective against race 6 (Elvira-Recuenco et al. 2003; Martín-Sanz et al. 2012). The major class of *R* genes are dominant in nature (De Ronde et al. 2014), and so highly suitable for use in marker-assisted selection (Tiwari and Singh 2012) and gene pyramiding. Previous studies (Martín-Sanz et al. 2012; Rodda et al. 2015) showed that the outcomes of phenotyping screening conducted under controlled conditions were positively correlated with field-based disease symptoms, and glasshouse-based leaf inoculation provides a useful method for the prediction of response under field conditions. For example, consistent resistance ratings were observed for the field pea cultivar PBA Percy under both glasshouse and field conditions while infected with *P. syringae* pv. *syringae* (Rodda et al. 2015).

For comparative purposes, the bacterial blight resistance QTLs were placed on the integrated map through use of common flanking markers. The incorporation of information from multiple independent QTL mapping studies provided an enhanced assessment of the genetic basis for resistance. Previous literature had reported the map positions of two QTLs determining resistance to *P. syringae* pv. *syringae* (Fondevilla et al. 2012). However, only one of these QTLs shared linked markers in common with the present study, revealing co-localisation with PsBB2_Psy on the integrated map. Identification of common QTLs between different segregating populations provides increased confidence in prospects for marker-assisted breeding, as well as allowing determination of a set of candidate markers for further enrichment of bacterial blight resistance QTL-containing regions on population-specific genetic maps in future. The high levels of concordance observed between marker locations in the integrated map and those of the population-specific genetic linkage maps confirm the value of the approach.

Functional annotation of sequences underpinning the genetic markers that flank the relevant QTL-containing intervals successfully identified genes associated with plant defence mechanisms in field pea. Serine/threonine kinases (SNP_100000290, Psy_KP1 region) play a role in recognition of pathogen-derived signal molecules and also different signalling levels in the context of non-race-specific elicitation, gene-for-gene interactions, or resistance to virulent pathogens in a number of different systems including tobacco, tomato, alfalfa, parsley, rice and *Arabidopsis thaliana* (Romeis 2001). Plant cysteine proteases (SNP_100000076, Psy_KP4 region) are proteolytic enzymes which are closely regulated during physiological functions (growth, development and accumulation–mobilization of storage proteins) and defence roles (Martínez et al. 2012).

Similarity searches of sequences associated with *P. syringae* pv. *pisi* race 3 QTL region-flanking markers against *M. truncatula* CDS facilitated identification of candidate genes such as plant ascorbate peroxidase (SNP_100000801, SNP_100000802 and SNP_100000347) and caffeic acid O-methyltransferase (SNP_100000063). Elevated expression of enzymes such as ascorbate peroxidase is also observed during defence mechanism of plants against a broad range of pathogens. Caffeic acid O-methyltransferase performs a key role in the biochemical pathway leading to lignin subunit (monolignol) synthesis (Ma and Xu 2008; Caverzan et al. 2012). During pathogen attack, defence reactions such as lignin production are activated, as a well-characterised mechanism of plant cell wall reinforcement against pathogen entry (Shimada et al. 2000).

The comparative analysis of the sequences underpinning the SNP markers flanking the QTL intervals (on Ps III and Ps VII) with the *M. truncatula* genome revealed the presence of multiple disease resistance genes within those intervals. A high level of conserved synteny was observed between the pea genetic map and the *M. truncatula* genome (Leonforte et al. 2013). It is therefore possible that the identified QTL regions may be conserved between these species and that the genes identified within the QTL-containing regions may be plausible candidates, although additional studies will be required for validation.

Implications for field pea breeding programs

The present study has developed valuable resources for pea linkage mapping, especially the integrated map which will enhance future mapping analysis studies and marker implementation in pea breeding programs. Important sources of bacterial blight resistance in field pea were identified, and the genetic markers flanking the QTL-containing regions identified in this study can be used for the development of linked and diagnostic polymorphisms for marker-assisted selection (MAS) of resistant cultivars, based on introgression of QTL-containing genomic regions from donor to recipient

germplasm. Moreover, as QTLs for resistance to both *P. syringae* pv. *syringae* and *P. syringae* pv. *pisi* race 3 are co-located on Ps III of Kaspa × PBA Oura, this region is an important target for improvement of bacterial blight resistance and provides the basis for co-selection in genomics-assisted breeding practice. Pea breeding programs that aim to develop bacterial blight resistance cultivars should target a combination of race-specific and non-specific resistance, in order to provide an optimised genetic background for protection against the disease.

Acknowledgments This work was supported by funding from the Victorian Department of Economic Development, Jobs, Transport and Resources and the Grains Research and Development Council, Australia.

References

Allen DJ, Lenne JM (1998) The pathology of food and pasture legumes. CAB International, Wallingford. ISBN 0-85199-166-1

Bevan JR, Taylor JD, Crute IR, Hunter PJ, Vivian A (1995) Genetics of specific resistance in pea (*Pisum sativum*) cultivars to seven races of *Pseudomonas syringae* pv. *pisi*. Plant Pathol 44:98–108

Caverzan A, Passaia G, Rosa SB, Ribeiro CW, Lazzarotto F, Margis-Pinheiro M (2012) Plant responses to stresses: role of ascorbate peroxidase in the antioxidant protection. Genet Mol Biol 35:1011–1019

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N et al (2009) Development and implementation of high-throughput SNP genotyping in barley. BMC Genom 10:582. doi:10.1186/1471-2164-10-582

De Ronde D, Butterbach P, Kormelink R (2014) Dominant resistance against plant viruses. Front Plant Sci 5:307. doi:10.3389/fpls.2014.00307

Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, Lavaud C, Lejeune-Hénaut I, Martinant JP, Pichon JP, Pilet-Nayel ML, Boutet G (2014) Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. BMC Genom 15:126

Duranti M (2006) Grain legume proteins and nutraceutical properties. Fitoterapia 77:67–82

Elvira-Recuenco M, Bevan JR, Taylor JD (2003) Differential responses to pea bacterial blight in stems, leaves and pods under glasshouse and field conditions. Eur J Plant Pathol 109:555–564

FAOSTAT (2014). http://faostat3.fao.org/home/E. Accessed 3 April 2015

Fondevilla S, Martín-Sanz A, Satovic Z, Fernández-Romero M, Rubiales D, Caminero C (2012) Identification of quantitative trait loci involved in resistance to *Pseudomonas syringae* pv. *syringae* in pea (*Pisum sativum* L.). Euphytica 186:805–812

Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM (2011) Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. BMC Genom 12:227

Galeano CH, Cortés AJ, Fernandez AC, Soler A, Franco-Herrera N et al (2012) Gene-based single nucleotide polymorphism markers for genetic and association mapping in common bean. BMC Genet 13:48

Gautami B, Foncéka D, Pandey MK, Moretzsohn MC, Sujay V et al (2012) An international reference consensus genetic map with 897 marker loci based on 11 mapping populations for tetraploid groundnut (*Arachis hypogaea* L.). PLoS ONE 7:e41213

Hollaway GJ, Bretag TW (1995) Occurrence and distribution of races of *P. syringae* pv. *pisi* in Australia and their specificity towards various field pea (*Pisum sativum*) cultivars. Aust J Exp Agr 35:629–632

Hollaway GJ, Bretag TW, Price TV (2007) The epidemiology and management of bacterial blight (*Pseudomonas syringae* pv. *pisi*) of field pea (*Pisum sativum*) in Australia: a review. Aust J Agr Res 58:1086–1099

Hunter PJ, Ellis N, Taylor JD (2001) Association of dominant loci for resistance to *Pseudomonas syringae* pv. *pisi* with linkage groups II, VI and VII of *Pisum sativum*. Theor Appl Genet 103:129–135

Kaur S, Pembleton L, Cogan NOI, Savin KW, Leonforte T, Paull J, Materne M, Forster JW (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genom 13:104

KnowPulse (2015). http://knowpulse.usask.ca/portal/search/sequences. Accessed 3 April 2015

Leonforte A, Sudheesh S, Cogan NOI, Salisbury PA, Nicolas ME, Materne M, Forster JW, Kaur S (2013) SNP marker discovery, linkage map construction and identification of QTLs for enhanced salinity tolerance in field pea (*Pisum sativum* L.). BMC Plant Biol 13:161

Loridon K, McPhee K, Morin J, Dubreuil P, Pilet-Nayel ML, Aubert G, Rameau C, Baranger A, Coyne C, Lejeune-Hènaut I, Burstin J (2005) Microsatellite marker polymorphism and mapping in pea (*Pisum sativum* L.). Theor Appl Genet 111:1022–1031

Ma QH, Xu Y (2008) Characterization of a caffeic acid 3-*O*-methyltransferase from wheat and its function in lignin biosynthesis. Biochimie 90:515–524

Martínez M, Cambra I, González-Melendi P, Santamaría ME, Díaz I (2012) C1A cysteine-proteases and their inhibitors in plants. Physiol Plant 145:85–94. doi:10.1111/j.1399-3054.2012.01569.x

Martín-Sanz A, Palomo JL, Pérez de la Vega M, Caminero C (2011) Identification of pathovars and races of *Pseudomonas syringae*, the main causal agent of bacterial diseases in pea in North-Central Spain, and the search for resistance. Eur J Plant Pathol 129:57–69

Martín-Sanz A, Pérez de la Vega M, Caminero C (2012) Resistance to *Pseudomonas syringae* in a collection of pea germplasm under field and controlled conditions. Plant Pathol 61:375–387

McConnell M, Mamidi S, Lee R, Chikara S, Rossi M, Papa R, 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. doi:10.1007/s00122-010-1375-9

Medicago Truncatula Genome Project. http://jcvi.org/medicago/. Accessed 3 April 2015

Richardson HJ, Hollaway GJ (2011) Bacterial blight caused by *Pseudomonas syringae* pv. *syringae* shown to be an important disease of field pea in south eastern Australia. Australas Plant Path 40:260–268

Rodda MS, Kant P, Lindbeck KD, Gnanasambandam A, Hollaway GJ (2015) A high-throughput glasshouse based screening method to evaluate bacterial blight resistance in field pea (*Pisum sativum*). Aust Plant Pathol. doi:10.1007/s13313-015-0374-3

Romeis T (2001) Protein kinases in the plant defence response. Curr Opin Plant Biol 4:407–414

Satovic Z, Avila CM, Cruz-Izquierdo S, Díaz-Ruíz R, García-Ruíz GM, Palomino C, Gutiérrez N, Vitale S, Ocaña-Moral S, Gutiérrez MV, Cubero JI, Torres AM (2013) A reference consensus genetic map for molecular markers and economically important traits in faba bean (*Vicia faba* L.). BMC Genom 14:932. doi:10.1186/1471-2164-14-932

Semagn K, Bjørnstad A, Ndjiondjop MN (2006) Principles, requirements and prospects of genetic mapping in plants. Afr J Biotechnol 5:2569–2587

Shimada N, Akashi T, Aoki T, Ayabe S (2000) Induction of isoflavonoid pathway in the model legume *Lotus japonicus*: molecular characterization of enzymes involved in phytoalexin biosynthesis. Plant Sci 160:37–47

Sindhu A, Ramsay L, Sanderson L-A, Stonehouse R, Li R, Condie J, Shunmugam ASK, Liu Y, Jha AB, Diapari M, Burstin J, Aubert G, Tar'an B, Bett KE, Warkentin TD, Sharpe AG (2014) Gene-based SNP discovery and genetic mapping in pea. Theor Appl Genet 127:2225–2241

Sudheesh S, Lombardi M, Leonforte A, Cogan NOI, Materne M, Forster JW, Kaur S (2014) Consensus genetic map construction for field pea (*Pisum sativum* L.), trait dissection of biotic and abiotic stress tolerance and development of a diagnostic marker for the *er1* powdery mildew resistance gene. Plant Mol Biol Rep. doi:10.1007/s11105-014-0837-7

Sudheesh S, Sawbridge TI, Cogan NOI, Kennedy P, Forster JW, Kaur S (2015) *De novo* assembly and characterisation of the field pea transcriptome using RNA-Seq. BMC Genom 16:611. doi:10.1186/s12864-015-1815-7

Tar'an B, Warkentin T, Somers DJ, Miranda D, Vandenberg A, Blade S, Woods S, Bing D (2004) Identification of quantitative trait loci for grain yield, seed protein concentration and maturity in field pea (*Pisum sativum* L). Euphytica 136:297–306

Taylor JD, Bevan JR, Crute IR, Reader SL (1989) Genetic relationship between races of *Pseudomonas syringae* pv. *pisi* and cultivars of *Pisum sativum*. Plant Pathol 38:364–375

Tiwari JK, Singh BP (2012) Marker-assisted selection for virus resistance in potato: options and challenges. Potato J 39:101–117

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

Wang S, Basten CJ, Zeng ZB (2012) Windows QTL Cartographer 2.5. Department of Statistics, North Carolina State University, Raleigh, NC. http://statgen.ncsu.edu/qtlcart/WQTLCart.htm

Wu Y, Close TJ, Lonardi S (2011) Accurate construction of consensus genetic maps via integer linear programming. IEEE/ACM Trans Comput Biol Bioinform 8:381–394. doi:10.1109/TCBB.2010.35

Yu J, Kohel RJ, Smith CW (2010) The construction of a tetraploid cotton genome wide comprehensive reference map. Genomics 95:230–240

CHAPTER 6

General discussion

6.1 Background and context of the research

Field pea is one of the oldest domesticated legume crops (Zohary *et al*. 2012). Pea currently ranks among the most important legumes, major production zones being in Canada, Russia, China, France, Australia, India and the United States (FAOSTAT 2014). Field peas are mainly grown for grain production. However, some varieties are also increasingly used for green manure, forage or hay. Field peas contain high levels of protein (22-25%), which are usually low in cereal grains (Zohary *et al*. 2012), and hence are valuable for protein provision to livestock.

One of the major challenges for field pea cultivation is adaptation to abiotic and biotic stresses. Major abiotic stresses include extremes of soil moisture (drought or water-logging), temperature extremes (heat stress or frost), nutrient deficiencies and toxicity due to salinity and boron (Dita *et al*. 2006). Soil salinity and B toxicity are very common in Australia, and field peas are severely restricted in their plant growth under such conditions, such that the combined effects cause a significant growth reduction (Grieve and Poss 2000). Field pea is also infested by insects, arachnids and nematodes, and infected by pathogens such as bacteria, fungus, viruses and nematodes (Allen and Lenne 1998). Diseases caused by these pathogens reduce grain quality through blemishes, as well as reduced seed size, and are the main limiting factor to expansion of field pea production. Powdery mildew in field peas is caused mainly by the biotrophic fungus *E. pisi* (Ondřej *et al*. 2005; Attanayake *et al*. 2010). Powdery mildew occurs sporadically and causes severe damage to pea cultivation, causing seed yield loss and quality effects. Bacterial blight, caused by the pathogens *P. syringae* pv. *pisi* and *P. syringae* pv. *syringae* (Richardson

and Hollaway 2011), is another serious and widespread disease of field peas, especially in the southern regions of Australia, with varying severity from crop to crop and between seasons, leading to severe epidemics which can result in crop failure. Control of these key diseases through development of resistant varieties is essential to maintain field pea productivity. Conventional breeding has delivered a large number of new field pea varieties, but is a slow and laborious process. In addition, it is often difficult to accurately assess plants for complex phenotypic traits or for those that require a specific environmental challenge, for example resistance to individual diseases. A combination of conventional and modern, genomics-based breeding practices will enhance this process.

The mechanisms of tolerance/resistance in field pea to many abiotic and biotic stresses are currently poorly understood, especially for multigenic traits. There has been a requirement for development of molecular genetic markers and linkage maps in field pea, followed by QTL analysis and candidate gene selection, to allow improvement of field pea germplasm through marker-assisted selection. SNP markers are one of the abundant source of genomic variation, providing polymorphisms for the construction of linkage maps. Compared to other types of markers, SNP markers provide the lowest cost and highest rate of throughput for genotypic analysis (Rafaski 2002). Several suites of SNP marker assays have been previously developed and successfully used for genotyping of mapping populations and genetic resources collections of pea (Deulvot *et al*. 2010; Duarte *et al*. 2014; Sindhu *et al.* 2014). However, additional SNP markers are still required for use in field pea molecular breeding programs.

Such challenges may be overcome through use of novel strategies and technologies. One of these is the generation of transcriptome data, which represents the complete set of transcripts that are present in cells or tissues at different developmental stages and physiological conditions. Genomic resources for field pea were previously developed through transcriptome sequencing using second-generation technologies, mainly in order to develop molecular genetic markers (Franssen *et al*. 2011; Kaur *et al*. 2012; Duarte *et*

al. 2014; Sindhu *et al.* 2014). However, decreases in cost and advances in throughput based on RNA-Seq, and increased efficiency of data analysis, now provide unprecedented opportunities for the production of enriched genomic resources for field pea. RNA-Seq permits comprehensive characterisation of the whole transcriptome of an organism, as well as providing a resource for development of SSR and SNP markers.

6.2 Overview of the research

The work presented in this thesis delivered a large suite of SNP markers, four detailed genetic linkage maps, and a high-density integrated reference map of field pea. As described in Chapter 1, the initial major objective of the project was to investigate aspects of bacterial blight resistance, and to identify genetic markers linked to resistance genes or QTLs. However, the generation of multiple linkage maps also provided an opportunity to dissect other key traits, such as soil salinity tolerance, B toxicity tolerance and powdery mildew resistance. As the final part of this study, a comprehensive field pea transcriptome data set was generated, providing a resource for new rounds of genetic marker discovery and candidate gene identification.

In the first instance, EST-derived SNP markers were developed for construction of comprehensive linkage maps. From a total of 36,188 variant nucleotide positions detected through *in silico* analysis, 768 were finally selected for high-throughput genotyping of four RIL populations. A total of 705 SNP markers (91.7%), successfully detected segregating polymorphisms, demonstrating that the pre-screening process had been highly efficient.

In the second area of research, genetic linkage maps were developed for the Kaspa x Parafield, Kaspa x ps1771, Kaspa x Yarrum and Kaspa x PBA Oura RIL populations. The parental genotypes of these populations differ for a range of important agronomic traits, allowing parallel dissection of multiple characters. SSR- and SNP-based genotyping enabled the generation of Kaspa x Parafield map containing 458 markers and a cumulative

total map length of 1,916 cM, Kaspa x ps1771 map containing 451 loci across 1545 cM, Kaspa x Yarrum map containing 428 loci across 1,910 cM and Kaspa x PBA Oura map containing 358 markers across 1,070 cM. Alignment of the maps was achieved through two strategies: comparison of sequences underpinning the map-assigned markers with the *M. truncatula* genome, and use of bridging SSR markers. By combining these trait-specific bi-parental maps with previously published consensus maps (Duarte *et al*. 2014; Sindhu *et al.* 2014), a comprehensive integrated structure was obtained, containing 2,857 markers on 7 LGs spanning 2,427 cM, with a high marker density of 1 locus per 0.85 cM. This integrated reference map of pea provides an important resource for detailed future studies of genome structure, trait architecture and candidate gene identification.

Genetic map construction permitted a detailed analysis of conserved syntenic relationships between the genomes of field pea and other legume species, namely *M. truncatula*, chickpea*, L. japonicus*, soybean and pigeon pea. Sequence information available from loci on the Kaspa x Parafield linkage map was used, revealing a highest number of matches (301:97%) with chickpea, followed by *M. truncatula* with 292 (94%) matches. The syntenic relationships related each of field pea LGs Ps II, Ps IV, Ps V, and Ps VII to chickpea pseudomolecules Ca4, Ca7, Ca3 and Ca6, respectively. Substantial blocks of conservation were observed between most *M. truncatula* chromosomes and field pea LGs. Mt 5, 1, 3, and 7 exhibited synteny and colinearity with pea linkage groups Ps I, Ps II, Ps III and Ps V respectively, indicating close evolutionary relationships between the two species. In contrast, large syntenic blocks spanning entire PsLGs were absent from the comparisons with the *L. japonicus*, soybean and pigeon pea genomes.

The third major area of research focused on improved understanding of key agronomic traits in field pea, namely salinity stress tolerance, B toxicity tolerance, powdery mildew resistance and bacterial blight resistance. In each case, either molecular markers flanking the QTL containing regions or diagnostic markers were developed.

Analysis of the Kaspa x Parafield RIL population revealed a quantitative basis for seedling-induced salinity tolerance in pea, identifying two QTL loci, with each accounted for moderate proportions of the phenotypic variation. Conserved synteny analysis with *M. truncatula* genome identified plausible candidate genes associated with saline stress tolerance. Genetic markers linked to these QTLs could be implemented for a MAS program, in order to enrich for the favourable alleles in selected progeny.

Trait dissection of B toxicity tolerance in the Kaspa x ps1771 RIL population identified a major QTL on Ps VI. QTL analysis for powdery mildew resistance was performed for both the Kaspa x ps1771 and Kaspa x Yarrum RIL populations, identifying a single common genomic region on Ps VI, which was inferred to correspond to the previously described *er1* gene. The powdery mildew resistance QTL coincided with the B tolerance locus, enabling co-selection for these desirable traits. Resequencing of the *PsMLO1* candidate gene from resistant and susceptible genotypes allowed the design and validation of a putative diagnostic marker for powdery mildew resistance.

Trait dissection of resistance to bacterial blight caused by *P. syringae* pv. *syringae* identified 4 QTL-containing regions on the Kaspa x Parafield genetic map, and 2 regions on the Kaspa x PBA Oura genetic map. In contrast, the mechanism of resistance to *P. syringae* pv. *pisi* was controlled by a single genomic region on both the Kaspa x Parafield and Kaspa x PBA Oura maps. An earlier study by Hunter *et al.* (2001) suggested a probable location of *Ppi3* [*R3*] on LG II, based on the possible linkage of *R3* to the anthocyanin gene, however, a direct comparative analysis between the maps could not be performed due to the lack of common markers. The co-localisation of the QTL for *P. syringae* pv. *pisi* resistance with one of the *P. syringae* pv. *syringae* resistance QTLs on the Kaspa x PBA Oura mapping population provides the basis for co-selection in genomics-assisted breeding practice. The conserved synteny analysis between the *M. truncatula* genome and sequences underpinning the SNP markers flanking several of the QTL-containing intervals revealed the presence of multiple disease resistance genes

within those intervals. These regions may be conserved in structure between the species, and therefore may be plausible candidates.

The final research activity aimed to develop a comprehensive transcriptome dataset for field pea. Transcriptome sequencing was performed on 23 cDNA libraries generated from multiple tissues of the field pea genotypes Kaspa and Parafield using Illumina high-throughput sequencing platforms. Both genotypes were selected from cultivars that are extensively used in field pea breeding. A total of 407 and 352 million paired-end reads from the Kaspa and Parafield transcriptomes were *de novo* assembled into 129,282 and 149,272 contigs, respectively. A final set of 126,335 contigs from Kaspa and 145,730 from Parafield were selected as references after filtering of the contigs on the basis of known gene annotations, presence of ORFs, reciprocal matches and degree of coverage. Reciprocal sequence analysis was performed to assess cultivar specificity, which applied to c. 23% of the contigs. Comparison of contig expression on a tissue-specific basis was performed by aligning reads from different libraries to the genotype-specific assemblies, revealing that 87% contigs were expressed in more than one tissue, while others showed distinct expression patterns in specific tissues, so providing unique transcriptome signatures.

In conclusion, the work presented here describes a comprehensive set of genetic and genomics resources for field pea, and has the potential to accelerate continuing efforts to improve field pea productivity and quality through exploitation of existing genetic variation. The transcriptome data generated in this study will prove valuable for future studies of field pea genome structure, including assembly of the gene-containing fraction, and annotation of gene function.

6.3 Future directions

The research described in this thesis provides a number of areas suitable for further investigations. In the case of the segregating mapping populations that were used and linkage maps that were generated, more extensive phenotypic screening of the parental genotypes could identify other divergent agronomic traits suitable for trait dissection. These may potentially include components of yield, seed-specific characters, plant architecture characters, herbicide tolerance, as well as disease-related traits such as resistance to downy mildew, PSbMV and BLRV. Future work should also include the enrichment of current linkage maps with additional genetic markers, including those selected from the integrated reference map. This will permit coverage of larger portions of the pea genome and an increased marker density in the vicinity of target QTLs, allowing more accurate placements, and enhanced value for trait-dissection. For a reference genome assembly process, the contigs and scaffolds must be globally ordered and orientated in order to generate chromosomal pseudomolecules. The high marker density on the integrated map will also assist chromosome-level assembly of whole-genome sequences of field pea in due course, both directly, and through comparative analysis with physical and genetic maps of closely related species, for successful genetic anchoring of the genome sequence assembly.

Application of marker-based selection has particular value for breeding of abiotic stress tolerant or disease resistant varieties, as such traits may be difficult or expensive to assess on the basis of phenotype alone. A detailed understanding of the control of quantitative traits will assist the development of appropriate breeding strategies. As previously described, QTL analysis of salinity tolerance revealed a complex mode of inheritance, such that no genomic region accounted for a predominant component of the phenotypic variation. As a consequence, selection for multiple QTLs representing different components of tolerance may be necessary for MAS implementation. Within the time-

frame of this project, phenotypic analysis of seedling growth stage salinity tolerance was assessed. Additional phenotypic analysis related to other growth stages, such as at germination (Cheng *et al*. 2008; Mano and Takeda 1997; Ma *et al*. 2007) or during reproductive development (Manneh *et al*. 2007; Xue *et al*. 2010; Villalta *et al*. 2007) are likely to be significant for field pea, and therefore require investigation. The previous literature has indicated (Cheng *et al*. 2008; Ma *et al*. 2007; Manneh *et al*. 2007; Xue *et al*. 2010; Villalta *et al*. 2007) significant variation in the degree and timing of salinity-induced growth responses within and between crop species. Consequently, selection for specific salinity-induced growth response QTLs in field pea should be undertaken separately for different growth stages as part of a targeted breeding strategy to pyramid useful genes. As the comparative genomic analysis identified a salt tolerance protein in the equivalent chromosomal region of the *M. truncatula* genome, it is plausible to speculate that variation at an orthologous gene may be responsible for the observed phenotypic variation in field pea, which could be characterised in order to develop candidate gene-based markers.

The powdery mildew resistance marker associated with *PsMLO1* is highly promising for MAS, due to its diagnostic character, and the large magnitude of effect exhibited by the resistance QTL. Diagnostic markers can facilitate selection processes in field pea breeding programs by direct identification of donor genotypes in germplasm collections, rather than requiring specific tests crosses on a case-by-case basis to determine linkage phase with the causal polymorphism. This will help to reduce the duration of the breeding cycle and also the possibility of lost association with the desirable trait due to recombination. Previously, *eIF4E* allele specific markers associated with PSbMV resistance was developed, which proved to be 100% reliable, faster and cost efficient compared to classical virological testing (Dostalova *et al*. 2010). Identification of a single genomic region conferring B toxicity tolerance and powdery mildew resistance ensures that the linked markers can co-select genes for tolerance, in multiple different genetic backgrounds thereby developing progeny with superior trait values. The resistance determinants for

each of the pathovars responsible for bacterial blight differ in terms of complexity of genetic control. The *P. syringae* pv. *pisi*-specific resistance is monogenic, and so MAS is both highly feasible and a beneficial option due to the laborious and expensive nature of the phenotypic assays. In contrast, implementation of markers associated with resistance to *P. syringae* pv. *syringae* will be highly challenging given the complex multigenic nature of the trait. This observation implies that multiple genomic regions, each contributing relatively small proportions of genetic variance, will require individual selection. Environmental factors such as frost, along with colder and wetter conditions, play a major role in expression of bacterial blight disease in field pea, and so must be considered in addition to genetic determinants of resistance. Several studies suggest that protection against *P. syringae* is related both to the presence of specific disease resistance genes and the frost tolerance of pea plants (Elvira-Recuenco *et al*. 2003; Hollaway *et al*. 2007; Martin-Sanz *et al*. 2012a). Pea breeding programs that aim to develop bacterial blight resistance cultivars should target a combination of race-specific and non-specific resistance with joint selection of frost tolerant accessions, in order to provide an optimised genetic background for protection against the disease. The analysis of bacterial blight resistance in the present study also identified transgressive segregants for each trait, implying that superior varieties may be produced by crossing between multiple parental genotypes.

In order to identify candidate genes for bacterial blight resistance, an integrated strategy that involves both functional and positional candidate gene approaches may be applied. The functional candidate gene approach would involve selection of genes related to *P. syringae* resistance from other crop and model species, as reported in the literature, followed by resequencing to identify SNPs, genetic map assignment and assessment of co-location with the pathogen resistance QTLs in field pea. The positional candidate gene approach, in contrast, would exploit the properties of the integrated reference map or known patterns of macrosynteny with *M. truncatula* to identify disease resistance genes,

or members of other gene classes such as those related to plant defence. The identified candidate genes could also be validated based on level of expression (Xu *et al*. 2014). Association of candidate gene sequences with a specific biological function or response would not only facilitates marker development, but also expand knowledge of the biological mechanisms that mediate given phenotypic traits in field pea.

Conventional disease resistance breeding programs are based on: (1) screening of germplasm collections to identify sources of resistance, and characterisation of the resulting phenotypes; (2) study of the mode of inheritance; (3) introgression of the resistance traits in elite cultivars; and (4) assessment of the performance of the new cultivars under pathogen challenge in the field (Pink 2002). For durable disease resistance, not only the major genes but also minor genes are deployed. R genes generally provide high levels of resistance and are relatively easy to manipulate. However, R genes experience limitations due to pathogen race-specificity and lack of durability, due to the process of continuous pathogen evolution. Higher durability may be achieved through a combination of resistance genes, based on incorporation of multiple R genes into single cultivars by 'gene pyramiding' or 'gene stacking' (Marone *et al*. 2013). The use of molecular markers is particularly suitable for gene pyramiding, in particular when different genes with similar phenotypes require introgression into the same genotype. R genes and R gene clusters have been identified in many species, including arabidopsis (Meyers *et al*. 2003) and soybean (Kanazin *et al.* 1996; Graham *et al.* 2002). PCR-based strategies which use degenerate primers to target the conserved domains of NBS-sequences have allowed isolation of candidate resistance genes or resistance gene analogues in several crops including potato (*Solanum tuberosum* L.) (Leister *et al.* 1996), soybean (Kanazin *et al.* 1996), chickpea (Huettel *et al*. 2002) and sorghum (*Sorghum bicolor* [L.] Moench) (Totad *et al.* 2005). Approaches based on degenerate primers to target the conserved domains of resistance genes, or comprehensive surveys based on high-throughput sequencing may assist characterisation of R genes in field pea. Characterisation of field pea

germplasm collections for R gene content and variation will highly facilitate the selection of plants with superior allelic combinations at multiple loci. When combined with precise phenotyping methods, such studies are capable of identifying candidate genes for several resistance loci, and providing valuable sources of closely linked molecular markers for implementation of MAS.

The transcriptome data presented within this thesis could be further exploited as a reference dataset for other studies, in which reads can be aligned to the reference with reduced requirement for sequence depth. In addition, the data could be re-analysed in order to improve the assembly once the pea genome sequence is available. Transcriptome analysis provides an opportunity to investigate plant response regulation and to identify genes involved in disease resistance or stress tolerance mechanisms. Several studies have investigated the changes in gene expression during abiotic (Raney *et al*. 2014; Belamkar *et al*. 2014; Shu *et al*. 2015) and biotic stresses (Goyer *et al*. 2015; Gao *et al*. 2013) in different plant species. To identify both major and minor genes associated with bacterial blight resistance and salt stress tolerance, a gene expression study using RNA-seq could be performed through comparisons between the transcriptomes of resistant/tolerant and susceptible/intolerant cultivars. Such an RNA-seq study would support identification of differentially expressed genes, along with potential causal polymorphisms located in candidate genes. As part of such gene expression analysis studies for disease resistance traits, efforts can be made to identify the specific genes that confer resistance. Such information would permit the design of diagnostic molecular genetic markers that are not vulnerable to decoupled association, as may occur between a linked genetic marker and the target gene locus. The combination of QTL mapping with transcriptome profiling is highly complementary, and capable of identifying both *cis* and *trans*-acting regulatory factors. Such genomic tools will accelerate the identification of diagnostic markers or markers in close linkage to genes for target traits and can be applied

in the MAS process in a rapid and economically feasible manner for the release of new varieties.

QTL analysis performed with a bi-parental genetic mapping population of limited size can detect only the differences between the parent-derived alleles that segregate within a population, and is subject to other constraints, such as underestimation of the magnitude of QTL effects. In order to fully characterise the extent of QTL variation, further studies within multiple genetic backgrounds should be performed. For multigenic traits like yield, grain characteristics, salinity tolerance and bacterial blight resistance, breeders will need to select for varying numbers of genetic markers in different germplasm, including wild relatives of the domesticated crop species. One of the most effective means for identification of marker-trait associations would be to perform a genome-wide association study (GWAS), in which a large number of genetic markers that are evenly distributed across the genome are individually assessed for correlation with trait variation in a complex multi-ancestral population. In crop species, such as rice, GWAS has been used to analyse traits similar to those targeted in the present study. Multi-parent advanced generation inter-cross (MAGIC) populations were used to identify QTLs for bacterial blight resistance and salinity, as well as blast resistance and submergence tolerance and grain quality traits (Bandillo *et al*. 2013). Such studies have also been performed for a pulse crop, chickpea, in order to determine the the genetic basis of tolerance to drought and heat stress (Thudi *et al*. 2014). In pea, association mapping was recently performed to investigate traits such disease resistance, flower colour and different seed characteristics, using a limited number of SNP markers (Cheng *et al*. 2015). The suite of SNP markers that was developed in this project could therefore be used for such a study, but a much higher marker density is anticipated to be required for GWASs.

As the cost of GoldenGate™ oligonucleotide pool assay (OPA) is high, alternative high-throughput SNP discovery approaches will be required to generate a larger number of SNP markers at a higher density throughout the field pea genome. Current technology platforms

such as the Illumina Infinium system would allow analysis of tens of thousands of SNPs in a chip-based assay. However, GBS methods provide the most attractive option, because of the potential to deliver hundreds of thousands of marker loci without any prior sequence information with simplified methodology at the required low cost (Elshire *et al.* 2011). GBS may be performed either by genomic complexity reduction (skim sequencing, restriction endonuclease [RE] digestion followed by adapter ligation, PCR and sequencing) or a whole-genome resequencing approach. Repetitive elements in plants can pose significant challenges for *de novo* assembly, alignment to a reference sequence and sequence comparison for variant discovery. Moreover, many of the identified SNPs could be located within non-genic regions, and hence be of reduced functional value. In contrast, GBS methods based on sequencing of cDNAs generated from expressed mRNAs offer an effective way of targeting coding regions of the genome (Deschamps *et al*. 2012; Harper *et al*. 2012). In whichever form, GBS methods will be necessary in future to permit cost-effective GWASs and genomic selection (GS) breeding strategies in pea.

GS provides an alternative to traditional MAS, with enormous potential to improve selection in breeding programs. GS integrates genome-wide marker data from a training population with phenotypic data and, also the pedigree data from the same population to generate a prediction model. The model outputs the genomic estimated breeding values (GEBVs) for all genotyped individuals within a breeding population for which genotypic information is available, and their phenotypic performance is predicted solely on the basis of that genotypic information (Meuwissen *et al*. 2001). The prediction accuracy depends upon the genetic relationship between the training population and the breeding population, the number of generations that separate them, the type and number of markers used, the accuracy of the phenotyping, and the heritability of the trait(s) (Crossa *et al*. 2014). GS is currently being applied to a range of crops including maize (Crossa *et al*. 2014), wheat (Rutkoski *et al*. 2011) and rice (Spindel *et al*. 2015). The GBS method can be applied in pea for the development of large number of genome-wide SNP markers, and so application

of GS in pea could be anticipated in the near future. The markers developed in this study could be incorporated into the marker set developed through GBS, allowing enhanced accuracy of prediction for the traits described in the present study, as well as GEBVs for other complex traits. As all of the genes contributing to a trait, even those with minor effects, are assessed for GS implementation, this approach is of highest value for complex characters controlled by multiple gene loci, such as some environmental stress tolerances, quantitative disease resistance, and aspects of grain yield and quality. There is much to be gained by combining the strengths of MAS and GS approaches in field pea breeding to reap maximum benefits from declining genotyping prices with minimal phenotyping, in order to enhance the productivity and sustainability of crop production in the face of climate change, evolving pathogen infestation and increasing human demand.

Plant breeding has traditionally exploited the natural allelic variation that is present within germplasm collections, or randomly induced variation generated by treatment with ionising radiation or by chemical mutagenesis. The work described in this thesis has developed and implemented molecular genetic tools to assist and refine this process. However, targeted genome engineering has also emerged as an alternative to both traditional plant breeding and pre-existing transgenesis methods (that generate genetically modified organisms [GMOs]) for improvement of crop plants. Genome editing describes a set of techniques that permit targeted changes to be made to genomic sequences, such as addition, deletion or replacement of nucleotide units at specified locations. Methods for introduction of site-specific double strand DNA breaks into the genome at or near to a target site for modification include four major classes of sequence-specific nucleases (SSNs) - engineered homing endonucleases or meganucleases, zinc finger nucleases (ZFNs) (Urnov *et al*. 2010), transcription activator-like effector nucleases (TALENs) (Joung and Sander 2013), and clustered regularly interspersed short palindromic repeats (CRISPR)/Cas9 reagents (Cong *et al*. 2013; Voytas and Gao 2014). Recently, TALEN technology has been used to modify the three *MLO* homoeologs in bread wheat (*TaMLO*-

A1, *TaMLO-B1* and *TaMLO-D1*) (Wang *et al.* 2014). This modification of all three *TaMLO* homoeologs conferred heritable broad-spectrum resistance to powdery mildew. This technology consequently has the potential to be applied to modification of MLO alleles in other crops to generate disease resistant varieties, such as for the *PsMLO1* gene described in the present study, as well as other target genes as they become available. An essential prerequisite of genome editing is the availability of precise genomic information and knowledge of gene function. For pea, whole-genome sequencing efforts are currently underway, and so the availability of genomic resources will surely facilitate elucidation of the control of important traits, allowing identification of precise target sequences for editing in the near future. However, application of genome editing in pea will require additional advances, as this species is highly recalcitrant to tissue culture-dependent transformation and regeneration processes, which are necessary for deployment of DNA constructs for editing (Svabova *et al.* 2008; Krejci *et al.* 2007). However, genome modification remains an attractive prospect for genetic improvement of field pea, providing the technological barriers can be overcome.

REFERENCES

ABARES (2014) Agricultural commodity statistics 2013. http://www.daff.gov.au/ABARES/Pages/display.aspx?url=http://143.188.17.20/anrdl/DAFFService/display.php?fid=pb_agcstd9abcc0022013_11a.xml. Accessed 29 August 2015.

Allen DJ, Lenne JM (1998) The pathology of food and pasture legumes. CABI, Wallingford, Oxon, UK.

Armstrong E, Bretag T, Ophel-Keller K, Panagiotopoulos K (2001) Strategies to minimise bacterial blight in field pea. www.agric.nsw.gov.au.

Arnold DL, Athey-Pollard A, Gibbon MJ, Taylor JD, Vivian A (1996) Specific oligonucleotide primers for the identification of *P. syringae* pv. *pisi* yield one of two possible DNA fragments by PCR amplification: evidence for phylogenetic divergence. Physiological and Molecular Plant Pathology. 49:233-245 (Erratum: (1997) 51, 213).

Arraouadi S, Chardon F, Huguet T, Aouani ME, Badri M (2011) QTL mapping of morphological traits related to salt tolerance in *Medicago truncatula*. Acta Physiologiae Plantarum. 33:917-929.

Attanayake RN, Glawe DA, McPhee KE, Dugan FM, Chen W (2010) *Erysiphe trifolii* - a newly recognized powdery mildew pathogen of pea. Plant Pathology. 59:712-720.

Aubert G, Morin J, Jacque F, Loridon K, Quillet MC, Petit A, Rameau C, Lejeune-Hénaut I, Huguet T, Burstin J (2006) Functional mapping in pea, as an aid to the candidate gene selection and for investigating synteny with the model legume *Medicago truncatula*. Theoretical and Applied Genetics. 112:1024-1041.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. PLoS ONE. 3: e3376.

Bandillo N, Raghavan C, Muyco PA, Sevilla MAL, Lobina IT, Dilla-Ermita CJ, Tung C-W, McCouch S, Thomson M, Mauleon R, Singh RK, Gregorio G, Redoña E, Leung H (2013) Multi-parent advanced generation inter-cross (MAGIC) populations in rice: progress and potential for genetics research and breeding. Rice. 6:1-15.

Basten CJ, Weir BS, Zeng ZB (1994) Zmap-a QTL cartographer. In: Smith C, Gavora JS, Benkel B, Chesnais J, Fairfull W, Gibson JP, Kennedy BW, Burnside EB (Eds.), *Computing Strategies and Software. Proceedings of the 5th World Congress on Genetics Applied to Livestock Production. Ontario, Canada.* 22:65-66.

Basten CJ, Weir BS, Zeng ZB (2002) QTL Cartographer, Version 1.16. Department of Statistics, North Carolina State University, Raleigh, NC.

Belamkar V, Weeks NT, Bharti AK, Farmer AD, Graham MA, Cannon SB (2014) Comprehensive characterization and RNA-Seq profiling of the HD-Zip transcription factor family in soybean (*Glycine max*) during dehydration and salt stress. BMC Genomics. 15:950.

Bender CL, Alarcon-Chaidez F, Gross DC (1999) *Pseudomonas syringae* phytotoxins: mode of action, regulation, and biosynthesis by peptide and polyketide synthetases. Microbiology and Molecular Biology Reviews. 63:266-292.

Bennetzen JL, Ma J, Devos KM (2005) Mechanisms of recent genome size variation in flowering plants. Annals of Botany. 95:127-132.

Bent AF (1996) Plant disease resistance genes: function meets structure. Plant Cell. 8:1757-1771.

Ben-Ze'ev N, Zohary D (1973) Species relationship in the genus *Pisum* L. Israel Journal of Botany. 22:73-91.

Bergelson J, Roux F (2010) Towards identifying genes underlying ecologically relevant traits in *Arabidopsis thaliana*. Nature Reviews Genetics. 11:867-879.

Bevan JR, Taylor JD, Crute IR, Hunter PJ, Vivian A (1995) Genetics of specific resistance in pea (*Pisum sativum*) cultivars to seven races of *Pseudomonas syringae* pv. *pisi.* Plant Pathology. 44:98-108.

Blixt S (1972) The pea. In: King RS (Ed.), *Handbook of genetics, vol. 2.* Plenum Press, New York.

Bogacki P, Peck DM, Nair RM, Howie J, Oldach KH (2013) Genetic analysis of tolerance to boron toxicity in the legume *Medicago truncatula*. BMC Plant Biology. 13:54.

Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics. 32:314-31.

Brar DS (2002) Molecular marker assisted breeding. In: Jain SM, Brar DS, Ahloowalia BS (Eds.), *Molecular techniques in crop improvement.* Kluwer Academic Publishers, Netherlands.

Broman KW, Wu H, Sen Ś, Churchill GA (2003) R/qtl: QTL mapping in experimental crosses. Bioinformatics. 19:889-890.

Brown PH, Hu H (1996) Phloem mobility of boron is species dependent: evidence for phloem mobility in sorbitol-rich species. Annals of Botany. 77:497-505.

Buschges R, Hollricher K, Panstruga R, Simons G, Wolter M, Frijters A, vanDaelen R, vanderLee T, Diergaarde P, Groenendijk J, Töpsch S, Vos P, Salamini F, Schulze-Lefert P (1997) The barley *Mlo* gene: a novel control element of plant pathogen resistance. Cell. 88:695-705.

Butler J, MacCallum I, Kleber M, Shlyakhter I, Belmonte MK, Lander ES, Nusbaum C, Jaffe DB (2008) ALLPATHS: *De novo* assembly of whole-genome shotgun microreads. Genome Research. 18:810-820.

Cannon WA (1903) Studies in plant hybrids: the spermatogenesis of hybrid cotton. Bulletin of the Torrey Botanical Club. 133-172.

Cannon SB, Sterck L, Rombauts S, Sato S, Cheung F, Gouzy J, Wang X, Mudge J, Vasdewani J, Schiex T, Spannagl M, Monaghan E, Nicholson C, Humphray SJ, Schoof H, Mayer KF, Rogers J, Quétier F, Oldroyd GE, Debellé F, Cook DR, Retzel EF, Roe BA, Town CD, Tabata S, Van de Peer Y, Young ND (2006) Legume genome evolution viewed through the *Medicago truncatula* and *Lotus japonicus* genomes. Proceedings of the National Academy of Sciences of the United States of America. 103:14959-14964.

Chamarthi SK, Kumar A, Voung TD, Blair MW, Gaur PM, Nguyen HT, Varshney RK (2011) Trait mapping and molecular breeding. In: Pratap A, Kumar J (Eds.), *Biology and breeding of food legumes*. CABI, Oxfordshire, UK.

Chen H, Jiang JG (2010) Osmotic adjustment and plant adaptation to environmental changes related to drought and salinity. Environmental Reviews. 18:309-319.

Chen W, Provart NJ, Glazebrook J, Katagiri F, Chang HS, Eulgem T, Mauch F, Luan S, Zou G, Whitham SA, Budworth PR, Tao Y, Xie Z, Chen X, Lam S, Kreps JA, Harper JF, Si-Ammour A, Mauch-Mani B, Heinlein M, Kobayashi K, Hohn T, Dangl JL, Wang X, Zhu T (2002) Expression profile matrix of Arabidopsis

transcription factor genes suggests their putative functions in response to environmental stresses. Plant Cell. 14:559-574.

Cheng HT, Jiang H, Xue DW, Guo LB, Zeng DL, Zhang GH, Qian Q (2008) Mapping of QTL underlying tolerance to alkali at germination and early seedling stages in rice. Acta Agronomica Sinica. 34:1719-1727.

Cheng P, Holdsworth W, Ma Y, Coyne CJ, Mazourek M, Grusak MA, Fuchs S, McGee RJ (2015) Association mapping of agronomic and quality traits in USDA pea single-plant collection. Molecular Breeding. 35:75.

Choi HK, Kim DJ, Uhm T, Limpens E, Lim H, Mun JH, Kalo P, Penmetsa RV, Seres A, Kulikova O, Roe BA, Bisseling T, Kiss GB, Cook DR (2004a) A sequence-based genetic map of *Medicago truncatul*a and comparison of marker colinearity with *M. sativa*. Genetics. 166:1463-1502.

Choi HK, Mun JH, Kim DJ, Zhu H, Baek JM, Mudge J, Roe B, Ellis N, Doyle J, Kiss GB, Young ND, Cook DR (2004b) Estimating genome conservation between crop and model legume species. Proceedings of the National Academy of Sciences of the United States of America. 101:15289-15294.

Cieslarova J, Hanacek P, Fialova E, Hybl M, Smykal P (2011) Estimation of pea (*Pisum sativum* L.) microsatellite mutation rate based on pedigree and single-seed descent analyses. Journal of Applied Genetics. 52:391-401.

Clarke WE, Parkin IA, Gajardo HA, Gerhardt DJ, Higgins E, Sidebottom C, Sharpe AG, Snowdon RJ, Federico ML, Iniguez-Luy FL (2013) Genomic DNA enrichment using sequence capture microarrays: a novel approach to discover sequence nucleotide polymorphisms (SNP) in *Brassica napus* L. PLoS ONE. 8:e81992.

Collard BCY, Jahufer MZZ, Brouwer JB, Pang ECK (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: the basic concepts. Euphytica. 142:169-196.

Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, Hsu PD, Wu X, Jiang W, Marraffini LA, Zhang F (2013) Multiplex genome engineering using CRISPR/Cas systems. Science. 339:819-823.

Cooper GO (1938) Cytological investigations of *Pisum sativum*. Botanical Gazette. 99:584-591.

Crossa J, Pérez P, Hickey J, Burgueño J, Ornella L, Cerón-Rojas J, Zhang X, Dreisigacker S, Babu R, Li Y, Bonnett D, Mathews K (2014) Genomic prediction in CIMMYT maize and wheat breeding programs. Heredity. 112:48-60.

Cunha L, McFall AJ, Mackey D (2006) Innate immunity in plants: a continuum of layered defenses. Microbes and Infection. 8:1372-1381.

Cunnac S, Lindeberg M, Collmer A (2009) *Pseudomonas syringae* type III secretion system effectors: repertoires in search of functions. Current Opinion in Microbiology. 12:53-60.

Dangl JL, Jones JD (2001) Plant pathogens and integrated defence responses to infection. Nature. 411:826-833.

Davidson JA, Krysinska-Kaczmarek M, Kimber RBE, Ramsey MD (2004) Screening field pea germplasm for resistance to downy mildew (*Peronospora viciae*) and powdery mildew (*Erysiphe pisi*). Australasian Plant Pathology. 33:413-417.

Davis PH (1970) *Pisum* L. In: Davis PH (Ed.), *Flora of Turkey and east Aegean Islands*. Edinburgh University Press, Edinburgh, UK.

de Givry S, Bouchez M, Chabrier P, Milan D, Schiex T (2005) Carhta Gene: multipopulation integrated genetic and radiation hybrid mapping. Bioinformatics. 21:1703-1704.

Deschamps S, Llaca V, May GD (2012) Genotyping-by-sequencing in plants. Biology. 1:460-483.

Deulvot C, Charrel H, Marty A, Jacquin F, Donnadieu C, Lejeune-Hénaut I, Burstin J, Aubert G (2010) Highly-multiplexed SNP genotyping for genetic mapping and germplasm diversity studies in pea. BMC Genomics. 11:468.

Devos KM, Gale MD (1997) Comparative genetics in the grasses. Plant Molecular Biology. 35:3-15.

Devos KM, Gale MD (2000) Genome relationships: the grass model in current research. Plant Cell. 12:637-646.

Dirlewanger E, Isaac PG, Ranade S, Belajouza M, Cousin R, de Vienne D (1994) Restriction fragment length polymorphism analysis of loci associated with disease resistance genes and developmental traits in *Pisum sativum* L. Theoretical and Applied Genetics. 88:17-27.

Dita MA, Rispail N, Prats E, Rubiales D, Singh KB (2006) Biotechnology approaches to overcome biotic and abiotic stress constraints in legumes. Euphytica. 147:1-24.

Dong ZC, Chen Y (2013) Transcriptomics: advances and approaches. Science China Life Science. 56:960-967.

Dostalova R, Navratil M, Safarova D, Smykal P (2010) Marker assisted pea breeding: *eIF4E* allele specific markers to *pea seed-borne mosaic virus* (PSbMV) resistance. Molecular Breeding. 26:425-438.

Doyle JJ, Luckow MA (2003) The rest of the iceberg. Legume diversity and evolution in a phylogenetic context. Plant Physiology. 131:900-910.

Duarte J, Rivière N, Baranger A, Aubert G, Burstin J, Cornet L, Lavaud C, Lejeune-Hénaut I, Martinant JP, Pichon JP, Pilet-Nayel ML, Boutet G (2014) Transcriptome sequencing for high throughput SNP development and genetic mapping in pea. BMC Genomics. 15:126.

Dunwell JM (2010) Haploids in flowering plants: origins and exploitation. Plant Biotechnology Journal. 8:377-424.

Ebbels DL (2003) Principles of plant health and quarantine. CABI, Wallingford, UK.

Ek M, Eklund M, Von Post R, Dayteg C, Heriksson T, Weibull P, Ceplitis A, Isaac P, Tuvesson S (2005) Microsatellite markers for powdery mildew resistance in pea (*Pisum sativum* L.). Hereditas. 142:86-91.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS ONE. 6:e19379.

Elvira-Recuenco M, Taylor JD (2001) Resistance to bacterial blight (*Pseudomonas syringae* pv. *pisi*) in Spanish pea (*Pisum sativum*) landraces. Euphytica. 118:305-311.

Elvira-Recuenco M, Bevan JR, Taylor JD (2003) Differential responses to pea bacterial blight in stems, leaves and pods under glasshouse and field conditions. European Journal of Plant Pathology. 109:555-564.

Falloon RE, Viljanen-Rollinson SLH (2001) Powdery mildew. In: Kraft JM, Pfleger FL (Eds.), *Compendium of pea disease and pests*. APS Press, St Paul, Minn, USA.

FAOSTAT (2014) http://faostat3.fao.org. Accessed 29 August 2015.

Faville MJ, Vecchies AC, Schrieber M, Drayton MC, Hughes LJ, Jones ES, Guthridge KM, Smith KF, Sawbridge T, Spangenberg GC, Bryan GT, Forster JW (2004) Functionally associated molecular genetic marker map construction in perennial ryegrass (*Lolium perenne* L.). Theoretical and Applied Genetics. 110:12-32.

Feechan A, Jermakow AM, Torregrosa L, Panstruga R, Dry IB (2009) Identification of grapevine *MLO* gene candidates involved in susceptibility to powdery mildew. Functional Plant Biology. 35:1255-1266.

Flint-Garcia S, Thornsberry J, Bukler E (2003) Structure of linkage disequilibrium in plants. Annual Review Plant Biology. 54:357-374.

Fondevilla S, Carver TLW, Moreno MT, Rubiales D (2006) Macroscopic and histological characterisation of genes *er1* and *er2* for powdery mildew resistance in pea. European Journal of Plant Pathology. 115:309-321.

Fondevilla S, Torres AM, Moreno MT, Rubiales D (2007) Identification of a new gene for resistance to powdery mildew in *Pisum fulvum*, a wild relative of pea. Breeding Science. 57:181-184.

Fondevilla S, Rubiales D, Moreno MT, Torres AM (2008) Identification and validation of RAPD and SCAR markers linked to the gene *Er3* conferring resistance to *Erysiphe pisi* DC in pea. Molecular Breeding. 22:193-200.

Fondevilla S, Cubero JI, Rubiales D (2011) Confirmation that the *Er3* gene, conferring resistance to *Erysiphe pisi* in pea, is a different gene from *er1* and *er2* genes. Plant Breeding. 130:281-282.

Fondevilla S, Martin-Sanz A, Satovic Z, Fernández-Romero M, Rubiales D, Caminero C (2012) Identification of quantitative trait loci involved in resistance to

Pseudomonas syringae pv. *syringae* in pea (*Pisum sativum* L.). Euphytica. 186:805-812.

Fondevilla S, Rubiales D (2012) Powdery mildew control in pea. A review. Agronomy for Sustainable Development. 32:401-409.

Francois LE, Maas EV (1994) Crop response and management on salt-affected soils. In: Pessarakli M (Ed.), *Handbook of plant and crop stress. Third Edition*. CRC Press, New York, USA.

Franssen SU, Shrestha RP, Bräutigam A, Bornberg-Bauer E, Weber APM (2011) Comprehensive transcriptome analysis of the highly complex *Pisum sativum* genome using next generation sequencing. BMC Genomics. 12:227.

Fuchs J, Kühne M, Schubert I (1998) Assignment of linkage groups to pea chromosomes after karyotyping and gene mapping by fluorescent in situ hybridization. Chromosoma. 107:272-276.

Gale MD, Devos KM (1998) Comparative genetics in the grasses. Proceedings of the National Academy of Sciences of the United States of America. 95:1971-1974.

Gao L, Tu ZJ, Millett BP, Bradeen JM (2013) Insights into organ-specific pathogen defense responses in plants: RNA-Seq analysis of potato tuber-*Phytophthora infestans* interactions. BMC Genomics. 14:340.

Góngora-Castillo E, Buell CR (2013) Bioinformatics challenges in *de novo* transcriptome assembly using short read sequences in the absence of a reference genome sequence. Natural Product Reports. 30:490-500.

Goyer A, Hamlin L, Crosslin JM, Buchanan A, Chang JH (2015) RNA-Seq analysis of resistant and susceptible potato varieties during the early stages of potato virus Y infection. BMC Genomics. 16:472.

Grabherr MG, Haas BJ, Yassour M, Levin JZ, Thompson DA, Amit I, Adiconis X, Fan L, Raychowdhury R, Zeng Q, Chen Z, Mauceli E, Hacohen N, Gnirke A, Rhind N, di Palma F, Birren BW, Nusbaum C, Lindblad-Toh K, Friedman N, Regev A (2011) Full-length transcriptome assembly from RNA-Seq data without a reference genome. Nature Biotechnology. 29:644-652.

Graham MA, Marek LF, Shoemaker RC (2002) Organization, expression and evolution of a disease resistance gene cluster in soybean. Genetics. 162:1961-1977.

Grieve CM, Poss JA (2000) Wheat response to interactive effects of boron and salinity. Journal of Plant Nutrition. 23:1217-1226.

Gritton ET, Ebert RD (1975) Interaction of planting date and powdery mildew on pea plant performance. American Society of Horticultural Science. 100:137-142.

Gupta B, Huang B (2014) Mechanism of salinity tolerance in plants: physiological, biochemical, and molecular characterization. International Journal of Genomics. 2014:18.

Gupta PK, Balyan HS, Sharma PC, Ramesh B (1996) Microsatellites in plants: a new class of molecular markers. Current Science. 70:45-54.

Gupta PK, Varshney RK (2000) The development and use of microsatellite markers for genetic analysis and plant breeding with emphasis on bread wheat. Euphytica. 113:163-185.

Gupta PK, Roy JK, Prasad M (2001) Single nucleotide polymorphism: a new paradigm for molecular marker technology and DNA polymorphism detection with emphasis on their use in plants. Current Science. 80:524-535.

Gupta PK, Varshney RK, Prasad M (2002) Molecular markers: principles and technology. In: Jain SM, Brar DS, Ahloowalia BS (Eds.), *Molecular techniques in crop improvement*. Kluwer Academic Publishers, Netherlands.

Guttman M, Garber M, Levin JZ, Donaghey J, Robinson J, Adiconis X, Fan L, Koziol MJ, Gnirke A, Nusbaum C, Rinn JL, Lander ES, Regev A (2010) Ab initio reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. Nature Biotechnology. 28:503-510.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, Macmanes MD, Ott M, Orvis J, Pochet N, Strozzi F, Weeks N, Westerman R, William T, Dewey CN, Henschel R, Leduc RD, Friedman N, Regev A (2013) *De novo* transcript sequence reconstruction from RNA-Seq using the Trinity platform for reference generation and analysis. Nature Protocols. 8:1494-1512.

Hammond-Kosack KE, Jones JD (1997) Plant disease resistance genes. Annual Review of Plant Physiology and Plant Molecular Biology. 48:575-607.

Harland SC (1948) Inheritance of immunity to mildew in Peruvian forms of *Pisum sativum*. Heredity. 2:263-269.

Harper AL, Trick M, Higgins J, Fraser F, Clissold L, Wells R, Hattori C, Werner P, Bancroft I (2012) Associative transcriptomics of traits in the polyploid crop species *Brassica napus*. Nature Biotechnology. 30:798-802.

Harter LZ, Zaumeyer WJ, Wade BL (1934) Pea diseases and their control. US Department of Agriculture. Farmers Bulletin No. 1735.

Hayes JE, Reid RJ (2004) B tolerance in barley is mediated by efflux of B from the roots. Plant Physiology. 136:3376-3382.

Heringa RJ, Van Nore A, Tazelaar MF (1969) Resistance to powdery mildew (*Erysiphe polygoni* D.C.) in peas (*Pisum sativum* L.). Euphytica. 18:163-169.

Hirano SS, Upper CD (2000) Bacteria in the leaf ecosystem with emphasis on *Pseudomonas syringae*: a pathogen, ice nucleus, and epiphyte. Microbiological Molecular Biology Review. 64:624-653.

Hollaway GJ, Bretag TW (1995) Occurrence and distribution of races of *P. syringae* pv*. pisi* in Australia and their specificity towards various field pea (*Pisum sativum*) cultivars. Australian Journal of Experimental Agriculture. 35:629-632.

Hollaway GJ, Bretag TW, Gooden JM, Hannah MC (1996) Effect of soil water content and temperature on the transmission of *Pseudomonas syringae* pv. *pisi* from pea seed (*Pisum sativum*) to seedling. Australasian Plant Pathology. 25:26-30.

Hollaway GJ, Bretag TW, Price TV (2007) The epidemiology and management of bacterial blight (*Pseudomonas syringae* pv. *pisi*) of field pea (*Pisum sativum*) in Australia: a review. Australian Journal of Agricultural Research. 58:1086-1099.

Hou S, Yang Y, Zhou JM (2009) The multilevel and dynamic interplay between plant and pathogen. Plant Signalling and Behaviour. 4:283-293.

Huettel B, Santra D, Muehlbauer F, Kahl G (2002) Resistance gene analogues of chickpea (*Cicer arietinum* L.): isolation, genetic mapping and association with a *Fusarium* resistance gene cluster. Theoretical and Applied Genetics. 105:479-490.

Humphry M, Reinstädler A, Ivanov S, Bisseling T, Panstruga R (2011) Durable broad-spectrum powdery mildew resistance in pea *er1* plants is conferred by natural loss-of-function mutations in *PsMLO1*. Molecular Plant Pathology. 12:866-878.

Hunter PJ, Ellis N, Taylor JD (2001) Association of dominant loci for resistance to *Pseudomonas syringae* pv*. pisi* with linkage groups II, VI and VII of *Pisum sativum.* Theoretical and Applied Genetics. 103:129-135.

Hussain K, Majeed A, Nawaz K, Khizar HB, Nisar MF (2009) Effect of different levels of salinity on growth and ion contents of black seeds (*Nigella sativa* L.). Current Research Journal of Biological Sciences. 1:135-138.

Huynh B-L, Ehlers JD, Ndeve A, Wanamaker S, Lucas MR, Close TJ, Roberts PA (2015) Genetic mapping and legume synteny of aphid resistance in African cowpea (*Vigna unguiculata* L. Walp.) grown in California. Molecular Breeding. 35:36.

Jacoby B (1999) Mechanisms involved in salt tolerance of plants. In: Pessarakli M (Ed.), *Information handbook of plant and crop Stress*. CRC Press, New York, USA.

Jain P, Krishnan NM, Panda B (2013) Augmenting transcriptome assembly by combining *de novo* and genome-guided tools. PeerJ 1:e133.

Janila P, Sharma B (2004) RAPD and SCAR markers for powdery mildew resistance gene *er* in pea. Plant Breeding. 123:271-274.

Jefferies SP, Barr AR, Karakousis A, Kretschmer JM, Manning S, Chalmers KJ, Nelson JC, Islam A, Langridge P (1999) Mapping of chromosome regions conferring boron toxicity tolerance in barley (*Hordeum vulgare* L.). Theoretical and Applied Genetics. 98:1293-1303.

Jefferies SP, Pallotta MA, Paull JG, Karakousis A, Kretschmer JM, Manning S, Islam A, Langridge P, Chalmers KJ (2000) Mapping and validation of chromosome regions conferring boron toxicity tolerance in wheat (*Triticum aestivum*). Theoretical and Applied Genetics. 101:767-777.

Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell. 60:473-485.

Jing R, Johnson R, Seres A, Kiss G, Ambrose MJ, Knox MR, Ellis THN, Flavell AJ (2007) Gene-based sequence diversity analysis of field pea (*Pisum*). Genetics. 177:2263-2275.

Jones ES, Sullivan H, Bhattramakki D, Smith JSC (2007) A comparison of simple sequence repeat and single nucleotide polymorphism marker technologies for the genotypic analysis of maize (*Zea mays* L.). Theoretical and Applied Genetics. 115:361-371.

Jones N, Ougham H, Thomas H (1997) Markers and mapping: we are all geneticists now. New Phytologist. 137:165-177.

Jørgensen JH (1994) Genetics of powdery mildew resistance in barley. Critical Reviews in Plant Sciences. 13:97-119.

Joung JK, Sander JD (2013) TALENs: a widely applicable technology for targeted genome editing. Nature Reviews Molecular Cell Biology. 14:49-55.

Kalo P, Seres A, Taylor SA, Jakab J, Kevei Z, Kereszt A, Endre G, Ellis THN, Kiss GB (2004) Comparative mapping between *Medicago sativa* and *Pisum sativum*. Molecular Genetics and Genomics. 272:235-246.

Kanazin V, Frederick ML, Shoemaker RC (1996) Resistance gene analogs are conserved and clustered in soybean. Proceedings of the National Academy of Sciences of the United States of America. 93:11746-11750.

Karimi-Kurdistani G, Harighi B (2008) Phenotypic and molecular properties of *P. syringae* pv*. syringae* the causal agent of bacterial canker of stone fruit trees in Kurdistan province. Journal of Plant Pathology. 90:81-86.

Katoch V, Sharma S, Pathania S, Banayal DK, Sharma SK, Rathour R (2010) Molecular mapping of pea powdery mildew resistance gene *er2* to pea linkage group III. Molecular Breeding. 25:229-237.

Kaur S, Pembleton L, Cogan NOI, Savin KW, Leonforte T, Paull J, Materne M, Forster JW (2012) Transcriptome sequencing of field pea and faba bean for discovery and validation of SSR genetic markers. BMC Genomics. 13:104.

Kaur S, Cogan NOI, Stephens A, Noy D, Butsch M, Forster JW, Materne M (2014) EST-SNP discovery and dense genetic mapping in lentil (*Lens culinaris* Medik.) enable candidate gene selection for boron tolerance. Theoretical and Applied Genetics. 127:703-713.

Kawahara Y, Oono Y, Kanamori H, Matsumoto T, Itoh T, Minami E (2012) Simultaneous RNA-Seq analysis of a mixed transcriptome of rice and blast fungus interaction. PLoS One. 7:e49423.

Keller B, Feuillet C, Yahiaoui N (2005) Map-based isolation of disease resistance genes from bread wheat: cloning in a supersize genome. Genetical Research. 85:93-100.

Keurentjes JJ, Bentsink L, Alonso-Blanco C, Hanhart CJ, Blankestijn-De Vries H, Effgen S, Vreugdenhil D, Koornneef M (2007) Development of a near-isogenic line population of *Arabidopsis thaliana* and comparison of mapping power with a recombinant inbred line population. Genetics. 175:891-905.

Kindiger B (2006) Cross-species amplification of *Lolium* microsatellites in *Poa* ssp. Grassland Science. 52:105-115.

Knight (1944) Pea disease in Idaho. Idaho Agricultural Experiment Station Bulletin No 253.

Knox MR, Ellis THN (2002) Map expansion in pea. Genetics. 162:861-873.

Kobe B, Deisenhofer J (1994) The leucine-rich repeat: a versatile binding motif. Trends in Biochemical Sciences. 19:415-421.

Kocer A, Albayrak S (2012) Determination of forage yield and quality of pea (*Pisum sativum* L.) mixtures with oat and barley. Turkish Journal of Field Crops. 17:96-99.

Komori T, Ohta S, Murai N, Takakura Y, Kuraya Y, Suzuki S, Hiei Y, Imaseki H, Nitta N (2004) Map-based cloning of a fertility restorer gene, *Rf-1*, in rice (*Oryza sativa* L.). The Plant Journal. 37:315-325.

Krejci P, Matuskova P, Hanacek P, Reinohl V, Prochazka S (2007) The transformation of pea (*Pisum sativum* L.): applicable methods of *Agrobacterium tumefaciens*-mediated gene transfer. Acta Physiologiae Plantarum. 29:157-163.

Kumar H, Singh RB (1981) Genetic analysis of adult plant resistance to powdery mildew in pea (*Pisum sativum* L.). Euphytica. 30:147-154.

Kumar J, Choudhary AK, Solanki RK, Pratap AP (2011) Towards marker-assisted selection in pulses: a review. Plant Breeding. 130:297-313.

Lacey A, Davies S (2009) Boron toxicity in WA soils. Department of Agriculture and Food, Western Australia, Farm note 388.

Lander EC, Green P, Abrahamson J, Barlow A, Daly MJ, Lincoln SE, Newburg L (1987) MAPMAKER: An interactive computer package for constructing primary genetic linkage maps of experimental and natural populations. Genomics 1:174-181.

Langmead B, Trapnell C, Pop M, Salzberg S (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biology. 10:R25.

Läuchli A, Grattan SR (2007) Plant growth and development under salinity stress. In: Jenks MA, Hasegawa PA, Jain SM (Eds.), *Advances in molecular-breeding towards salinity and drought tolerance*. Springer-Verlag, Netherlands.

Lavin M, Herendeen PS, Wojciechowski MF (2005) Evolutionary rates analysis of Leguminosae implicates a rapid diversification of lineages during the tertiary. Systematic Biology. 54:575-594.

Lawyer AS (1984) Diseases caused by bacteria In: Hagedorn DJ (Ed.), C*ompendium of pea diseases*. APS Press, St Paul, Minn, USA.

Lee GJ, Boerma HR, Villagarcia MR, Zhou X, Carter Jr, TE, Li Z, Gibbs MO (2004) A major QTL conditioning salt tolerance in S-100 soybean and descendent cultivars. Theoretical and Applied Genetics. 109:1610-1619.

Leister D, Ballvora A, Salamini S, Gebhardt CA (1996) PCR based approach for isolating pathogen resistance genes from potato with potential for wide application in plants. Nature Genetics. 14:421-429.

Leonforte A (2013) A study of salinity tolerance in field pea. PhD thesis, Department of Agriculture and Food Systems, Melbourne School of Land and Environment, The University of Melbourne.

Lewis GB, MacKinder SB, Lock M (2005) Legumes of the world. Royal Botanical Gardens, Kew, UK.

Liu B (1998) Statistical genomics: linkage, mapping and QTL analysis. CRC Press, Boca Raton, Florida, USA.

Liu G, Bernhardt JL, Jia MH, Wamishe YA, Jia Y (2008) Molecular characterization of the recombinant inbred line population derived from a japonica-indica rice cross. Euphytica. 159:73-82.

Liu J, Liu X, Dai L, Wang G (2007) Recent progress in elucidating the structure, function and evolution of disease resistance genes in plants. Journal of Genetics and Genomics. 34:765-776.

Liu JJ, Sniezko RA, Sturrock RN, Chen H (2014) Western white pine SNP discovery and high-throughput genotyping for breeding and conservation applications. BMC Plant Biology. 14:380.

Liu S, Lin L, Jiang P, Wang D, Xing Y (2011) A comparison of RNA-Seq and high-density exon array for detecting differential gene expression between closely related species. Nucleic Acids Research. 39:578-588.

Liu SR, Li WY, Long D, Hu CG, Zhang JZ (2013) Development and characterization of genomic and expressed SSRs in citrus by genome-wide analysis. PLoS One. 8:e75149.

Lockhart DJ, Winzeler EA (2000) Genomics, gene expression and DNA arrays. Nature. 405:827-835.

Losada LC, Hutcheson SW (2005) Type III secretion chaperones of *Pseudomonas syringae* protect effectors from Lon-associated degradation. Molecular Microbiology. 55:941-953.

Ma LQ, Zhou EF, Huo NZ, Zhou RH, Wang GY, Jia JZ (2007) Genetic analysis of salt tolerance in a recombinant inbred population of wheat (*Triticum aestivum* L.). Euphytica. 153:109-117.

Maas EV (1986) Salt tolerance of plants. Applied Agricultural Research. 1:12-26.

Macas J, Neumann P, Navratilova A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics. 8:427.

Maher L, Armstrong R, Connor D (2003) Salt tolerant lentils - a possibility for the future? In: *Solutions for better environment*. *Proceedings of the 11th Australian Agronomy Conference. Geelong, Victoria. Australian Society of Agronomy*.

Mammadov J, Aggarwal R, Buyyarapu R, Kumpatla S (2012) SNP markers and their impact on plant breeding. International Journal of Plant Genomics. 2012:11.

Manikanda BN (2013) Genetic mapping and marker assisted selection: basics, practice and benefits. Springer, India.

Manly KF, Cudmore Jr RH, Meer JM (2001) Map Manager QTX, cross-platform software for genetic mapping. Mammalian Genome. 12:930-932.

Manneh B, Stam P, Struik PC, Bruce-Oliver S, van Eeuwijk FA (2007) QTL-based analysis of genotype-by-environment interaction for grain yield of rice in stress and non-stress environments. Euphytica. 156:213-226.

Mano Y, Takeda K (1997) Mapping quantitative trait loci for salt tolerance at germination and the seedling stage in barley (*Hordeum vulgare* L.). Euphytica. 94:263-272.

Mardis ER (2008) The impact of next-generation sequencing technology on genetics. Trends in Genetics. 24:133-141.

Marone D, Russo MA, Laidò G, De Leonardis AM, Mastrangelo AM (2013) Plant nucleotide binding site–leucine-rich repeat (NBS-LRR) genes: active guardians in host defense responses. International Journal of Molecular Sciences. 14:7302-7326.

Martin-Sanz A, Palomo JL, Pérez de la Vega M, Caminero C (2011) Identification of pathovars and races of *Pseudomonas syringae*, the main causal agent of bacterial diseases in pea in North-Central Spain, and the search for resistance. European Journal of Plant Pathology. 129:57-69.

Martin-Sanz A, Pérez de la Vega M, Caminero C (2012a) Resistance to *Pseudomonas syringae* in a collection of pea germplasm under field and controlled conditions. Plant Pathology. 61:375-387.

Martin-Sanz A, Pérez de la Vega M, Murillo J, Caminero C (2012b) Genetic, biochemical and pathogenic diversity of *Pseudomonas syringae* pv. *pisi* strains. Plant Pathology. 61:1063-1072.

Martin-Sanz A, Pérez de la Vega M, Murillo J, Caminero C (2013) Strains of *Pseudomonas syringae* pv. *syringae* from pea are phylogenetically and pathogenically diverse. Phytopathology. 103:673-681.

Marx GA (1986) Location of er proving elusive. Pisum Newsletter. 18:39-41.

McPhee K (2003) Dry pea production and breeding, a mini-review. The Journal of Food, Agriculture and Environment. 1:64-69.

Mendel G (1866) Versuche über Plflanzenhybriden. In: *Verhandlungen des naturforschenden Vereines in Brünn [Proceedings of the Natural History Society of Brünn] IV (1865): 3-47.*

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157:1819-1829.

Meyers BC, Dickerman AW, Michelmore RW, Sivaramakrishnan S, Sobral BW, Young ND (1999) Plant disease resistance genes encode members of an ancient and diverse protein family within the nucleotide-binding superfamily. The Plant Journal. 20:317-332.

Meyers BC, Kozik A, Griego A, Kuang H, Michelmore RW (2003) Genome-wide analysis of NBS-LRR–encoding genes in Arabidopsis. The Plant Cell Online. 15:8.

Miwa K, Tanaka M, Kamiya T, Fujiwara T (2010) Molecular mechanisms of boron transport in plants: involvement of Arabidopsis NIP5; 1 and NIP6; 1. In: Jahn TP, Bienert GP (Eds.), *Mips and their role in the exchange of metalloids.* Springer, New York, USA.

Monforte AJ, Tanksley SD (2000) Development of a set of near isogenic and backcross recombinant inbred lines containing most of the *Lycopersicon hirsutum* genome in a *L. esculentum* genetic background: a tool for gene mapping and gene discovery. Genome. 43:803-813.

Moreau S, Verdenaud M, Ott T, Letort S, de Billy F, Niebel A, Gouzy J, de Carvalho-Niebel F, Gamas P (2011) Transcription reprogramming during root nodule development in *Medicago truncatula*. PLoS ONE. 6:e16463.

Morozova O, Marra MA (2008) Applications of next-generation sequencing technologies in functional genomics. Genomics. 92:255-264.

Mudge J, Cannon SB, Kalo P, Oldroyd GE, Roe BA, Town CD, Young ND (2005) Highly syntenic regions in the genomes of soybean, *Medicago truncatula*, and *Arabidopsis thaliana*. BMC Plant Biology. 5:15.

Munjal RL, Chenulu VV, Hora TS (1963) Assessment of losses due to powdery mildew (*Erysiphe polygoni*) on pea. Indian Phytopathology. 19:260-267.

Murray GM, Brennan JP (2012) The Current and potential costs from diseases of pulse crops in Australia. Grains Research and Development Corporation. www.grdc.com.au/Resources/Publications.

Murray MG, Cuellar RE, Thompson WF (1978) DNA sequence organisation in the pea genome. Biochemistry. 17:5781-5890.

Murray MG, Peters DL, Thompson WF (1981) Ancient repeated sequences in the pea and mung bean genomes and implications for genome evolution. Journal of Molecular Evolution. 17:31-42.

Muthamilarasan M, Prasad M (2013) Plant innate immunity: an updated insight into defence mechanism. Journal of Biosciences. 38:433-449.

Mysore KS, Ryu MC (2004) Nonhost resistance: how much do we know? Trends in Plant Science. 9:97-104.

Neumann P, Pozárková D, Vrána J, Doležel J, Macas J (2002) Chromosome sorting and PCR-based physical mapping in pea (*Pisum sativum* L.). Chromosome Research. 10:63-71.

Nikolopoulou D, Grigorakis K, Stasini M, Alexis MN, Iliadis K (2007) Differences in chemical composition of field pea (*Pisum sativum*) cultivars: effect of cultivation area and year. Food Chemistry. 103:847-852.

Nimchuk Z, Rohmer L, Chang JH, Dangl JL (2001) Knowing the dancer from the dance: R-gene products and their interactions with other proteins from host and pathogen. Current Opinion in Plant Biology. 4:288-294.

O'Brien HE, Desveaux D, Guttman DS (2011) Next generation genomics of *Pseudomonas syringae.* Current Opinion in Microbiology. 14:24-30.

Okcu G, Kaya MD, Atak M (2005) Effects of salt and drought stresses on germination and seedling growth of pea (*Pisum sativum* L.). Turkish Journal of Agriculture and Forestry. 29:237-242.

Ondřej M, Dostálová R, Odstrčilová L (2005) Response of *Pisum sativum* germplasm resistant to *Erysiphe pisi* to inoculation with *Erysiphe baeumleri*, a new pathogen of peas. Plant Protection Science. 41:95-103.

Panstruga R (2005) Serpentine plant MLO proteins as entry portals for powdery mildew fungi. Biochemical Society Transactions. 33:389-392.

Paran I, Goldman I, Tanksley SD, Zamir D (1995) Recombinant inbred lines for genetic mapping in tomato. Theoretical and Applied Genetics. 90:542-548.

Paull JG, Rathjen AJ, Cartwright B (1991) Major gene control of tolerance of bread wheat (*Triticum aestivum* L.) to high concentrations of soil boron. Euphytica. 55:217-228.

Pavan S, Schiavulli A, Appiano M, Miacola C, Visser RG, Bai Y, Lotti C, Ricciardi L (2013) Identification of a complete set of functional markers for the selection of *er1* powdery mildew resistance in *Pisum sativum* L. Molecular Breeding. 31:247-253.

Pearson CE, Sinden RR (1998) Trinucleotide repeat DNA structures: dynamic mutations from dynamic DNA. Current Opinion Structural Biology. 8:321-330.

Pfeil BE, Schlueter JA, Shoemaker RC, Doyle JJ (2005) Placing paleopolyploidy in relation to taxon divergence: a phylogenetic analysis in legumes using 39 gene families. Systematic Biology. 54:441-454.

Pink DAC (2002) Strategies using genes for non-durable resistance. Euphytica. 1:227-236.

Pottorff M, Ehlers JD, Fatokun C, Roberts PA, Close TJ (2012) Leaf morphology in cowpea (*Vigna unguiculata* (L.) Walp.): QTL analysis, physical mapping and identifying a candidate gene using synteny with model legume species. BMC Genomics. 13:234.

Provvidenti R, Hampton RO (1991) Chromosomal distribution of genes for resistance to seven potyviruses in *Pisum sativum*. Pisum Genetics. 23:26-28.

Pulse Australia (2009) Field pea *(Pisum sativum)* General Information. http://www.pulseaus.com.au/index.aspx.

Rafalski A (2002) Applications of single nucleotide polymorphism in crop genetics. Current Opinion in Plant Biology. 5:94-100.

Ramos AR, Morello JE, Ravindran S, Deng WL, Huang HC, Collmer A (2007) Identification of *Pseudomonas syringae* pv. *syringae* 61 Type III secretion system Hrp proteins that can travel the type III pathway and contribute to the translocation of effector proteins into plant cells. Journal of Bacteriology. 189:5773-5778.

Raney JA, Reynolds DJ, Elzinga DB, Page J, Udall JA, Jellen EN, Bonfacio A, Fairbanks DJ, Maughan PJ (2014) Transcriptome analysis of drought-induced stress in *Chenopodium quinoa*. American Journal of Plant Sciences. 2014:338-357.

Redden B, Leonforte T, Ford R, Croser J, Slattery J (2005) Pea (*Pisum sativum* L.). In: Singh RJ (Ed.), *Genetic resources, chromosome engineering, and crop improvement.* CRC Press, Florida, USA.

Reid R (2007) Identification of Boron transporter genes likely to be responsible for tolerance to Boron toxicity in wheat and barley. Plant Cell Physiology. 48:1673-1678.

Rengasamy P (2006) World salinization with emphasis on Australia. Journal of Experimental Botany. 57:1017-1023.

Rennie WJ, Cockerell V (2006) Seedborne Disease. In: Cook BM, Jones DG, Kaye B (Eds.), *The epidemiology of plant diseases*. Springer, Netherlands.

Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science. 273:1516-1517.

Richardson HJ, Hollaway GJ (2011) Bacterial blight caused by *Pseudomonas syringae* pv. *syringae* shown to be an important disease of field pea in south eastern Australia. Australasian Plant Pathology. 40:260-268.

Roa AC, Chavarriaga-Aguirre P, Duque MC, Maya MM, Bonierbale MW, Iglesias C, Tohme J (2000) Cross-species amplification of cassava (*Manihot esculenta*) (Euphorbiaceae) microsatellites: allelic polymorphism and degree of relationship. American Journal of Botany. 87:1647-1655.

Roberts SJ (1992) Effect of soil moisture on the transmission of pea bacterial blight (*Pseudomonas syringae* pv. *pisi*) from seed to seedling. Plant Pathology. 41:136-140.

Roberts SJ (1993) Effect of bacterial blight (*Pseudomonas syringae* pv. *pisi*) on the growth and yield of single pea (*Pisum sativum*) plants under glasshouse conditions. Plant Pathology. 42:568-576.

Roberts SJ, Phelps K, McKeown BM, Heath MC, Cockerell V (1995) Effect of bacterial blight (*Pseudomonas syringae* pv. *pisi*) on the yield of spring sown combing peas (*Pisum sativum*). Annals of Applied Biology. 126:61-73.

Roberts SJ (1997) Effect of weather conditions on local spread and infection by pea bacterial blight (*Pseudomonas syringae* pv. *pisi*). European Journal of Plant Pathology. 103:711-719.

Robertson G, Schein J, Chiu R, Corbett R, Field M, Jackman SD, Mungall K, Lee S, Okada HM, Qian JQ, Griffith M, Raymond A, Thiessen N, Cezard T, Butterfield YS, Newsome R, Chan SK, She R, Varhol R, Kamoh B, Prabhu A-L, Tam A, Zhao YJ, Moore RA, Hirst M, Marra MA, Jones SJM, Hoodless PA, Birol I (2010) *De novo* assembly and analysis of RNA-Seq data. Nature Methods. 7:909-912.

Robinson AJ, Love CG, Batley J, Barker G, Edwards D (2004) Simple sequence repeat marker loci discovery using SSR primer. Bioinformatics. 20:1475-1476.

Rozema J, De Bruin J, Broekman RA (1992) Effect of boron on growth and mineral economy of some halophytes and nonhalophytes. New Phytologist. 121:249-256.

Rutkoski JE, Heffner EL, Sorrells ME (2011) Genomic selection for durable stem rust resistance in wheat. Euphytica. 179:161-173.

Sadiki M, Rabih K (2001) Selection of chickpea (*Cicer arietinum*) for yield and symbiotic nitrogen fixation ability under salt stress. Agronomie. 21:659-666.

Samatadze TE, Zelenina DA, Shostak NG, Volkov AA, Popov KV, Rachinskaya OV, Borisov Alu, Tihonovich IA, Zelenin AV, Muravenko OV (2008) Comparative genome analysis in pea *Pisum sativum* L. varieties and lines with chromosomal and molecular markers. Russian Journal of Genetics. 44:1424-1430.

Sato S, Nakamura Y, Kaneko T, Asamizu E, Kato T, Nakao M, Sasamoto S, Watanabe A, Ono A, Kawashima K, Fujishiro T, Katoh M, Kohara M, Kishida Y, Minami C, Nakayama S, Nakazaki N, Shimizu Y, Shinpo S, Takahashi C, Wada T, Yamada M, Ohmido N, Hayashi M, Fukui K, Baba T, Nakamichi T, Mori H, Tabata S (2008) Genome structure of the legume, *Lotus japonicus*. DNA Research. 15:227-39.

Sauna ZE, Kimchi-Sarfaty C (2011) Understanding the contribution of synonymous mutations to human disease. Nature Reviews Genetics. 12:683-691.

Saxena NP, Saxena MC, Ruckenbauer P, Rana RS, El-Fouly MM, Shabana R (1994) Screening techniques and sources of tolerance to salinity and mineral nutrient imbalances in cool season food legumes. Euphytica. 73:85-93.

Scarth GW (1944) Cell physiological studies of frost resistance - a review. New Phytologist. 43:1-12.

Schaefer H, Hechenleitner P, Santos-Guerra A, de Sequeira MM, Pennington RT, Kenicer G, Carine MA (2012) Systematics, biogeography, and character evolution of the legume tribe Fabeae with special focus on the middle-Atlantic island lineages. BMC Evolutionary Biology. 12:250.

Schmit J, Taylor JD, Roberts SJ (1993) Sources of resistance to pea bacterial blight (*Pseudomonas syringae* pv. *pisi*) in pea germplasm. In: *Proceedings of the 6th International Congress of Plant Pathology. Montreal, Canada.*

Schmutz J, Cannon SB, Schlueter J, Ma J, Mitros T, Nelson W, Hyten DL, Song Q, Thelen JJ, Cheng J, Xu D, Hellsten U, May GD, Yu Y, Sakurai T, Umezawa T, Bhattacharyya MK, Sandhu D, Valliyodan B, Lindquist E, Peto M, Grant D, Shu S, Goodstein D, Barry K, Futrell-Griggs M, Abernathy B, Du J, Tian Z, Zhu L, Gill N, Joshi T, Libault M, Sethuraman A, Zhang XC, Shinozaki K, Nguyen HT, Wing RA, Cregan P, Specht J, Grimwood J, Rokhsar D, Stacey G, Shoemaker RC, Jackson SA (2010) Genome sequence of the palaeopolyploid soybean. Nature. 463:178-183.

Schneider K (2005) Mapping populations and principles of genetic mapping. In: Meksem K, Kahl G (Eds.), *The handbook of plant genome mapping: genetic and physical mapping*. WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim, Germany.

Schnurbusch T, Hayes J, Hrmova M, Baumann U, Ramesh SA, Tyerman SD, Langridge P, Sutton T (2010) Boron toxicity tolerance in barley through reduced expression of the multifunctional aquaporin HvNIP2;1. Plant Physiology. 153:1706-1715.

Schroeder HE, Schotz AH, Wardley-Richardson T, Spencer D, Higgins TJV (1993) Transformation and regeneration of two cultivars of pea (*Pisum sativum* L.). Plant Physiology. 101:751-757.

Schulz MH, Zerbino DR, Vingron M, Birney E (2012) Oases: robust *de novo* RNA-Seq assembly across the dynamic range of expression levels. Bioinformatics. 28:1086-1092.

Schwim F (1994) Seed treatment - a panacea for plant production? In: TJ Martin (Ed.) *Seed treatment: progress and prospects*. BCPC Monograph N0 57. British Crop Protection Council, Farnham, Surrey, UK.

Semagn K, Bjørnstad A, Ndjiondjop MN (2006) Principles, requirements and prospects of genetic mapping in plants. African Journal of Biotechnology. 5:2569-2587.

Semagan K, Bjornstad A, Xu Y (2010) The genetic dissection of quantitative traits in crops. Electronic Journal of Biotechnology. 13. http://dx.doi.org/10.2225/vol13-issue5-fulltext-14.

Shastry BS (2009) SNPs: impact on gene function and phenotype. Methods in Molecular Biology. 578:3-22.

Shi S, Chen T, Zhao X (2013) Comparative transcriptome analysis for metabolic engineering. Methods in Molecular Biology. 985:447-458.

Shirasawa K, Bertioli DJ, Varshney RK, Moretzsohn MC, Leal-Bertioli SC, Thudi M, Pandey MK, Rami JF, Foncéka D, Gowda MV, Qin H, Guo B, Hong Y, Liang X, Hirakawa H, Tabata S, Isobe S (2013) Integrated consensus map of cultivated peanut and wild relatives reveals structures of the A and B genomes of Arachis and divergence of the legume genomes. DNA Research. 20:173-184.

Shu Y, Zhang J, Ao Y, Song L, Guo C (2015) Analysis of the *Thinopyrum elongatum* transcriptome under water deficit stress. International Journal of Genomics. 2015. Article ID 265791.

Sim SC, Durstewitz G, Plieske J, Wieseke R, Ganal MW, Van Deynze A, Hamilton JP, Buell CR, Causse M, Wijeratne S, Francis DM (2012) Development of a large SNP genotyping array and generation of high-density genetic maps in tomato. PLoS ONE. 7:e40563.

Sindhu A, Ramsay L, Sanderson L-A, Stonehouse R, Li R, Condie J, Shunmugam ASK, Liu Y, Jha AB, Diapari M, Burstin J, Aubert G, Tar'an B, Bett KE, Warkentin TD, Sharpe AG (2014) Gene-based SNP discovery and genetic mapping in pea. Theoretical and Applied Genetics. 127:2225-2241.

Singh R (2012) Pros and cons of utilizing major, race-specific resistance genes versus partial resistance in breeding rust resistant wheat. In: McIntosh R (Ed.), *Proceedings of the Borlaug Global Rust Initiative (BGRI) 2012 Technical Workshop. Beijing, China*.

Skoric V (1927) Bacterial blight of field pea: overwintering, dissemination and pathological histology. Phytopathology. 17:611-627.

Smith JSC, Chin ECL, Shu H, Smith OS, Wall SJ, Senior ML, Mitchell SE, Kresovich S, Ziegle JN (1997) An evaluation of the utility of SSR loci as molecular markers in maize (*Zea mays* L.): comparisons with data from RFLPS and pedigree. Theoretical and Applied Genetics. 95:163-173.

Smith PH, Foster EM, Boyd LA, Brown JKM (1996) The early development of *Erysiphe pisi* on *Pisum sativum* L. Plant Pathology. 45:302-309.

Smýkal P, Kenicer G, Flavell AJ, Corander J, Kosterin O, Redden RJ, Ford R, Coyne CJ, Maxted N, Ambrose MJ, Ellis NTH (2011) Phylogeny, phylogeography and genetic diversity of the *Pisum* genus. Plant Genetic Resources. 9:4-18.

Smýkal P, Aubert G, Burstin J, Coyne CJ, Ellis NTH, Flavell AJ, Ford R, Hýbl M, Macas J, Neumann P, McPhee KE, Redden RJ, Rubiales D, Weller JL, Warkentin TD (2012) Pea (*Pisum sativum* L.) in the genomic era. Agronomy. 2:74-115.

Solh MB, Halila HM, Hernandez-Bravo G, Malik BA, Mihov MI, Sadri B (1994). Biotic and abiotic stresses constraining the productivity of cool season food legumes in different farming systems: Specific examples. In: *Expanding the production and use of cool season food legumes.* Springer Netherlands.

Sokhi SS, Jhooty JS, Bains SS (1979) Resistance in pea against powdery mildew. Indian Phytopathology. 32:571-574.

Sonah H, Bastien M, Iquira E, Tardivel A, Légaré G, Boyle B, Normandeau E, Laroche J, Larose S, Jean M, Belzile F (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. PLoS ONE. 8:e54603.

Soto-Cerda BJ, Cloutier S (2012) Association mapping in plant genomes. In: Caliskan M (Ed.), *Genetic diversity in plants*. ISBN: 978-953-51-0185-7. http://www.intechopen.com/books/genetic-diversity-in-plants/association-mapping-in-plant-genomes.

Spindel J, Begum H, Akdemir D, Virk P, Collard B, Redoña E, Atlin G, Jannink J-L, McCouch SR (2015) Genomic selection and association mapping in rice (*Oryza sativa*): effect of trait genetic architecture, training population composition, marker number and statistical model on accuracy of rice genomic selection in elite, tropical rice breeding lines. PLoS Genetics. 11:e1004982.

Stam P (1993) Construction of integrated genetic linkage maps by means of a new computer package: Join Map. The Plant Journal. 3:739-744.

Stelpflug SC, Sekhon RS, Vaillancourt B, Hirsch CN, Robin Buell C, de Leon N, Kaeppler SM (2015) An expanded maize gene expression atlas based on RNA-sequencing and its use to explore root development. The Plant Genome. doi:10.3835/plantgenome2015.04.0025.

Steppuhn H, Volkmar KM, Miller PR (2001) Comparing canola, field pea, dry bean, and durum wheat crops grown in saline media. Crop Science. 41:1827-1833.

Storz G (2002) An expanding universe of noncoding RNAs. Science. 296:1260-1263.

Sutton T, Baumann U, Hayes J, Collins NC, Shi B-J, Schnurbusch T, Hay A, Mayo G, Pallotta M, Tester M, Langridge P (2007) Boron-toxicity tolerance in barley arising from efflux transporter amplification. Science. 318:1446-1449.

Suzuki A, Togawa M, Ohta K (2003) Occurrence of white top of pea caused by a new strain of *Pseudomonas syringae* pv. *pisi.* Plant Disease. 87:1404-1410.

Svabova L, Griga M (2008) The effect of cocultivation treatments on transformation efficiency in pea (*Pisum sativum* L). Plant Cell, Tissue and Organ Culture. 95:293-304.

Tanksley SD (1993) Mapping polygenes. Annual Review of Genetics. 27:205-233.

Taylor JD, Dye DW (1976) Evaluation of streptomycin seed treatments for the control of bacterial blight of peas (*Pseudomonas pisi* Sackett 1916). New Zealand Journal of Agricultural Research. 91:91-95.

Taylor JD, Bevan JR, Crute IR, Reader SL (1989) Genetic relationship between races of *Pseudomonas syringae* pv. *pisi* and cultivars of *Pisum sativum*. Plant Pathology. 38:364-375.

Thompson JD, Shibahara G, Rajan S, Pel J, Marziali A (2012) Winnowing DNA for rare sequences: highly specific sequence and methylation based enrichment. PLoS ONE. 7:e31597.

Thomson MJ (2014) High-throughput SNP genotyping to accelerate crop improvement. Plant Breeding and Biotechnology. 2:195-212.

Thudi M, Upadhyaya HD, Rathore A, Gaur PM, Krishnamurthy L, Roorkiwal M, Nayak SN, Chaturvedi SK, Basu PS, Gangarao NVPR, Fikre A, Kimurto P, Sharma PC, Sheshashayee MS, Tobita S, Kashiwagi J, Ito O, Killian A, Varshney RK (2014) Genetic dissection of drought and heat tolerance in chickpea through genome-wide and candidate gene-based association mapping. PLoS ONE. 9:e96758:e96758.

Timmerman GM, Frew TJ, Weeden NF, Miller AL, Goulden DS (1994) Linkage analysis of *er-1*, a recessive *Pisum sativum* gene for resistance to powdery mildew fungus (*Erysiphe pisi* D.C.). Theoretical and Applied Genetics. 88:1050-1055.

Tiwari KR, Penner GA, Warkentin TD (1997) Inheritance of powdery mildew resistance in pea. Canadian Journal of Plant Science. 77:307-310.

Tiwari KR, Penner GA, Warkentin TD (1998) Identification of coupling and repulsion phase RAPD markers for powdery mildew resistance gene *er1* in pea. Genome. 41:440-444.

Tiwari KR, Penner GA, Warkentin TD (1999) Identification of AFLP markers for powdery mildew resistance gene *er2* in pea. Pisum Genetics. 31:27-29.

Tonguc M, Weeden NF (2010) Identification and mapping of molecular markers linked to *er1* gene in pea. Journal of Plant Molecular Biology and Biotechnology. 1:1-5.

Totad AS, Fakrudin B, Kuruvinashetti MS (2005) Isolation and characterization of resistance gene analogs (RGAs) from sorghum (*Sorghum bicolor* L. Moench). Euphytica. 143:179-188.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-Seq experiments with TopHat and Cufflinks. Nature Protocols. 7:562-578.

Urnov FD, Rebar EJ, Holmes MC, Zhang HS, Gregory PD (2010) Genome editing with engineered zinc finger nucleases. Nature Reviews Genetics. 11:636-646.

Vaid A, Tyagi P (1997) Genetics of powdery mildew resistance in pea. Euphytica. 96:203-206.

Vallejos CE, Tanksley SD (1983) Segregation of isozyme markers and cold tolerance in an interspecific backcross of tomato. Theoretical and Applied Genetics. 66:241-247.

Varshney RK, Chen W, Li Y, Bharti AK, Saxena RK, Schlueter JA, Donoghue MT, Azam S, Fan G, Whaley AM, Farmer AD, Sheridan J, Iwata A, Tuteja R, Penmetsa RV, Wu W, Upadhyaya HD, Yang SP, Shah T, Saxena KB, Michael T, McCombie WR, Yang B, Zhang G, Yang H, Wang J, Spillane C, Cook DR, May GD, Xu X, Jackson SA (2011) Draft genome sequence of pigeonpea (*Cajanus cajan*), an orphan legume crop of resource-poor farmers. Nature Biotechnology. 30:83-89.

Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, Cannon S, Baek J, Rosen BD, Tar'an B, Millan T, Zhang X, Ramsay LD, Iwata A, Wang Y, N W, Farmer AD, Gaur PM, Soderlund C, Penmetsa RV, Xu C, Bharti AK, He W, Winter P, Zhao S, Hane JK, Carrasquilla-Garcia N, Condie JA, Upadhyaya HD, Luo MC, Thudi M, Gowda CLL, Singh NP, Lichtenzveig J, Gali KK, Rubio J,

Nadarajan N, Dolezel J, Bansal KC, Xu X, Edwards D, Zhang G, Kahl G, Gil J, Singh KB, Datta SK, Jackson SA, Wang J, Cook DR (2013) Draft genome sequence of chickpea (*Cicer arietinum*) provides a resource for trait improvement. Nature Biotechnology. 31:240-246.

Vavilov NI (1926) Studies on the origin of cultivated plants. Bulletin of Applied Botany and Plant Breeding. 26:139-248.

Verdier J, Dessaint F, Schneider C, Abirached-Darmency M (2013a) A combined histology and transcriptome analysis unravels novel questions on *Medicago truncatula* seed coat. Journal of Experimental Botany. 64:459-470.

Verdier J, Torres-Jerez I, Wang M, Andriankaja A, Allen SN, He J, Tang Y, Murray JD, Udvardi MK (2013b) Establishment of the *Lotus japonicus* gene expression atlas (LjGEA) and its use to explore legume seed maturation. The Plant Journal. 74:351-362.

Vershinin AV, Allnutt TR, Knox MR, Ambrose MJ, Ellis THN (2003) Transposable elements reveal the impact of introgression, rather than transposition, in *Pisum* diversity, evolution, and domestication. Molecular Biology and Evolution. 20:2067-2075.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. Genetics Selection Evolution. 34:275-305.

Villalta I, Bernet GP, Carbonell EA, Asins MJ (2007) Comparative QTL analysis of salinity tolerance in terms of fruit yield using two *Solanum* populations of F-7 lines. Theoretical and Applied Genetics. 114:1001-1017.

Vos P, Hogers R, Bleeker M, Reijans M, van de Lee T, Hornes M, Frijters A, Pot J, Peleman J, Kuiper M, Zabeau M (1995) AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research. 23:4407-4414.

Voytas DF, Gao C (2014) Precision genome engineering and agriculture: opportunities and regulatory challenges. PLoS Biology. 12:e1001877.

Wade GC (1951) Pea diseases in Tasmania. Tasmanian Journal of Agriculture. 22:40-48.

Wang H, Hwang SF, Turnbull GD, Chang KF, Howard RJ (1999) Ascochyta blight and other diseases of field pea in north eastern Alberta in 1998. Canadian Plant Disease Survey. 79:116-118.

Wang H, Hwang SF, Chang KF, Li SX, Gossen BD, Turnbull GD, Howard RJ (2007) Reaction of pea cultivars to metabolites of *Mycosphaerella pinodes* detected and separated using thin-layer chromatography. International Journal of Agriculture and Biology. 9:387-391.

Wang X, Warkentin TD, Briggs CJ, Oomah BD, Campbell CG, Woods S (1998) Total phenolics and condensed tannins in field pea (*Pisum sativum* L.) and grass pea (*Lathyrus sativus* L.). Euphytica. 101:97-102.

Wang Y, Cheng X, Shan Q, Zhang Y, Liu J, Gao C, Qiu JL (2014) Simultaneous editing of three homoeoalleles in hexaploid bread wheat confers heritable resistance to powdery mildew. Nature Biotechnology. 32:947-951.

Wang Z, Gerstein M, Snyder M (2009) RNA-Seq: a revolutionary tool for transcriptomics. Nature Reviews. 10:57-63.

Ward JA, Ponnala L, Weber CA (2012) Strategies for transcriptome analysis in nonmodel plants American Journal of Botany. 99:267-276.

Wark DC (1954) Factors influencing the development of bacterial blight of peas. Crop and Pasture Science. 5:365-371.

Warkentin TD, Rashid KY, Xue AG (1996) Fungicidal control of powdery mildew in field pea. Canadian Journal of Plant Science. 76:933-935.

Webb A, Cottage A, Wood T, Khamassi K, Hobbs D, Gostkiewicz K, White M, Khazaei H, Ali M, Street D, Duc G, Stoddard FL, Maalouf F, Ogbonnaya FC, Link W, Thomas J, O'Sullivan DM (2015) A SNP-based consensus genetic map for synteny-based trait targeting in faba bean (*Vicia faba* L.). Plant Biotechnology Journal. doi:10.1111/pbi.12371.

Weller JL, Liew LC, Hecht VF, Rajandran V, Laurie RE, Ridge S, Wenden B, Vander Schoor JK, Jaminon O, Blassiau C, Dalmais M, Rameau C, Bendahmane A, Macknight RC, Lejeune-Hénaut I (2012) A conserved molecular basis for photoperiod adaptation in two temperate legumes. Proceedings of the National Academy of Sciences of the United States of America. 109:21158-21163.

Williams JG, Kubelik AR, Livak KJ, Rafalski JA, Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Research. 18:6531-6535.

Wimmer MA, Bassil ES, Brown PH, Lauchli A (2005) Boron response in wheat is genotype-dependent and related to boron uptake, translocation, allocation, plant phenological development and growth rate. Functional Plant Biology. 32:507-515.

Wing R, Zhang H, Tanksley S (1994) Map-based cloning in crop plants: tomato as a model system: I. genetic and physical mapping of jointless. Molecular Genetics and Genomics. 242:681-688.

Wolf JBW (2013) Principles of transcriptome analysis and gene expression quantification: an RNA-Seq tutorial. Molecular Ecology Resources. 13:559-572.

Wright JM, Bentzen P (1994) Microsatellites: genetic markers for the future. Reviews in Fish Biology and Fisheries. 4:384-388.

Wu Y, Bhat PR, Close TJ, Lonardi S (2008) Efficient and accurate construction of genetic linkage maps from the minimum spanning tree of a graph. PLoS Genetics. 4:e1000212.

Wu YQ, Wang YW, Samuels TD (2011) Development of 1,030 genomic SSR markers in switchgrass. Theoretical and Applied Genetics. 122:677-686.

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, Zhou X, Lam T-W, Li Y, Xu X, Wong GK-S, Wang J (2014) SOAPdenovo-Trans: *De novo* transcriptome assembly with short RNA-Seq reads. Bioinformatics. eprint arXiv:1305.6760.

Xu J, Yuan Y, Xu Y, Zhang G, Guo X, Wu F, Wang Q, Rong T, Pan G, Cao M, Tang Q, Gao S, Liu Y, Wang Y, Lan H, Lu Y (2014) Identification of candidate genes for drought tolerance by whole-genome resequencing in maize. BMC Plant Biology. 14:83.

Xue DW, Zhou MZ, Zhang XQ, Chen S, Wei K, Zeng FR, Mao Y, Wu FB, Zhang GP (2010) Identification of QTLs for yield and yield components of barley under different growth conditions. Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology). 11:169-176.

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. The Plant Journal. 37:528-538.

Young JM (2010) Taxonomy of *Pseudomonas syringae.* Journal of Plant Pathology. 92:(1, Supplement), S1.5-S1.14.

Young ND, Cannon SB, Sato S, Kim D, Cook DR, Town CD, Roe BA, Tabata S (2005) Sequencing the genespaces of *Medicago truncatula* and *Lotus japonicus*. Plant Physiology. 137:1174-1180.

Yuan Y, SanMiguel PJ, Bennetzen JL (2003) High-Cot sequence analysis of the maize genome. The Plant Journal. 34:249-255.

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Research. 7:821-829.

Zhang W, Li F, Nie L (2010) Integrating multiple 'omics' analysis for microbial biology: application and methodologies. Microbiology. 156:287-301.

Zhao QY, Wang Y, Kong YM, Luo D, Li X, Hao P (2011) Optimizing *de novo* transcriptome assembly from short-read RNA-Seq data: a comparative study. BMC Bioinformatics. 12(Suppl 14):S2.

Zhu H, Choi HK, Cook DR, Shoemaker RC (2005) Bridging model and crop legumes through comparative genomics. Plant Physiology. 137:1189-1196.

Zhu JK (2001) Plant salt tolerance. Trends in Plant Science. 6:66-71.

Zohary D, Hopf M (1973) Domestication of pulses in the old world. Science. 182:887-894.

Zohary D, Hopf M, Weiss E (2012) Domestication of pulses in the old world: The origin and spread of domesticated plants in Southwest Asia, Europe, and the Mediterranean Basin. Fourth Edition. Oxford University Press, Oxford, UK.